Advances in Intelligent Systems and Computing 685

Thanaruk Theeramunkong
Andrzej M. J. Skulimowski
Takaya Yuizono
Susumu Kunifuji *Editors*

Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems

Selected Revised Papers from the Tenth International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2015), 12–14 November 2015, Phuket, Thailand

Springer

Advances in Intelligent Systems and Computing

Volume 685

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Thanaruk Theeramunkong
Andrzej M. J. Skulimowski
Takaya Yuizono · Susumu Kunifuji
Editors

Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems

Selected Revised Papers from the Tenth International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2015), 12–14 November 2015, Phuket, Thailand

Springer

Editors
Thanaruk Theeramunkong
School of Information, Computer, and Communication Technology (ICT)
Sirindhorn International Institute of Technology
Thammasat University
Pathum Thani
Thailand

Andrzej M. J. Skulimowski
Department of Automatic Control and Biomedical Engineering
AGH University of Science and Technology
Kraków
Poland

Takaya Yuizono
Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
Nomi-shi, Ishikawa
Japan

Susumu Kunifuji
School of Knowledge Science
Japan Advanced Institute of Science and Technology
Nomi-shi, Ishikawa
Japan

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-70018-2 ISBN 978-3-319-70019-9 (eBook)
https://doi.org/10.1007/978-3-319-70019-9

Library of Congress Control Number: 2017956755

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains a number of selected papers that were presented at the 10th International Conference on Knowledge, Information and Creativity Support Systems (KICSS'2015) in Phuket, Thailand, during November 12–14, 2015. Among 45 submitted papers, we carefully reviewed and accepted 13 regular papers (28.89% of them), five short papers (11.11% of them), and six workshop papers with aiming to explore diversity of current issue across various themes. All selected papers included in this series deeply discuss the latest advances in blending of innovative thinking support systems, information science, knowledge science, service science, and creativity science in order to bringing together an innovation design model combined with knowledge management, management of technology, and management of service.

Finally, we would be grateful to all authors for their contribution through academic researches. We also would like to thank the International Conference on Knowledge, Information and Creativity Support Systems (KICSS'2015) Executive Committees and Program Co-chairs for entrusting us to process and deliver valuable academic researches to envision a shape of future aspects in the Creativity Support System research area.

Pathum Thani, Thailand — Thanaruk Theeramunkong
Kraków, Poland — Andrzej M. J. Skulimowski
Nomi-shi, Japan — Takaya Yuizono
Nomi-shi, Japan — Susumu Kunifuji
June 2017

Organization

Organizing Committee

Honorary Co-chairs

Somnuk Tangtermsirikul, Thammasat University, Thailand
Sarun Sumriddetchkajorn, NECTEC, Thailand
Hiroshi Motoda, Osaka University, Japan

General Co-chairs

Susumu Kunifuji, JAIST, Japan
George Angelos Papadopoulos, University of Cyprus, Cyprus
Vilas Wuwongse, Asian University, Thailand

International Advisory Committee

Chang Duk Jung, Korea University, Korea
Ho Tu Bao, JAIST, Japan
Kok-Leong Ong, Deakin University, Australia
Pham Tho Hoan, Hanoi National University of Education, Vietnam
Sanparith Marukatat, NECTEC, Thailand
Takeo Higuchi, IMS Institute, Japan
Thomas Koehler, Dresden University of Technology, Germany
Vincent C.S. Lee, Monash University, Australia
Xijin Tang, Chinese Academy of Sciences, China

Program Co-chairs

Thanaruk Theeramunkong, Thammasat University, Thailand
Takaya Yuizono, JAIST, Japan
Andrzej M.J. Skulimowski, P&BF/AGH, Poland

Publicity Co-chair

Kiyota Hashimoto, PSU, Phuket Campus, Thailand

Organizing Co-chairs

Thepchai Supnithi, NECTEC, Thailand
Rattana Wetprasit, PSU, Phuket Campus, Thailand
Nattapong Tongtep, PSU, Phuket Campus, Thailand

Workshop Co-chairs

Virach Sortlertlamvanich, Thammasat University, Thailand
Teerayut Horanont, Thammasat University, Thailand
Ryoju Hamada, Thammasat University, Thailand
Takayuki Ito, Nagoya Institute of Technology, Japan
Naoki Fukuta, Shizuoka University, Japan
Aziz Nanthaamornphong, PSU, Phuket Campus, Thailand

Publication Co-chairs

Prachya Boonkwan, NECTEC, Thailand
Nongnuch Ketui, RMUTL, Nan, Thailand
Ilya Sadeghi Ardakani, Thammasat University, Thailand
Christoph Senn, Thammasat University, Thailand
Uraiwan Buatoom, Thammasat University, Thailand
Siriwon Taewijit, Thammasat University, Thailand

Post-Proceedings Production Assistant

Siriwon Taewijit, Thammasat University, Thailand

Financial Co-chair

Nichnan Kittiphattanabawon, Walailak University, Thailand

Contents

Detecting TV Program Highlight Scenes Using Twitter Data Classified by Twitter User Behavior 1
Tessai Hayama

BrainTranscending: A Hybrid Divergent Thinking Method that Exploits Creator Blind Spots 14
Aya Hasebe and Kazushi Nishimoto

Open Data Integration for Lanna Cultural Heritage e-Museums 29
Watchira Buranasing, Marut Buranarach, and Chai Wutiwiwatchai

RDF Dataset Management Framework for Data.go.th 42
Pattama Krataithong, Marut Buranarach, and Thepchai Supnithi

Probabilistic Argumentation for Service Restoration in Power Distribution Networks 54
Nguyen Duy Hung, Santi Kaisaard, and Saung Hnin Pwint Oo

Experimental Teaching of Collaborative Knowledge Creation Using Analogue Games 69
Tomomi Kaneko, Ryoju Hamada, and Masahiro Hiji

Ontology-Based Framework to Support Recommendation Rule Management Using Spreadsheet 77
Marut Buranarach, Thanyalak Rattanasawad, and Taneth Ruangrajitpakorn

Factuality Classification Using Multi-facets Based on Elementary Discourse Units for News Articles 87
Khaing Swe Wynn and Sasiporn Usanavasin

A Recommender of Transportation Planning in Campus Using Ontology 101
Suwit Somsuphaprungyos, Salin Boonbrahm, Poonpong Boonbrahm, and Taneth Ruangrajitpakorn

TLex+: A Hybrid Method Using Conditional Random Fields and Dictionaries for Thai Word Segmentation 112
Sarawoot Kongyoung, Anocha Rugchatjaroen, and Krit Kosawat

An Expert System Using Ontology as Knowledge Base for Personalized Rice Cultivation Suggestion 126
Wichuphon Chariyamakarn, Poonpong Boonbrahm, Salin Boonbrahm, and Taneth Ruangrajitpakorn

A Knowledge-Based Approach for Generating UML Sequence Diagrams from Operation Contracts 141
Nasith Laosen and Ekawit Nantajeewarawat

Designing a Social Innovation Based Knowledge Support System: A Preliminary Guideline 156
Ahmet Suerdem, Baskin Yenicioglu, and Ozge Demir

A Study of Cost, Revenue, and Profit Efficiency of Commercial Banks in Cambodia Using DEA Approach 168
Sokim Phang and Morrakot Raweewan

Mapping Knowledge Management for Technology Incubation 179
Jarot S. Suroso, Abba Suganda Girsang, and Ford Lumban Gaol

Competition Boosts Innovativeness: The Case of CMS Evolution 188
Andrzej M. J. Skulimowski and Inez Badecka

Large Scale Mobility Analysis: Extracting Significant Places Using Hadoop/Hive and Spatial Processing 205
Apichon Witayangkurn, Teerayut Horanont, Masahiko Nagai, and Ryosuke Shibasaki

Modeling Traffic Accidents Occurrences Based on Land Use and Road Factors Using Geographically Weighted Regression Models 220
Paweenuch Songpatanasilp, Teerayut Horanont, Harutoshi Yamada, and Ryosuke Shibasaki

An Innovative Outdoor IoT System Architecture for Service Oriented Things 233
Takanobu Otsuka, Yoshitaka Torii, and Takayuki Ito

Ontology-Based Architecture to Improve Driving Performance Using Sensor Information for Intelligent Transportation Systems 245
Susel Fernandez and Takayuki Ito

An Analysis of Allocation Stability on Approximation-Based Pricing for Multi-unit Combinatorial Auctions 256
Naoki Fukuta

Revenue Improvement Mechanisms Based on Advertisement's Values and Layouts 270
Satoshi Takahashi and Tokuro Matsuo

A Model-Based Method for Modeling and Verifying Event-Based Aspect-Oriented Applications 281
Hong Anh Le, Ninh Thuan Truong, and Aziz Nanthaamornphong

A Retrospective on Constructing a Personal Narrative on Agile Development 290
Roy Morien

Detecting TV Program Highlight Scenes Using Twitter Data Classified by Twitter User Behavior

Tessai Hayama(✉)

School of Information and Management Systems Engineering,
Nagaoka University of Technology, 1603-1, Kamitomioka Nagaoka, Niigata 940-2188, Japan
t-hayama@kjs.nagaokaut.ac.jp
http://www.nagaokaut.ac.jp

Abstract. This paper presents a novel TV event detection method for automatically generating TV program digests by using Twitter data. Previous studies of TV program digest generation based on Twitter data have developed TV event detection methods that analyze the frequency time series of tweets that users made while watching a given TV program; however, in most of the previous studies, differences in how Twitter is used, e.g., sharing information versus conversing, have not been taken into consideration. Since these different types of Twitter data are lumped together into one category, it is difficult to detect highlight scenes of TV programs and correctly extract their content from the Twitter data. Therefore, this paper presents a highlight scene detection method to automatically generate TV program digests for TV programs based on Twitter data classified by Twitter user behavior. To confirm the effectiveness of the proposed method, experiments using a TV soccer program were conducted.

Keywords: Highlight-scene detection · TV digest · Twitter data
Text analysis

1 Introduction

This paper describes a novel TV-event detection method for automatically generating TV program digests using Twitter data. TV program digests are useful for understanding the overview of a given TV program and for quickly searching for interesting scenes in the TV program. They are therefore often used in news program, program propaganda, and so on; however, high costs in terms of time and labor are required to manually create digests for such TV programs. Here it is necessary to find and highlight scenes in the TV programs and then create indices for these highlighted scenes. To reduce costs, automatic video digest generation methods have been developed using the audiovisual features of the

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_1

video [2,12,13]. Though such methods successfully extract well-defined scenes from videos with high accuracy, it is difficult for these methods to extract undefined highlight scenes; moreover, it is also difficult to assign indices that include various types of information, including keywords and impression words, to the extracted scenes.

In recent years, the number of users who use Twitter while watching TV programs has increased. Users discuss and share details and opinions of what happens in the given TV program on Twitter in real time, such that there are a huge number of live tweets regarding the TV programs[1] [7]. Previous studies have developed TV event detection methods for TV program digest generation based on live tweets corresponding to the given TV programs [4–6,8,9]. These methods analyze trends in the frequency time series of the tweets that users made while watching a given TV program. Since Twitter users behave in different ways, for example, conversing or sharing information, Twitter content is not always synchronized with the TV program content in real time. It is therefore difficult to correctly detect the highlight scenes of TV programs from the Twitter data.

Given the above problems, this paper proposes a highlight scene detection method for TV programs based on Twitter data categorized by Twitter user behavior. Using the categorized Twitter data, the proposed method detects highlight scenes of TV programs and assigns metadata encompassing the keywords included in the tweets to the detected scenes. The effectiveness of the proposed method has been confirmed by experiments conducted using a TV soccer program.

2 Related Studies

To automatically generate sports TV digests, some researchers have analyzed visual and audio features of sports TV program videos, for example, recognizing goal posts within soccer video scenes to detect goal scenes of a soccer game [13], developing an image/speech hybrid recognition method for detecting highlight scenes of soccer game videos, developing a Hidden Markov Model (HMM) method that assigns labels to highlight scenes of soccer game TV programs by using the speech features of the videos [12], and developing an emotion model constructed from visual and audio features of videos to detect highlight scenes of soccer game TV programs [2]. Although the above studies have been able to accurately detect the main scenes within well-defined visual and audio features, it has been impossible for them to detect undefined main scenes or their achievements have been prone to false positives despite high recall rates.

Also, in recent years, researchers have developed highlight scene detection methods for TV programs using Twitter data, based primarily on the analysis of the frequency time series of tweets that users made while watching a given TV program. The literature [6] describes the ability of such methods to detect highlight scenes of TV programs using Twitter data with accuracies similar to

[1] https://www.nielsen.com/us/en/press-room/2012/nielsen-and-twitter-establish-social-tv-rating.html.

that of the aforementioned audio and visual analysis. Some approaches have analyzed Twitter data using categories of tweets and Twitter users to detect highlight scenes of TV programs based on a variety of viewpoints. As an example of tweet categorization, tweets may be categorized into groups based on such emotional representations as emotion words [1] and emoticons [14]; then the main scenes of the given TV program are differently detected on the basis of these categories. As an example of user categorization, Twitter users may be separated by the fun side of each team to detect highlight scenes from opposite viewpoints in such team games as soccer [10], baseball [4], and football [11]. Based on the tweets of one of these separated users, enjoyable and exciting scenes of the game can be detected for each fun side.

These previous studies have analyzed relationships between highlight scenes of a TV program and tweets made during the time of a given scene by observing the trends of frequency time series of tweets and adapting Twitter data with a burst detection method. Twitter data used in previous studies have been collected by hashtags and keywords; then all such data has been used for analysis. Since Twitter user behavior manifests in different ways, Twitter content is not always synchronized with the content of the TV program in real time. It is therefore difficult for the previous approaches to correctly detect highlight scenes of TV programs from Twitter data. As noted above, this paper introduces the categorization of user behavior into highlight scene detection. The approach is different from the previous studies in that tweet categorization is based on user behavior, not on linguistic representation.

3 Proposed Method

This paper proposes a novel method to use Twitter data to detect highlight scenes of TV programs and extract a bag-of-words as metadata of the scenes. To achieve this, the method categorizes Twitter data into user types during a timeslot of the TV program, and then properly applies one of the categorized Twitter data to the detection method of the highlight scenes. As shown in Fig. 1, the procedure of the proposed method consists of four steps. In the first step, Twitter users generate tweets while watching the TV program; these tweets are categorized into groups based on user types during the given timeslot of the TV program. In the second step, the frequency time series of the tweets for each user type is created. In the third step, the burst points of the frequency time series are automatically detected to decide the time periods of the highlight scenes of the TV program. In the final step, keywords for each the highlight scene of the TV program are extracted from the tweets of the detected time period and assigned to each highlighted scene as metadata.

In the given procedure, highlight scenes of the TV program and their metadata are identified with a burst detection method using the Twitter data. If this approach properly adapts the tweets of certain user types to the burst detection method, the approach performs better than an approach that lumps user types into one category. For example, when a user shares information regarding a TV

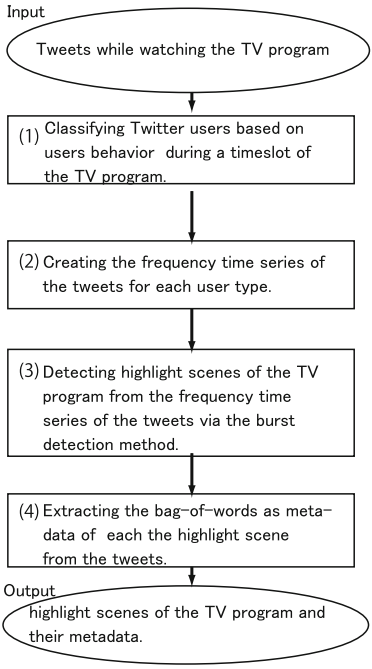


Fig. 1. Procedure of the proposed method

program on Twitter while watching the TV program, the user's tweets are likely to include content phrases of the TV program. Conversely, when a user communicates his or her impression of the TV program to other users, the user's tweets are likely to be out of sync with the content of the TV program and likely not include enough information regarding the TV program, instead including brief feedback, off-topic conversations, and the like-though it is interesting to include words that represent the user's impression of the TV program. Thus, the proposed method introduces analysis of user behavior into the detection method of the highlight scenes of TV programs such that remarkable scenes and their metadata are extracted with high accuracy from the categorized Twitter data.

Details of each step of the procedure are described below:

(1) Classifying Twitter users based on user behavior during a timeslot of the TV program.

Twitter users tweeting while watching the given TV program are classified into types of similar tweet behavior using cluster analysis. Considering how Twitter is used, the following four features were chosen for cluster analysis and investigated for an hour-long TV program for every user:

- The number of retweets and reply-tweets, which represents the degree of relationships between the given user and other users. If users often generate

tweets that include retweets or reply-tweets, users place a priority on the Twitter conversation.
- The number of hashtags, which represents the degree of contribution of information sharing. The hashtag of a topic is useful for retrieving tweets of the given topic. If users often generate tweets that include hashtags with regard to the TV program, users put a priority on information sharing among Twitter users who are interested in the same TV program.
- The average number of characters, which represents the degree of informative content. If users often generate tweets that include large numbers of characters, the tweets are likely to include large numbers of keywords regarding the given TV program.
- The average number of tweets, which represents the sensitivity in regards to the TV program and other tweets. If users make a large number of tweets during the timeslot of the TV program, they tend to sensitively respond to the main scenes of the TV program.

Note that the current system uses the Ward method for calculating distances in the cluster analysis.

(2) Creating the frequency time series of the tweets for each user type.

To create the frequency time series of the tweets for each user type, tweet frequencies for certain a time span are calculated and arranged into the timeline of the TV program. The time span of the current system is set to 5 s, which was determined by investigating the time gaps between highlight scenes of the TV program and the tweets regarding these scenes.

(3) Detecting highlight scenes of the TV program from the frequency time series of the tweets via the burst detection method.

To detect highlight scenes of the TV program, the unusual increases in the frequency time series of the tweets must be found. The current system adopts Kleinberg's method [3] for burst detection; this method is based on a HMM algorithm that automatically detects abnormal time spans and their burst level on time-series data. The burst detection procedure consists of the following three steps: (a) an emergence possibility for each burst-level state of every time span position is calculated based on the tweet frequency of the position; (b) all time sequences of the burst-level states are found and their costs are calculated based on the emergence possibilities included in the time sequences; and (c) the time sequence of the burst-level with the minimum cost is chosen from all time sequences, and the time periods with the burst in the sequences are determined as the highlight scenes. Note that the system adopts the Viterbi algorithm to choose the sequence with the minimum cost.

(4) Extracting the bag-of-words for each highlight scene from the tweets.

To extract the bag-of-words of each detected scene, the words included in the tweets are investigated using the burst detection method; then the unusual increasing words in the time window of the detected scene are determined as the metadata of the scene. The words for the analysis are manually chosen based

on category, such as “object,” “situation,” “impression,” and “viewpoint.” The burst detection method of the system uses the same algorithm as the highlight scene detection algorithm.

4 Experiment

4.1 Overview

This paper proposes a method to detect remarkable scenes and their metadata with high accuracy using Twitter data categorized by user behavior. In the experiment, the proposed method was compared to a burst detection method that adopts noncategorized Twitter data; the focus of the experiment was the accuracy of highlight scene detection of a TV program.

For the experimental data, an international soccer game(i.e., Japan versus Netherlands) broadcast in Japan on November 16, 2013 was used, as was Japanese Twitter data consisting of 376,656 tweets collected from 51,565 Twitter followers of soccer player accounts. The properties of the experimental data are as shown in Figs. 2, 3 and 4.

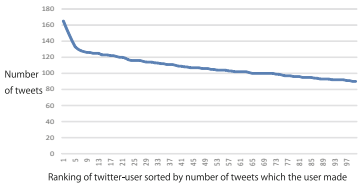


Fig. 2. Number of tweets which each twitter-user made in the experimental data

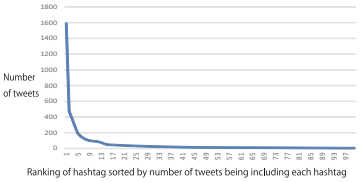


Fig. 3. Number of hashtags being including in the tweets which each twitter-user made in the experimental data

As shown in Fig. 2, the highest number of tweets which a twitter-user made are 165 and the average number of tweets of top 100 users is 108.38 in the experimental data. As shown in Fig. 3, the highest number of hashtags being including in the tweets which each twitter-user made is 1586 and the average number of hashtags of top 100 hashtags is 51.35 in the experimental data. As shown in Fig. 4, the highest number of retweet being including in the tweets which each

twitter-user made is 537 and the average number of retweets of top 100 users is 63.5 in the experimental data.

As shown in Table 1, the word lists for the metadata extraction were chosen from the top 300 high frequency words included in the Twitter data.

Table 1. Categories of metadata and corresponding example words used in the experiment

Category	Examples of word list
Object (Player name)	'本田 (Honda)', '大迫 (Osako)', '内田 (Uchida)', '長友 (Nagatomo)', '柿谷 (Kakitani)', '遠藤 (Endo)', '長谷部 (Hasebe)', '山口 (Yamaguchi)', '岡崎 (Okazaki)', '西川 (Nishikawa)', '香川 (Kagawa)', 'ロッベン (Robben)', 'ファンデルファールト (Ferdinand Van Der Vaart)'
Situation	'ゴール (goal)', 'シュート (shoot)', '点 (score)', 'パス (pass)', 'サイド (wing)', '交代 (substite)'
Impression	'ああ/あー (ah)', '笑 (ha-ha)', 'おお (oh)', 'すごい (fine)', 'やばい (excellent)', 'ミス (miss)', 'ナイス/素晴らしい (nice)', 'おしい (almost)'
Veiwpoint:	'日本 (Japan)', 'オランダ (Netherland)'

The dataset for this experiment was generated from live news feeds of the soccer game on the Web,[2] which reported the main events, such as shots, goals, and substitutes, as well as their times. To evaluate the validity of the metadata, the bag-of-words for each highlight scene was manually judged as being at one of three levels: match, partial match, and nonmatch. For these judgments, the evaluations of metadata tags "object" and "situation" were determined by the corresponding degree between the extracted words and the words included in the live news feed document. For example, the evaluation of "match" was assigned if the words were included in the live news feed document of the event, while the evaluation of "partial match" was assigned if most of the extracted words were included in the live news feed document, and "non-match" was assigned if the extracted words were not included in the live news feed document at all (or if the words included in the live news feed document were not extracted).

The evaluation of metadata tag "impression" was determined by the corresponding degree of positive and negative words between the TV program event and the extracted words of the event. For example, if the positive words are extracted in the case of the Japanese team's goals and shots and the negative

[2] http://live.sportsnavi.yahoo.co.jp/live/soccer/japan/jpn_20131116_01.

words are extracted in the case of the opposing team's goals and shots, the judgment is assigned to "match." The evaluation of metadata tag "viewpoint" was determined according to information of each event of the live news feed (i.e., Japan or Netherlands).

4.2 Results

(1) Classifying Twitter users based on user behavior

Table 2 shows the results of classifying Twitter users based on user behavior.

Table 2. Results of classifying Twitter users based on user behavior

Group ID	Number of the users	Average number of tweets	Use rate of hashtag	Use rate of retweet	Average number of characters	Label
1	1,474	9.01	0.84	0.07	42.09	Heavy use of hashtags
2	19,669	11.00	0.03	0.49	19.84	Heavy use of retweets
3	11,944	4.09	0.11	0.98	16.41	parallel use of retweets and plain-tweets
4	18,482	5.31	0.00	0.01	26.10	Heavy use of plain-tweets

In the results of the classification, user behavior was categorized into four distinct types with the following features. Users in Group ID 1 generated tweets that included hashtags at the highest rate (i.e., 0.84) and contained the longest sentences (i.e., average number of characters was 42.09). Users in Group ID 2 generated tweets that included retweets and reply-tweets in the middle (i.e., 0.49). Users in Group ID 3 generated tweets that included retweets and reply-tweets at the highest rate (i.e., 0.98). And users in Group ID 4 generated tweets that included hashtags at the lowest rate (i.e., 0.00) and retweets at the lowest rate (i.e., 0.01). Therefore, user behavior while watching the TV program was categorized by the obvious features. In this study, user behavior of Group IDs 1, 2, 3, and 4 were labeled "heavy use of hashtags," "heavy use of retweets," "parallel use of both retweets and plain tweets," and "heavy use of plain tweets," respectively.

(2) Detecting highlight scenes of the TV program from the frequency time series of tweets based on the classification of user behavior

Figure 5 shows the results of detecting highlight scenes of the TV program from the frequency time series of tweets based on the classification of user behavior. The TV program of the experimental data included nine highlight scenes of the following types: goal scenes of both teams, shot scenes of the Japanese team, and substitute scenes. In the results, using Twitter data with a user type of "heavy use of plain tweets," all goal scenes of the TV program were detected and the burst levels of the detected scenes were assigned a 1 or 2. Using Twitter data with the other user types, fewer scenes of the TV program were detected, and the burst levels of the detected scenes were assigned only a 1. Thus, to detect highlight scenes of the TV program from the frequency time series of the tweets, using Twitter data with the user type of "heavy use of plain tweets" was more sensitive than using Twitter data with the other user types. Conversely, nothing was detected using Twitter data with a type of "heavy use of retweets," and half of the highlight scenes were detected using "parallel use of both retweets and plain tweets." Using the noncategorized Twitter data, five of the nine highlight scenes were detected. Therefore, we conclude here that it is useful to detect highlight scenes of a given TV program by using Twitter data classified based on user behavior.

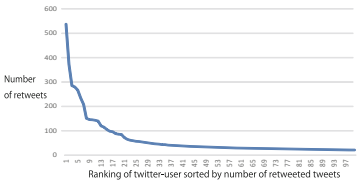


Fig. 4. Number of retweets being including in the tweets which each twitter-user made in the experimental data

(3) Extracting the bag-of-words of each highlighted scene

Table 3 shows evaluation results of extracting the bag-of-words for each high-lighted scene from the tweets.

In these results, the bag-of-words for "Object," "Situation," and "Viewpoint" were extracted most correctly using the tweets made by users of the type "heavy use of plain tweets." For "Object," the corresponding degree between extracted words and high-lighted scenes was evaluated with four scenes as a "match" and five scenes as a "partial match." In "Situation," the corresponding degree was evaluated with four scenes as a "match" and five scenes as a "partial match." In "Viewpoint," the corresponding degree was evaluated with five scenes as a "match," three scenes as a "partial match," and one scene as a "non-match." The extraction using tweets generated by users of the type "heavy use of plain tweets" was more accurate than the extraction using tweets of the other user types. The bag-of-words for "Emotion" was extracted most accurately using

Table 3. Evaluation results of extracting the bag-of-words of each highlighted scene from the tweets

Category of metadata	Type of heavy use of hashtags			Type of parallel use of retweets and plain-tweets			Type use of retweets			Type of heavy use of plain-tweets			Noncategorized tweets		
	M	PM	UM	M	PM	UM	M	PM	UM	M	PM	UM	M	PM	UM
Object	4	3	2	2	7	0	3	3	3	4	5	0	3	6	0
Situation	2	4	3	7	0	2	3	1	5	8	1	0	7	0	2
Impression	4	2	3	7	2	0	1	2	6	4	2	3	7	2	0
Viewpoint	5	3	1	3	2	4	2	1	6	5	3	1	3	2	4

M: match, PM: partial match, UM: unmatch

tweets generated by users of the type "parallel use of retweets and plain tweets"; the corresponding degree was evaluated with seven scenes as a "match" and two scenes as a "partial match."

Conversely, though the bag-of-words for the categories of metadata were extracted correctly to some extent using noncategorized tweets, the bag-of-words for these categories were extracted with higher accuracies using tweets with one of the user types.

4.3 Discussion

Through this experiment, the effectiveness of the proposed method was confirmed as follows:

- Twitter user behavior while watching the given TV program was categorized into four types, i.e., "heavy use of hashtags," "heavy use of retweets," "parallel use of retweets and plain tweets," and "heavy use of plain tweets."
- It is useful to detect highlight scenes of the given TV program by using tweets generated by users of the type "heavy use of plain tweets."
- It is also useful to extract the bag-of-words as the metadata for "Object," "Situation," and "Viewpoint" from the tweets generated by users of the types "heavy use of plain tweets" and "parallel use of retweet and plain tweets."

The summarized results of applying the proposed method to the given TV program are shown in Fig. 6. In the results, the content of the TV program is offered by the detected highlighted scenes and the extracted metadata. Therefore, it is useful to introduce tweet categorization based on user behavior into detecting highlight scenes of a TV program and assigning the metadata of the bag-of-words to the scenes via the Twitter data.

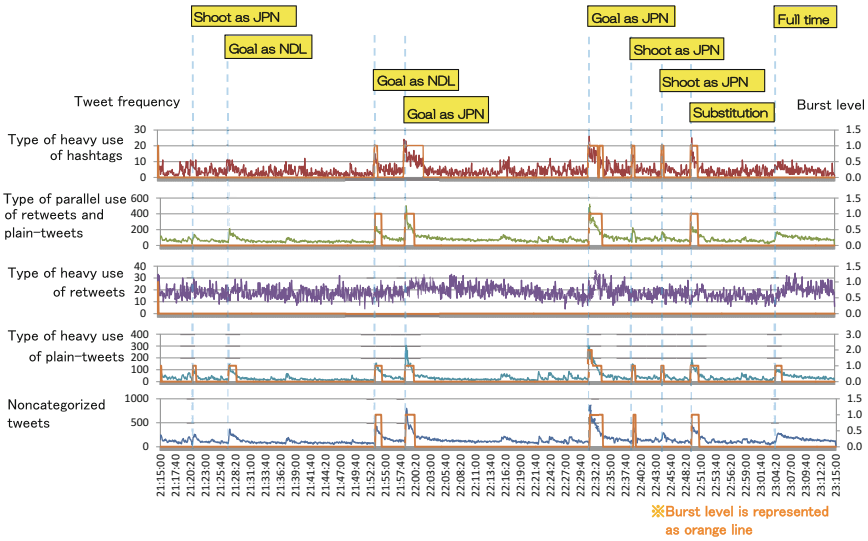


Fig. 5. Results of detecting highlight scenes of a TV program from the frequency time series of tweets based on the classification of user behavior

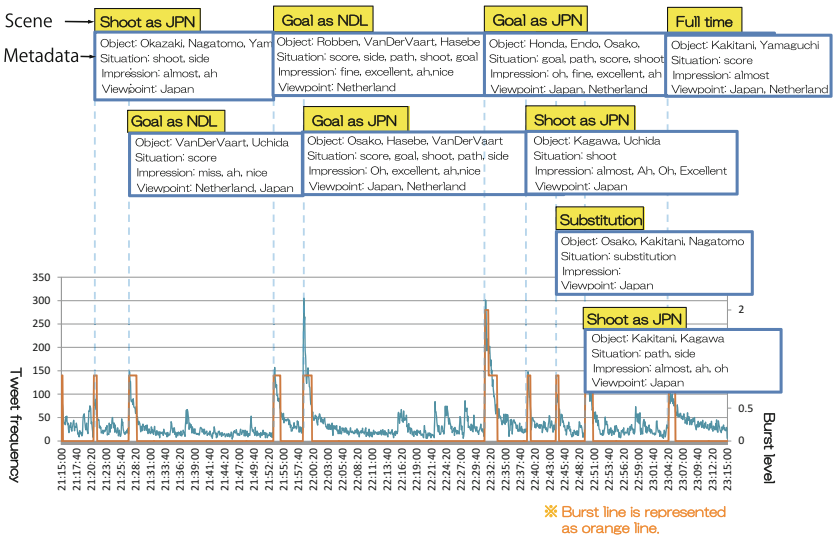


Fig. 6. Summarized results of detecting highlighted scenes of the given TV program and extracting their metadata from tweets based on the classification of the Twitter user behavior

5 Conclusions

In this paper, a novel method for detecting highlight scenes of a TV program is proposed; this method also assigns the bag-of-words as metadata to the corresponding scenes. Previous related studies have dealt with Twitter data by lumping such data into one category and then detecting the highlight scenes of a TV program. The approach proposed in this paper introduced the analysis of user behavior into the method used to detect highlight scenes of TV programs. More specifically, the proposed method properly adapts tweets of one user type for correctly detecting highlight scenes and extracting their metadata. By conducting an experiment using a TV soccer program, the effectiveness of the proposed method was confirmed.

In future work, the proposed method will be applied to various TV programs, thus investigating the possibility of detecting highlight scenes of these programs and extracting the corresponding metadata. Further, a digest generation method for TV programs will be developed based on various kinds of viewpoints using the results the proposed method generates.

Acknowldgements. This work was supported by JSPS KAKENHI Grant Number 25730210.

References

1. Doman, K., Tomita, T., Ide, I., Deguchi, D., Murase, H.: Event detection based on Twitter Enthusiasm Degree for generating a sports highlight video. In: Proceedings 22nd ACM International Multimedia Conference (ACM-MM2014), pp. 949–952 (2014)
2. Hanjali, A.: Adaptive extraction of highlights from a sport video based on excitement modeling. IEEE Trans. Multimed. **7**(6), 1114–1122 (2005)
3. Kleinberg, J.: Bursty and hierarchical structure in streams. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1–25 (2002)
4. Kobayashi, T., Takahashi, T., Deguchi, D., Ide, I., Murase, H.: Detection of biased broadcast sports video highlights by attribute-based tweets analysis. Lect. Note Comput. Sci. **7733**, 364–373 (2013)
5. Kubo, M., Sasano, R., Takamura, H., Okumura, M.: Generating live sports updates from twitter by finding good reporters. The 2013 IEEE/WIC/ACM International Conference on Web Intelligence (WI 13) (2013)
6. Lanagan, J., Smeation, A.F.: Using twitter to detect and tag important events in live sports. In: ICWSM 2011, pp. 542–545 (2011)
7. Mizunuma, Y., Yamamoto, S., Yamaguchi, Y., Ikeuchi, A., Satoh, T., Shimada, S.: Twitter bursts: analysis of their occurrences and classifications. The Eighth International Conference on Digital Society, ICDS 2014, pp. 182–187 (2014)
8. Nakahara, T., Hamuro, Y.: Detecting topics from Twitter posts during TV program viewing. In: Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW '13), pp. 714–719 (2013)

9. Nakazawa, M., Erdmann, M., Hoashi, K., Ono, C.: Social indexing of TV programs: detection and labeling of significant TV scenes by Twitter analysis. In: Proceedings of the 2012 26th International Conference on Advanced Information Networking and Applications Workshops, pp. 141–146 (2012)
10. van Oorschot, G., van Erp, M., Dijkshoorn, C.: Automatic extraction of soccer game events from Twitter. In: Proceedings of the Workhop on Detection, Representation, and Exploitation of Events in the Semantic Web DeRiVE 2012, pp. 21–30 (2012)
11. Tang, A., Boring, S.: #EpicPlay: crowd-sourcing sports video highlights. In: ACM International Conference on Human Factors in Computing Systems (CHI'12), pp. 1569–1572 (2012)
12. Wang, J., Xu, C., Chng, E., Tian, Q.: Sport highlight dectection from keyword sequences using HMM. In: Proceedings of the IEEE ICME, pp. 27–30 (2004)
13. Yamamoto, T., Shimizu, D., Watanabe, M.: Research of automatic scene analysis for soccer broadcast images. Technical report of IEICE. PRMU, vol. 104(573), pp. 73–78 (2005). (in Japanese)
14. Yamauchi, T., Yukiko, H., Nakano, I.: Searching emotional scenes in TV programs based on twitter emotion analysis. In: Proceedings of the 5th international conference on Online Communities and Social Computing, pp. 21–26 (2013)

BrainTranscending: A Hybrid Divergent Thinking Method that Exploits Creator Blind Spots

Aya Hasebe(✉) and Kazushi Nishimoto

Japan Advanced Institute of Science and Technology, 1-1, Asahidai, Nomi, Ishikawa 923-1292, Japan
aya.hasebe.7@gmail.com, knishi@jaist.ac.jp
http://www.jaist.ac.jp/ks/labs/knishi/

Aya Hasebe is currently with Yahoo Japan Co., Ltd.

Abstract. Divergent thinking methods play a very important role in creating new ideas for planning and/or development activities based on specific themes or existing objects. However, even when using existing divergent thinking methods, it remains difficult to conceptualize truly novel ideas because creators are typically constrained by their fixed ideas. Therefore, we propose a novel divergent thinking method named "Brain-Transcending (BT)" that exploits Brainstorming (BS), a typical divergent thinking method, as a way to identify the creator's fixed ideas, rather than generate ideas, and thus support the further expansion of those ideas. We conducted user studies and confirmed that the number of idea groups that were created by BT significantly increased compared to typical brainstorming. Furthermore, the subjects evaluated the final ideas created with BT and determined that the quality was better than that of BS. We also propose the "Reduced BrainTranscending (RBT)" method for alleviating the heavy cognitive load that BT imposes on creators. We conducted user studies of RBT to investigate its usefulness and, consequently, concluded that BT and RBT are effective methods for supporting divergent thinking aimed at improving existing products.

Keywords: Divergent thinking · Brainstorming · Fixed ideas · Blind spots

1 Introduction

Creation of novel ideas and products based on existing ideas and products is often done in the fields of planning and development. In such activities, it is very important to collect as many various seeds of new ideas as possible from a broad

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_2

range of viewpoints [10]. To support it, various divergent thinking methods have been developed and utilized. Brainstorming (BS) [10] is one of the most famous methods, and it has been widely used. A BS session is usually conducted by a group, whose members try to create ideas while following four rules:

1. Focusing on quantity, not on quality,
2. Withholding criticism,
3. Welcoming unusual ideas, and
4. Combining and improving ideas.

By following these rules and referring to the other members' ideas, it is expected that each member becomes able to create many novel ideas.

However, a lot of experimental studies indicated that people in face-to-face brainstorming meetings are less efficient at generating ideas than when working alone [12]. These results have been mainly attributed to features of group activities, that is, social loafing, evaluation apprehension and production blocking [2]. Furthermore, if there is a member with a dominant personality, the direction of the entire thinking may be led by his/her ideas, or some other members may feel pressure to withhold their ideas.

To solve these problems deriving from group thinking, a lot of various attempts have been done. Involving a facilitator in the group's discussion, to coordinate the balance of utterance frequencies and encourage inactive members, is one solution to these problems [10]. The Brain-writing (BW) [13] method, which is an improved BS method, can be considered superior to BS because the problems of the dominant person and unbalanced utterance frequencies never arise. Therefore, social loafing and production blocking can be avoided by BW.

Not only such apporaches to improve the BS methodology itself, but also many approaches to create support systems for the group BS have been attempted. Gallupe et al. [2] showed that an on-line communication tool for group BS sessions (so-called "electronic brainstormring") is effective to solve these problems. Presmo Brainstormer [7] is a web-based electronic brainstoming system that supports wider phases in idea generation process including problem definition, idea generation, group discussion and presentation. Idea Expander [14] dynamically retrieves and shows pictorial stimuli based on conversational content in a group brainstorming session. Meeting Mediator [6] detects social interactions among people. For example, it can detect who are dominant people in a group and measure their influence on other people. By utilizing this system, abovementioned dominant people problem can be alleviated. Furthermore, recently, large-scale ideation by hundreds of (or more) people has been attempted by utilizing crowdsourcing [1]. This can be another promising approach to solve the three problems.

Thus, solving the problems of BS has been mainly addressed from perspectives on group. However, it is also very important to address to problems deriving from the individual to enhance the productivity of BS. Regardless of either nominal groups or interactive groups, productivity of idea creation depends on individuals performances. Generally, everybody has some fixed ideas. Such fixed ideas are formed from our experiences and the knowledge we have obtained. In this sense, fixed ideas can be regarded as an intellectual basis for creating

new ideas. On the other hand, they also constrain the free creation of ideas. In creating ideas, even if there are actually possibilities of new ideas, the creator's thinking tends to adhere to his/her fixed ideas, which leads to the creator going around in circles and finally arriving at a deadlock situation [9]. Therefore, to support and enhance the individual idea creation, it is important to provide effective clues that support him/her in thinking beyond his/her fixed ideas.

Osborn's checklist method [10], which is one of the forced relationship techniques, provides nine viewpoints, such as "Magnify" and "Reverse," in a checklist and forces a creator to create ideas based on each viewpoint. As a result, he/she can avoid overlooking some of these viewpoints. However, it is difficult to fully adapt such a generic checklist to someone's own ways of thinking and to specific problems.

Providing related information from knowledge resources outside can be one of the good support methods. Kantorovitch et al. [5] utilized annotated knowledge-base and ontology to support the initial stage of the conceptual product design. The second author of this paper developed an outsider agent [8] that extracts keywords from the ideas submitted by the creators and retrieves pieces of information that even have weak relationships with the creators' ideas. Such information from the database can be located outside of the creators' fixed ideas. Therefore, by referring to such pieces of information, it is expected that the creators would notice their fixed ideas. However, it is not always assured that the pieces of information are located outside of the creators' fixed ideas. In order to effectively support human creative thinking considering the fixed ideas, we should take into account various human cognitive phenomena [11].

This paper proposes a novel divergent thinking method named "BrainTranscending (BT)." BT is a method mainly for improving and developing existing popular products such as consumer electrical appliances like refrigerators and vacuum cleaners. The most prominent feature of BT is that it exploits BS as a method for finding "blind spots" of the creators, rather than being a method for creating new ideas. Here, the "blind spot" is a potential target that is overlooked due to some cognitive biases and thus not regarded as a target of idea creation by the creators, although it should be regarded as a target of idea creation. In other words, the blind spot is a target of idea creation that is located outside of the creators' fixed ideas.

The rest of this paper is organized as follows. Section 2 describes the BT method in detail. In addition, it presents user studies to investigate whether BT is effective, by comparing it with a baseline method, and discusses the results of those studies. Section 3 proposes an alternative BT for alleviating the creators' cognitive load, and investigates its effectiveness based on user studies. Section 4 concludes this paper.

2 BrainTranscending

2.1 BT Method

As mentioned in Sect. 1, in the ordinary BS method the creators tend to create ideas that are strongly influenced by their fixed ideas. It has generally been

assumed that such situations should be avoided, and various modified BS methods have been developed so far. However, many informative clues about the creators' fixed ideas can be found in the submitted ideas of a BS session. Therefore, based on this feature of BS, we can utilize BS as a method for extracting the fixed ideas of the creators who join a BS session, rather than using it as a method for creating new ideas. This is the most fundamental idea of this study.

We propose, in this paper, a novel divergent thinking method named "BrainTranscending (BT)." BT is a method mainly for improving and developing widely used products such as consumer electrical appliances. BT consists of the following five steps:

1. **Creating initial ideas**: Creating ideas to improve and develop a target product (e.g., a vacuum cleaner) by the ordinary BS approach. Each idea is written down on a label.
2. **Grouping ideas**: Pasting each label onto the part of the target product that the idea on the label refers to and finally grouping all labels based on the parts of the target product. For example, the ideas that relate to the flexible hose of the vacuum cleaner are gathered in one group relating to the hose.
3. **Making blind spot list**: Listing up the parts onto which no label is pasted. Such parts are called "overlooked elements."
4. **Creating further ideas**: Creating further ideas in a similar way to the ordinary BS conducted in the first step. However, in this step, the creators are required to refer to the blind spot list made in the previous step and to create ideas that improve and/or develop the overlooked elements in particular.
5. **Crystalizing final idea**: Finally, integrating all ideas obtained in the first step and the fourth step toward crystalizing a final idea.

The parts listed in the blind spot list made in the third step are parts of the target product. Therefore, the creators should be able to recognize them. However, they are not recognized as objects that should be improved and/or developed: no ideas are created for these objects. In this sense, this list can be regarded as a list of the creator's "blind spots." Thus, the BT method utilizes BS in the first step as a method to extract the blind spots, not as a method to create new ideas, although the creators may think that they are creating new ideas. In addition, the second BS session conducted in the fourth step can be regarded as a "forced relationship technique" in which the blind spot list is used as a checklist. In this sense, BT is a hybrid divergent thinking method that consists of a free-association technique (i.e., conventional BS) and a forced relationship technique.

2.2 User Study

We conducted a user study to investigate the effectiveness of the proposed "BrainTranscending (BT)" method described in Sect. 2.1. We prepared a baseline method to confirm the effectiveness of BT through a comparison. The baseline method consists of the following steps:

1. **Creating initial ideas:** The same as the first step of BT.
2. **Grouping ideas:** The same as the second step of BT.
3. **Creating further ideas:** Creating further ideas in a similar way to the conventional BS conducted in the first step. However, in this step, the creators are required to refer to "the grouping results" (not the list of blind spots) made in the previous step.
4. **Crystalizing final idea:** Finally, integrating all ideas obtained in the first and third steps and then crystalizing these into a final idea.

Briefly speaking, the baseline method is one that omits the "Making blind spot list" step from BT.

Procedure We employed six subjects in their twenties who are students of design and media development. All of them had experience with brainstorming in their course work. We asked each subject to individually improve two target products, i.e., an electric fan and a vacuum cleaner, by using the baseline method and BT. If we had conducted BT first and then the baseline method, the subjects would have inevitably come to pay attention to the originally overlooked parts, even in the baseline method, making it impossible to accurately estimate the effectiveness of BT. Therefore, to avoid this problem, we always conducted the baseline experiment before the BT experiment for all subjects.

At the beginning of the experiment, we showed the subjects the persona data shown in Fig. 1. The persona data illustrated a typical family as the users of the target products. We asked the subjects to create novel features to be used by the persona family.

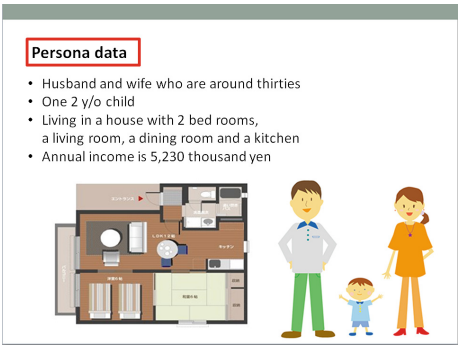


Fig. 1. Persona of a typical family as the users of the target products

We required the subjects to write down each idea on a label with the name of the part to be improved, the reason why it should be improved, and a sequence number to manage the labels sequentially. As reference materials, we provided them some pictures of the target products such as those shown in Fig. 2. The pictures include entire images of the products as well as detailed partial images of the products. In addition, we orally explained to the subjects how the products

Fig. 2. Example pictures of the target product

work and detailed functions that cannot be perceived only by looking at the still pictures.

As they were thinking of ideas, we asked the subjects to speak aloud everything that they were thinking, such as on what they were focusing at a given moment. All speech protocol data were recorded. After finishing each experiment, we interviewed the subjects about their introspection and changes in impressions of the product. In order to investigate the effectiveness of BT, we used the following data: the number of the labels and the number of the groups as quantitative data, and speech protocol data as well as interview data as qualitative data.

Table 1. Sequence and time of each step in the baseline method

Step	Time (min)
Creating initial ideas	20
Grouping ideas	10
Creating further ideas	25
Crystalizing final idea	20

Procedure of the baseline method experiment Table 1 shows the sequence and time of each step in the baseline method. In the first step, each subject individually created ideas in a conventional BS manner for 20 min, while referring

Table 2. Sequence and time of each step in BT method

Step	Time (min)
Creating initial ideas	20
Grouping ideas	10
Making blind spot list	10
Creating further ideas	25
Crystalizing final idea	20

to the sample pictures of the target product. In the second step, each subject grouped all of the labels created in the first step by him/herself for 10 min, using the sample pictures of the target product based on the relations between each idea and the corresponding part of the product. In the third step, each subject individually created ideas in a conventional BS manner for 25 min, again referring to the grouping results. Finally, in the fourth step, each subject crystalized a final idea for 20 min, while referring to the ideas created in the first step as well as in the third step. After finishing all of this work, we asked the subject to add the labels created in the third step to the grouping results obtained in the second step, and we counted the number of newly created groups.

Procedure of BT experiment Table 2 shows the sequence and time of each step in BT. In the first step, each subject individually created ideas in a conventional BS manner for 20 min, while referring to the sample pictures of the target product. In the second step, each subject grouped all of the labels created in the first step by him/herself for 10 min, using the sample pictures of the target product based on the relations between each idea and the corresponding part of the product (An example is shown in Fig. 3). In the third step, each subject listed up the parts where no ideas were created, indicated by the no-idea-parts in the pictures (An example is shown in Fig. 4) and made a blind spot list for 10 min. In the fourth step, each subject individually created ideas by a forced relationship technique for 25 min, while referring to the blind spot list. Finally, in the fifth step, each subject crystalized a final idea for 20 min, while referring to the ideas created in the first step as well as in the fourth step. After finishing all of this work, we asked the subject to add the labels created in the fourth step to the grouping

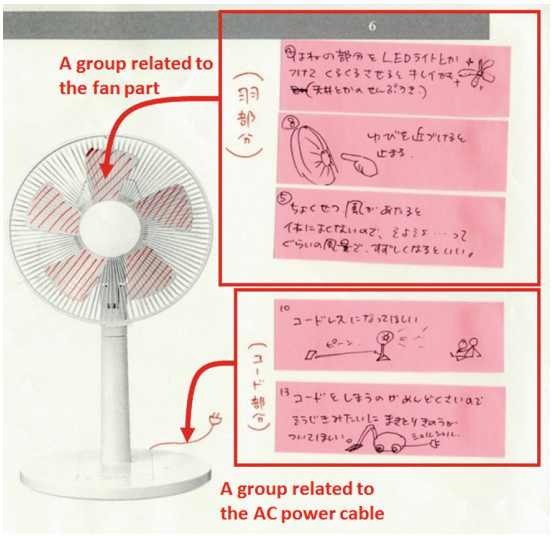


Fig. 3. Example of provided picture of target product and grouping labels related to the fan part and AC power cable part

Table 3. Results: number of labels

Subject	Total num. of labels		Num. of labels (1st step) (Num. of ideas reflected to the final idea)		Num. of labels (3rd or 4th step) (Num. of ideas reflected to the final idea)	
	Baseline	BT	Baseline	BT	Baseline	BT
a	27	22	16 (9)	11 (8)	11 (3)	11 (5)
b	28	22	19 (11)	11 (10)	9 (3)	11 (10)
c	19	20	7 (5)	8 (4)	12 (6)	12 (8)
d	41	27	20 (8)	16 (5)	21 (6)	13 (2)
e	48	52	22 (12)	28 (15)	26 (13)	24 (13)
f	19	24	11 (5)	15 (9)	8 (6)	11 (6)
Av.	30.3	27.8	15.8 (8.3)	14.8 (8.5)	14.5 (6.2)	13.7 (7.3)

results obtained in the second step, and we counted the number of newly created groups.

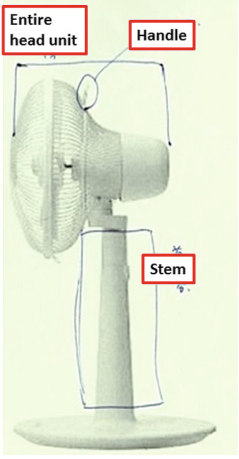


Fig. 4. Example of listing up overlooked parts

Results Table 3 shows the number of labels created in both experiments and Table 4 shows the number of groups obtained in both experiments.

Common results for both experiments All subjects had experience in BS sessions by a group but no experience in BS sessions by an individual person. Therefore, in the first step (first BS session) of both experiments, they felt uneasy about not being able to obtain any stimuli from other participants. They also felt that

Table 4. Results: Number of groups

Subject	Num. of groups (2nd step)		Num. of additional groups (After exp.)	
	Baseline	BT	Baseline	BT
a	9	6	1	4
b	7	6	2	8
c	7	8	1	12
d	9	6	3	5
e	9	9	2	5
f	8	5	2	4
Av.	8.2	6.7	1.8	6.3

it was difficult to expand and to go beyond the scope of their ideas in the first step's BS session.

Results of the baseline method experiment In the interview, the subjects answered, "Categories of ideas became evident by grouping the labels" and "I could review the created ideas by the grouping." These opinions suggest that the grouping task is effective for organizing the created ideas. However, in contrast, the grouping task is seemingly not effective for providing novel viewpoints on the subjects. As shown in Table 4, the increased number of groups after the third step (second BS session) was only 1–3 ($Av. = 1.8, STDV = 0.69$). Most of the newly added groups were ambiguous ones that were difficult to relate to specific parts in the pictures of the target object. Furthermore, self-evaluation results on the quality of the final ideas were not so good.

Results of BT experiment As shown in Table 4, the increased numbers of the groups after the fourth step was 4–12 ($Av. = 6.33, STDV = 2.87$). We compared this result with the result obtained in the baseline experiment method by Mann-Whitney test and confirmed that the difference is significant ($p < 0.01$). Some subjects answered in the interview that the ideas created in the fourth step were incorporated into the final idea crystalized in the fifth step. Self-evaluation results on the quality of the final ideas were higher than those in the baseline method experiment. Thus, the results of the BT experiment were better overall than those of the baseline method experiment. On the other hand, as for mental aspects, some subjects pointed out that it was difficult to create new ideas related to the overlooked parts in the fourth step of BT, but the difficulty was different from that of the third step in the baseline method.

Discussion Based on the results of the user study, we confirmed that the list of the blind spots is useful to support creation of further ideas. Consequently, we can conclude that the proposed "BrainTranscending" method is basically effective.

The subjects' cognitive load in the second idea-creation step for both methods was quite high. In both methods, the time spent for creating a single idea in the

second idea-creation step was longer than that in the first idea-creation step. However, the reasons for the high cognitive load were different between the methods.

For the baseline method, the reason for the high cognitive load was that participants fell into a deadlock situation. In the first step of the BS session, they extracted ideas that came easily to their minds but fell within the limits of their fixed ideas (the same situation also happened in the BT experiment). In such a case, the creators' thinking usually goes around in circles, and some assistance to get out of the deadlock is necessary. The grouping task of the labels in the second step would be useful not only for organizing the extracted ideas but also for finding overlooked points and new directions in idea creation. In this sense, the grouping task has the potential to support participants in getting out of a deadlock. However, although these subjects pointed out that the grouping was useful for organizing the extracted ideas, they did not use it for finding overlooked points and new directions. As a result, they could not get out of the deadlock in the third step, and the number of additional label groups did not increase so much. Furthermore, the ideas included in the additional groups were not used in the crystalizing task of the fourth step. Accordingly, the cognitive load in the baseline method is a simple deadlock of idea creation, and the baseline method itself could not provide an effective solution to this problem.

In contrast, with BT, the reason for the high cognitive load was that participants were forced to think about elements outside the scope of their fixed ideas. In other words, completely different from the baseline, they were forced to think beyond their fixed ideas. The blind spots listed up in the third step of BT were originally out of their limited scope of ideas. Actually, some subjects answered in the interview that they did not regard the blind spots as targets for which they should create new ideas and, moreover, that they only did so after they were directed to create ideas using the blind spots in the fourth step of BT. This is apparently the principal reason for the high cognitive load in the second idea-creation step of BT. However, the blind spot list works as a support for them to think beyond their fixed ideas. The list clearly and concretely shows the novel targets for creating ideas. The ideas related to the blind spots naturally form several new groups. Therefore, as shown in Table 4, the number of additional groups increased by much more than it did in the baseline method. Thus, the proposed BT method makes the blind spots available for use, which has not been achieved so far due to the barrier imposed by fixed ideas.

Fredrik Haren [3] pointed out that the initial ideas are very similar, even if the idea creators are different. Such initial ideas are usually stale and do not lead to any novel idea. Rikie Ishii [4] pointed out that novel ideas can be obtained only after first extracting all ideas that easily come to the creator's mind. He named this stage the "next zone," which the creators can reach when they feel they have completely extracted possible ideas. Therefore, it is important for the creators to move to the next zone to obtain novel ideas. Naturally, however, this is not easy. Some support methods are required, but conventional BS is not a good method for this purpose, as abovementioned. BT can be a good method to

effectively lead creators to the next zone, although the applicable area of BT is restricted to improving the existing products. One of the subjects answered in the interview that he could create an idea that was unexpected, even by himself; such a user evaluation supports the potential of BT.

3 Reduced BrainTranscending

As we described in the previous section, BT is an effective method for creating diverse ideas that go beyond the fixed ideas of the creator. However, the step for creating further ideas imposes a very high cognitive load. Therefore, in this section, we propose an alternative BT for alleviating this cognitive load and investigate its effectiveness through user studies.

3.1 Method

Our alternative BT, called Reduced BT (RBT), is a method that simply omits the creating further ideas step from the original BT. As mentioned in the previous section, some subjects felt that it was difficult to create new ideas in the creating further ideas step of BT. Therefore, if we can obtain similar results to BT without this step, the creators' cognitive load will lessen and time will be saved. RBT consists of the following four steps:

1. **Creating initial ideas**: The same as the first step of BT.
2. **Grouping ideas**: The same as the second step of BT.
3. **Making blind spot list**: The same as the third step of BT.
4. **Crystalizing final idea**: Integrating ideas obtained in the first step while referring to the blind spots obtained in the third step to crystalize a final idea.

3.2 User Study

In this section, we describe the RBT user study employed to investigate whether RBT can achieve similar results to BT, but with a reduced cognitive load.

Procedure We used five subjects in their twenties who are students of media informatics. All had experience in group BS sessions in their course work, but no experience in individual BS sessions. We provided several pictures of a vacuum cleaner similar to that in the BT experiment and asked each subject to individually improve it using RBT. At the beginning of the experiment, we also showed the subjects the same persona data shown in Fig. 1. We asked the subjects to create novel features to be used by the persona family.

The procedure of the user study was almost the same as that for BT. In the first step, each subject individually created ideas in a conventional BS manner for 20 min while referring to the provided sample pictures of the target product. Each subject was required to write down each idea on a label with the name of the part to be improved, the reason why it should be improved, and a sequence

number to manage the labels sequentially. In the second step, each subject spent 10 min grouping all of his/her labels created in the first step, using the sample pictures of the target product, based on the relations between each idea and the corresponding product part. In the third step, each subject was given 10 min to list up the parts where no ideas were created, indicated by the no-idea parts in the pictures, and make a blind spot list. Before starting the fourth step, we asked the subjects to be sure to create ideas for improving the overlooked elements included in their blind spot list. To do so, we allocated 45 min to this step and allowed them to freely draw sketches for creating ideas. Lastly, each subject crystalized a final idea that improved the overlooked elements. Figure 5 shows an example of a subject's crystalized idea by a subject.

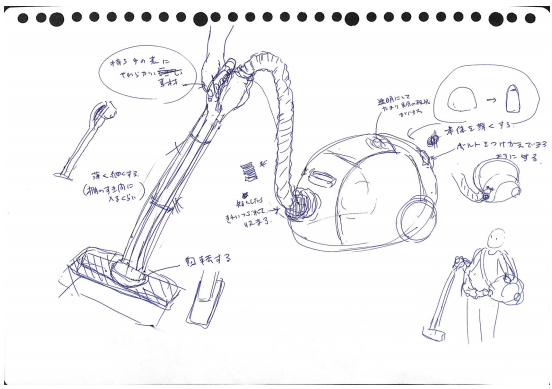


Fig. 5. Example of a subject's crystalized idea

As they were thinking of ideas, we asked the subjects to speak aloud everything that they were thinking, such as what they were focusing on at a given moment. All speech protocol data were recorded. After finishing each experiment, we interviewed the subjects on the following: (1) Total impressions of the RBT task, (2) Impressions of the fourth step, and (3) Self evaluation of the finally crystalized idea.

Results The total impressions were almost the same as those of BT. The subjects found the individual BS difficult; it was a totally new experience for them to arrange the idea labels based on their related parts, to list the overlooked elements, and to create further ideas based on these overlooked elements. The subjects also pointed out that, through these processes, they thought about things they had not been conscious of.

The fourth step of RBT, that is, crystalizing final ideas based on the blind spot list, was not easy for many of them; they were somewhat embarrassed about what they should think or do at first. Those who drew idea sketches found that these sketches were effective for glancing over all of the ideas and for creating further diverse ideas that were useful for crystalizing the final idea. Thus, in the

RBT process, most subjects represented their ideas visually, while, in the BT process, the ideas were represented verbally.

The self-evaluation results were divided into good and not good. The subjects who evaluated their final ideas as good said that they were able to come up with a novel idea because they were forced to focus on areas they previously did not consider. These subjects were able to crystalize final ideas within 25 min. In contrast, the subjects who evaluated their final idea as not good said that thinking about what they usually do not pay attention to was difficult and, hence, their final ideas were unsatisfactory. These subjects spent more than 30 min crystalizing final ideas.

Discussion It is difficult for anyone to concept new ideas to improve overlooked elements. Therefore, both BT and RBT impose the same cognitive load on the creators. However, the effectiveness of RBT depended on each creator, while BT was generally effective for almost all of the subjects, as shown in the previous section.

The BT process explicitly requires creators to generate further ideas based on the blind spot list. Therefore, no significant differences among the subjects were observed. In contrast, the RBT process does not explicitly require them to create further ideas based on the blind spot list. As a result, people who are good at creating ideas related to the overlooked elements were able to produce ideas even without any explicit instructions and eventually crystalize final ideas based on them. However, people who lacked skills for creating further ideas without explicit instructions were unable to achieve satisfactory final ideas. Thus, RBT chooses the users; it is suitable for people who are good at thinking about what is outside of their scope. We can expect that such ability can be fostered by continuously using BT. Hence, people should utilize RBT after obtaining sufficient experience with BT.

4 Concluding Remarks

This paper proposed a novel hybrid divergent-thinking method called "Brain-Transcending (BT)" that consists of a free-association technique (brainstorming) and a forced-relation technique (checklist method). BT exploits the BS session to find creator blind spots and is not used for creating new ideas. When using the forced-relation technique, the creators are compelled to create ideas related to the identified blind spots, which effectively leads the creators to the "next zone." We conducted user studies and confirmed that BS can be used as a method for identifying blind spots and that BT allows the creators to develop ideas related to their overlooked elements, which is difficult to achieve using conventional BS. Furthermore, we proposed the "Reduced BrainTranscending (RBT)" method, which omits the "creating further ideas" step of BT in order to alleviate a heavy cognitive load for the creators. User studies of RBT revealed that it is as effective as BT; however, RBT chooses the users. The users are required to accustom

themselves to "stepping outside" of their scope. Consequently, we concluded that BT and RBT are effective methods for supporting divergent thinking aimed at improving existing products. In the future, we would like to expand the applicable targets of BT to intangible objects (e.g., concepts) and software.

Acknowledgments. The authors sincerely thank the subjects for their cordial cooperation. This study was partially supported by JSPS KAKENHI Grant Number JP26280126.

References

1. Chan, J., Dang, S., Dow, S.P.: Comparing different sensemaking approaches for large-scale ideation. In: Proceedings of the CHI'16, pp. 2717–2728 (2016)
2. Gallupe, R.B., Bastianutti, L.M., Cooper, W.H.: Unblocking brainstorming. J. Appl. Psychol. **76**(1), 137–142 (1991)
3. Haren, F.: The Idea Book. Interesting.org (2006)
4. Ishii, R.: Ideamation—Everybody Creates Similar Ideas at First. Rikie Ishii's Activity Report (2006). https://www.ishiirikie.jpn.org/article/1299127.html
5. Kantorovitch, J., Niskanen, I., Zafeiropoulos, A., Liapis, A., Gonzalez, J.M.G., Motta, E.: Knowledge extraction and annotation tools to support creativity in the initial stage of product design: requirements and assessment. In: Proceedings of the 9th International Conference on Knowledge, Information and Creativity Support Systems, pp. 129–140 (2014)
6. Kim, T., Chang, A., Holland, L., Pentland, A.: Meeting mediator: enhancing group collaboration using sociometric feedback. In: Proceedings of the CSCW'08, pp. 457–466 (2008)
7. Liikkanen, L.A., Kuikkaniemi, K., Lievonen, P., Ojala, P.: Next step in electronic brainstorming: collaborative creativity with the web. CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 2029–2034 (2011)
8. Nishimoto, K., Sumi, Y., Mase, K.: Enhancement of creative aspects of a daily conversation with a topic development agent. In: Conen, W., Neumann, G. (eds.) Coordination Technology for Collaborative Applications—Organizations, Processes, and Agents. LNCS, vol. 1364, pp. 63–76. Springer (1998)
9. Noguchi, H.: An idea generation support system for industrial designers (idea sketch processor). Knowledge-Based Systems, vol. 10, Issue 1, pp. 37–42. Elsevier (1997)
10. Osborn, A.F.: Applied Imagination: Principles and Procedures of Creative Problem-solving, 3rd revised edn. Charles Scribner's Sons (1979)
11. Skulimowski, A.M.J.: Freedom of choice and creativity in multicriteria decision making. In: Theeramunkong, T., et al. (eds.) KICSS 2010, LNAI 6746, pp. 190–203. Springer (2011)
12. Sutton, R.I., Hargadon, A.: Brainstorming groups in context: effecti veness in a product design firm. Administrative Science Quarterly, vol. 41, no. 4 (Dec 1996), pp. 685–718. Sage Publications (1996)

13. van Gundy, A.B.: Brain writing for new product ideas: an alternative to brainstorming. J. Consum. Mark. **2**, 67–74 (1984)
14. Wang, H.C., Cosley, D., Fussell, S.R.: Idea expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. Proceedings of the CSCW **2010**, 103–106 (2010)

Open Data Integration for Lanna Cultural Heritage e-Museums

Watchira Buranasing(✉), Marut Buranarach, and Chai Wutiwiwatchai

National Electronics and Computer Technology Center, National Science and Technology Development Agency, Thailand Science Park, Klong Luang, Pathum Thani 12120, Thailand
{watchira.buranasing, marut.buranarach, chai.wutiwiwatchai}@nectec.or.th

Abstract. Cultural heritage is the way of life. Using digital technology to manage the information of cultural heritage has become an important issue on the perseverance of cultural heritage. Several cultural heritage institutions have their own large databases with different metadata schemas. Linked Open Data is a mechanism of publishing structured data that allows metadata to be connected. This paper presents a methodology for integrating, converting Lanna cultural heritage metadata and linking data to external data sources. Our framework uses OAI-PMH (Open Archives Initiative—Protocol for Metadata Harvesting) by converting the various metadata formats to Dublin Core metadata standard. The Dublin Core/XML was used for the integrated metadata to create a central repository of data sources. The OAI harvester was developed to extract data from the various repositories and gather data into the repository. A framework for publishing the repository data as a Linked Data source in the RDF format using the OAM Framework is described. In this framework, ontology is used as a common schema for publishing the RDF data from the database sources by mapping database schema to ontology. Finally, we demonstrate Linked Data consumption by creating a cultural heritage portal application prototype from the published RDF data sources.

Keywords: Open Data · Cultural heritage · Ontology · e-Museum
Metadata crosswalk

1 Introduction

Cultural Heritage [1, 2] is the way of life of group of people. It can be seen in ways of behaving, beliefs, values, customs, languages and traditions that are passed on from one generation to the next. Cultural heritage includes tangible culture such as buildings, monuments, inscription, manuscript, landscapes, books, works of art and artifacts, and intangible culture such as folklore, traditions, language, and knowledge and natural heritage. Using digital technology to manage the information of cultural heritage has become an important issue on the perseverance of cultural heritage. Besides written

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_3

documents, drawings and paintings, cultural heritage also includes various media such as video, photo, object virtual reality, panoramic photo and audio. Several cultural heritage institutions have their own large databases with difference metadata schemas. Some e-museum projects that focused on linking metadata from various sites include Europeana, Museum Finland, Amsterdam Museum, Smithsonian Museum, Bangladesh Museum and LODAC Museum. These projects focused on designing a common framework using ontology or other common metadata models for integrating various types of media formats, subjects, and metadata standards. The Resource Description Framework (RDF) standard [3] is often used as the metadata interchange format.

This paper proposes a new approach for interoperability between the datasets from different institutions using OAI-PMH and Linked Data framework. The methodology covers data management, crosswalk metadata, metadata harvesting, conversion and linked open data publishing. The framework was conducted over e-museum systems related to Lanna (Northern Thailand) cultural heritage. The project was conducted in collaboration with Cultural Heritage institutions in north of Thailand, including museums, temples, local agents and Cultural Information center to publish their own data to Linked Open Data. The remainder of the paper is organized as follows. Section 2 gives an overview of the methodology for integration cultural database and cultural ontology. Section 3 shows an application prototype of linked Lanna e-museums. Section 4 concludes and discusses some future directions.

2 Related Work

Linked Open Data [4] is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources. In previous research, "Linked Open Data for Cultural Heritage: Evolution of an Information Technology", Edelstein et al. [5] have made a survey on the landscape of linked open data projects in cultural heritage, examining the work of groups from around the world. Traditionally, linked open data has been ranked using the five star method proposed by Tim Berners-Lee. This research developed a six-stage life cycle based on the five-star method, describing both dataset development and dataset usage. It uses this framework to describe and evaluate fifteen linked open data projects in the realm of cultural heritage.

In [6], Constantia Kakali et al. have presented a method of ontology-based metadata integration using CIDOC/CRM ontology and a methodology for mapping Dublin Core elements to CIDOC/CRM ontology. In addition, "Supporting Linked Data Production for Cultural Heritage institutes: The Amsterdam Museum Case Study" [7] developed by Victor de Boer et al. present the methodology to convert the small cultural heritage as Amsterdam Museum metadata to a Linked Data version. "data.europeana.eu The Europeana Linked Open Data Pilot" [8] developed by Bernhard Haslhofer and Antoine Isaac produce the Linked Data version of Europeana and published the resulting datasets on the Web. "MUSEUM FINLAND-Finnish Museums

Table 1 A comparison of strategies to create linked open data

	Type	Format	Model/Ontology	Metadata	Tool	Interlinking
Amsterdam	Tangible	RDF			ClioPatria, Amalgamc	
Europeana	Tangible/Intangible	RDF	EDM	ESE		
Bangladesh	Tangible	RDF	EDM			
Finland	Tangible	RDF	CIDOC-CRM			
Smithsonian	Tangible	RDF	EDM		Karma	DBPedia, Wikipedia, Newyork time
LODAC	Tangible	RDF	CIDOC/CRM/LODAC	Dublin Cure		DBPedia
MAC-USP	Tangible	RDF	CIDOC-CRM			
NHMs	Tangible	RDF	EDM		MMAT,OAI- PMH	
Russian	Tangible		CIDOC-CRM			DBPedia, British Museum
Gothenburg	Tangible		proton, CIDOC-CRM			GeoNames, Dbpedia, CoL, Uniprol:
ThaiLanna.	Tangible/In tangible	RDF	Thailand Cultural experts	ISAD++, CDWA+, specific metadata	QAM, e-museum, OAl-PMH	Cultural Information center

on the Semantic Web" [9] developed by Eero Hyvönen et al. produce the semantic portal MUSEUM FINLAND for publishing heterogeneous museum collections on the semantic web. "Connecting the Smithsonian American Art Museum to the Linked Data Cloud" [10] developed by Pedro Szekely et al., present the process of publishing data from the Smithsonian American Art Museum (SAAM) and linking dataset to DBpedia and the Getty Vocabularies. "Linked Open Data Representation of Historical Heritage of Bangladesh" [11] developed by Shima Chakraborty et al. use semantic web technology for data management of historical heritage in Bangladesh, linking data to Geo-Bangladesh and using SPARQL to retrieve and inference specific information. "Sharing Cultural Heritage Information using Linked Open Data at a Museum of Contemporary Art" [12] developed by Erika Guetti Suca and Flávio Soares Corrêa da Silva present the architecture for sharing cultural heritage data based linked data for the Museum of Contemporary Art at the University of São Paulo (MAC-USP). Using RDF as a schema language for creating logical relationships among cultural heritage items. "Elevating Natural History Museum's Cultural Collections to the Linked Data Cloud" [13] developed by Giannis Skevakis et al. present the architecture for transition the Natural History Museum repositories (Cultural heritage content is collected from six Natural History Museums around Europe) to the semantic web and methodology for converting metadata into Linked Data. "Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing" [14] developed by Dmitry Mouromtsev et al. present the architecture and methodology for publishing open linked data from Russian Museum by using CIDOC-CRM ontology and linking data to Dbpedia and British Museum. "Reasonable View of Linked Data for Cultural Heritage" [15] developed by Mariana Damova and Dana Dannells present an application for data integration based on semantic web technologies from the Gothenburg City Museum by using PROTON and CIDOC-CRM ontology and linking data with GeoNames, DBpedia, Catalogue of Life (CoL) and Uniprot.

The above-mentioned works proposed various approaches to create linked open data from data-base with the same metadata schema. However, they can be complementary to each other. As another research work, Matsumura et al. [16] collected information of museums from their websites and generated Linked Open Data of museum information in Japan. Table 1 compares different approaches of existing e-museum projects based on the LOD framework.

3 Our Framework for Integrating Lanna Cultural Datas for Linked Open Data Publishing

3.1 System Architecture Overview

One of the challenges of integrating cultural heritage data is system and metadata heterogeneity. Every cultural heritage institution has its own collection management system such as e-museum management system, manuscript management system, mural management system. Furthermore, they usually occupy their own metadata schedule.

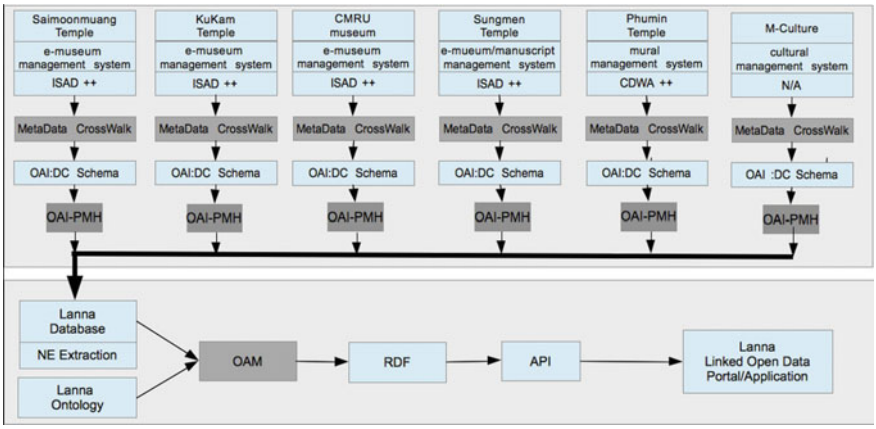


Fig. 1 Architecture and workflow for harvesting and converting to linked open data

Figure 1 shows our system architecture for integrating various e-museum systems of the Lanna culture. Our architecture is primarily based on OAI-PH and ontology frameworks. The interoperability and exchange of metadata is further facilitated by metadata crosswalks. This methodology crosswalks metadata from ISAD++, CDWA++ and specific metadata (cultural information center) dataset to Dublin Core metadata schema and export to a simple dublin core xml format. (OAI: DC schema). Using OAI-PMH to harvest data from xml. To extract information from the result and insert into the database, we map the extracted information with the cultural heritage ontology under the ontology-based application management (OAM) framework [17], then export them in the RDF format, and finally create a cultural heritage portal application by using such RDF data sources.

3.2 Lanna Cultural Data Management Systems

This section describes existing systems utilized in various Lanna cultural heritage sites. These systems are e-Museum, Manuscript and Mural management systems.

3.2.1 e-Museum Management System

Our e-Museum Management System is designed to aid users to archive and to catalog cultural heritage items in museums, temples, and cultural agents for digital exhibition and digital asset management. e-Museum Management System is based on ISAD (G) metadata standard and added some elements from cultural experts. This application allows for the upload, description, management and access of digital collections and also converts and exports OAI-DC xml format.

3.2.2 Manuscript Management System

Manuscript Management System is a web-based software application for the description, management, and dissemination of manuscript collections information and translation from local language to native language and also translation from native

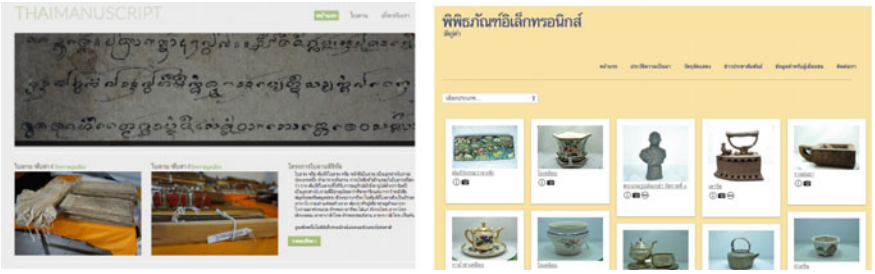


Fig. 2 e-Museum management system and manuscript management system

language to an interpretive approach. This application is based on ISAD (G) metadata standard and added some elements from manuscript expert interpreter. This application allows for the upload, description, management and access of manuscript digital collections (Fig. 2).

3.2.3 Mural Management System

Mural Management System is an application to administrate wall painting (mural) photo galleries. This application is based on CDWA metadata standard and added some elements from cultural expert. This application allows for the upload, description, management and access of manuscript digital collections.

3.2.4 Thailand Cultural Information center

Thailand Cultural Information center (http://www.m-culture.in.th) is a national cultural archive, which is one of an important database for education, economy and society. This archive was developed under the collaboration between Ministry of Culture and National Electronics and Computer Center. The content database is associated with person, organization, place and artifact. This archive is based on special metadata elements from cultural experts from Ministry of Culture and the content is approved by cultural agents from 76 provinces in Thailand (Fig. 3).

Fig. 3 Mural management system and Thailand cultural information center

3.3 Metadata Crosswalk and Open Archives Initiative Protocol for Metadata Harvesting

3.3.1 Metadata Crosswalk

A metadata crosswalk is a specification for mapping one metadata standard to another. A complete specified crosswalk consists of both a semantic mapping and a metadata conversion specification and all implementations of the crosswalk on a specific source content result in the same target content. Each archive has its own database with difference metadata schemas and the Dublin Core metadata schemes is a standard for cross-domain information resource description. The approach is to convert its own metadata schemas to Dublin core metadata schemas.

The following steps are mapping instructions for a *complete specified* crosswalk.

Step 1: extract terminology and properties. A lack of common terminology currently exists among the different metadata standards. Some of terminology has the same or nearly the same meaning. For example, ISAD metadata is identified using <identifier>, whereas Dublin Core metadata is identified using <reference code>. Besides, some of the metadata standards use similar properties in the definition of their metadata. The similarities need to be extracted and the concepts generalized and used in a common way across all metadata standards.

Step 2: element to element mapping including one-to-one and one-to-many transformation. Each metadata standard was built for a different propose and some elements were built from expert, so some of elements could match in Dublin Core, but there are some elements could not match in the other (Dublin Core) metadata standard. The importance of mapping method is extra elements to element mapping. For example, manuscript has <Pariwat> element (translate from the old northern *Thai language* to native Thai language), the solution to handle this problem is take <Pariwat> element maps to <description> element in Dublin Core metadata standard.

Step 3: content conversion and combination. Each metadata standard restricts the content format of each metadata element such as data type or range of values. It is important to convert between text and numeric values or text and date values. Some specific elements was built by cultural experts for example, Chula Sakarat (Minor Era) and Rattanakosin era is *necessary to convert to date format, Besides, some of source element values is free text, that must be converted to same format such as Anno Domini (A.D.) is necessary to convert to MM-DD-YY in Buddhist Era (B.E.).* A general content conversion could map element to element with one value, but some element must be combinations of value for example, values of <source> elements could be a source of location and URL (Watkukam: www.emusuem.in.th/watkukam).

After the crosswalk change, the result of conversion is complete metadata elements and values, then export to OAI-DC XML file format for sharing. Table 2 shows the mapping of existing metadata schemes of the Lanna cultural systems to OAI-DC.

Table 2 Example of metadata crosswalk

	OAI-DC	Antique	Manuscript	Mural	Cultural Information center
	DC	ISAD++	1SAD++	CDWA++	N/A
1	Identifier	Reference code(s)	Reference code(s)	Unique id	id
2	Title	Title	Title	Title	Title
3	Date	Date	Date	Creation period	
4	Format	Extent	Extent	Work type	
				Classification	
5	Creator	Name of creator(s)	Name of creator(s)	Creator	
6	Description	Administrative/biographical history	Administrative/biographical history	Description	Description
		Archive history	Archive history	Shape	
		Finding aids	Finding aids	Dimension	
		Note	Note	Material	
		Archivist'sNote	Archivist'sNote	Technique	
		Date(s) of descriptions	Date(s) of descriptions	Culture	
		System of arrangement	System of arrangement	Inspiration	
		Appraisal, destruction and scheduling information	Appraisal, destruction and scheduling information	Inspiration	
			Burana phasa	Detail	
			Pariwat		

3.3.2 Open Archives Initiative Protocol for Metadata Harvesting

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a protocol developed by the Open Archives Initiative. It is used to harvest the metadata descriptions of the records in an archive so that services can be built using metadata from many archives. An implementation of OAI-PMH representations metadata in Dublin Core. Based on OAI-PMH using Metadata Integration application applied from the PKP Metadata Harvester for centralized repository. Metadata can be harvested at any time, and as frequently as required. After the centralized repository is completed, then extract data from the result of repository using tag extraction. The method to extract the data find Dublin Core element tag patterns from the xml and insert into the MySQL Database. Figure 4 shows an example of the exported XML data of OAI-DC for OAI-PMH harvesting.

```
<oai:metadata>
  <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>ตาลปัตร</dc:title>
    <dc:creator></dc:creator>
    <dc:subject></dc:subject>
    <dc:description>
      ตาลปัตร ใหยธรรมที่ระลึกถวายพาน ที่ทานบดีเจ้าภาพถวายไว้ณ วัดภู่ค่า ไม่ปรากฏปีที่สร้าง เป็นพัด
      หน้านาง พิมพ์ลายสลับสี เหลือง เขียว เป็นลายกรอบ ใจกลางเป็นรูปเสือ คงเป็นนักษัตรประจำปีเกิด
      ของทานบดีเอง ด้ามเป็นไม้กลึง
    </dc:description>
    <dc:description></dc:description>
    <dc:description></dc:description>
    <dc:publisher></dc:publisher>
    <dc:contributor></dc:contributor>
    <dc:date></dc:date>
    <dc:type></dc:type>
    <dc:format></dc:format>
    <dc:identifier>
      พิพิธภัณฑ์วัดภู่ค่า : www.emuseum.in.th/watkukan : museum :
      http://emuseum.in.th/watkukan/emuseum/projects/objects/indexzoom.php?
      objref=MA-M-2557-0018
    </dc:identifier>
    <dc:source></dc:source>
    <dc:language></dc:language>
    <dc:relation></dc:relation>
    <dc:relation></dc:relation>
    <dc:relation>
      www.emuseum.in.th/watkukan/object/pic/thumb/thumb-DSC_1519.jpg
    </dc:relation>
    <dc:coverage></dc:coverage>
    <dc:rights></dc:rights>
  </oai_dc:dc>
</oai:metadata>
```


Fig. 4 Example of exported XML data of OAI-DC for OAI-PMH harvesting

3.4 Publishing Lanna Cultural Data as Linked Open Data

Although OAI-DC provides a common metadata representation using the Dublin Core standard, it has limited expressivity. Specifically, the DC elements are general and can not easily represent domain-specific metadata, e.g. paintings, museum objects, traditional archive, etc. Thus, domain-specific metadata representation is needed in addition to DC. In addition, Linked Open Data (LOD) is now accepted as an effective mechanism of data sharing to enable data integration based on RDF standard. Thus, we design a LOD data publishing framework that extends OAI-DC. In this framework,

data from DC elements can be further extracted for key concepts and relations. The extracted results can be mapped with the Lanna Cultural Heritage ontology and published as RDF data as a LOD data source. The ontology design approach and prototype system that utilizes the RDF data are described as follows.

3.4.1 Cultural Heritage Ontology Design and Mapping

Cultural heritage involves knowledge coherence and knowledge variety. Thus, it can be difficult to classify some knowledge into domains and design a complete cultural heritage ontology. Some of knowledge such as NOK-Hat-Sa-Dee-Ling (a legendary bird of Himavanta) can be classified into three domains: Antiques, Tradition and Ritual Instruments and tool. Therefore, it is important to limit the scope of the construction of domain ontology. Our initial ontology design focuses on three domains: antiques, manuscript and mural. The ontology model of cultural heritage was designed and evaluated with the involvement of cultural experts from Ministry of Culture, Thailand. The ontology was created in the OWL format (Web Ontology Language) using Hozo ontology editor. Figure 5 shows an initial design of Lanna Cultural Heritage ontology.

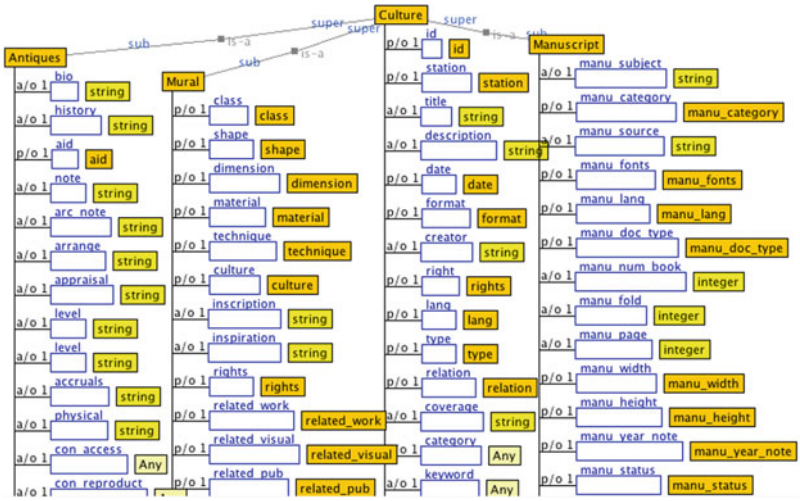


Fig. 5 Cultural heritage ontology (antiques, manuscript and mural)

The RDF data publishing used the Ontology Application Management (OAM) framework [18] for schema mapping and vocabulary mapping between OWL format and a database source. OAM facilitated defining mapping between ontology classes and database tables by class-table mapping and property-column mapping, then exports to RDF format.

4 Lanna Cultural Heritage Portal Application using RDF Data Sources

After the data sources are transformed to RDF data based on the schema defined in the ontology, the OAM framework provides the RESTFul API that allows applications to utilize the RDF data sources in a uniform fashion. The "Lanna Cultural Heritage" portal is a web portal that provides a semantic view-based search engine where the knowledge base consisting of ontologies and data. The portal combines data from several data sources of museums and temples in the northern part of Thailand, for example, cultural objects from WatKukam, manuscripts and documents from Watsungmen, murals from Watphumin and cultural information center (www.m-culture.in.th) and so on (Fig. 6).

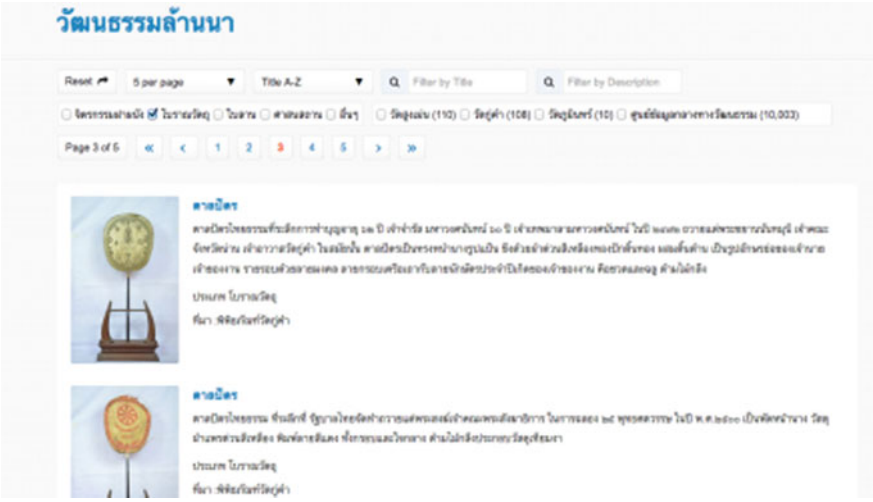


Fig. 6 Cultural heritage portal application using RDF data sources

5 Conclusion

This paper presents a methodology for integrating, converting Lanna cultural heritage metadata and sharing data to external data sources as linked open data. Case studies for this methodology are museum archives in north of Thailand and cultural information center from the Ministry of Culture. Our methodology is designed to convert different metadata formats to Dublin Core metadata standard as OAI-DC XML format, and to incorporate metadata to form a central repository of data sources. The extraction of key information from the result of repository is made by tag extraction and mapping of database schema using ontology under the Ontology-based Application Management (OAM) framework. The result is formulated in the form of RDF format. The data were

integrated to create cultural heritage portal application by using RDF data sources. Our directions for future work include applying the methodology to extract key concepts from the textual data and more complex ontology design for the Lanna cultural heritage domains.

References

1. Wikipedia: Culture (2015). https://simple.wikipedia.org/wiki/Culture
2. Culture. https://www.tamu.edu/faculty/choudhury/culture.html
3. Selçuk Candan, K., Liu, H., Suvarna, R.: Resource description framework: metadata and its applications. SIGKDD Explor. **3**(1), 6 (2001)
4. The Europeana Foundation: Europeana Linked Open Data Culture. http://labs.europeana.eu
5. Edelstein, J., Galla, L., Li-Madeo, C., Marden, J., Rhonemus, A., Whysel, N. : Linked open data for cultural heritage: evolution of an information technology. In: The 31st ACM International Conference on Design of Communication (2013)
6. Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., Gergatsoulis, M.: Integrating Dublin Core metadata for cultural heritage collections using ontologies. In: DCMI International Conference on Dublin Core and Metadata Applications, 27–31 August 2007
7. de Boer, V., Wielemaker, J., van Gent, J., Hildebrand, M., Isaac, A., van Ossenbruggen, J., Schreiber, G.: Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study, ESWC, vol. 7295 of Lecture Notes in Computer Science, pp 733–747 (2012)
8. Haslhofer, B., Isaac, A.: data.europeana.eu The Europeana Linked Open Data Pilot. In: Proceedings of the International Conference on Dublin Core and Metadata Applications (2011)
9. Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: Museum Finland—Finnish Museums on the Semantic Web, Web Semantics: Science, Services and Agents on the World Wide Web, vol. 3, issue 2–3, October, pp. 224–241 (2005)
10. Szekely, P., Knoblock, C.A., Yang, F., Zhu, X., Fink, E.E., Allen, R., Goodl, G.: Connecting the Smithsonian American Art Museum to the Linked Data Cloud. In: The 10th Extended Semantic Web Conference (ESWC) (2013)
11. Chakraborty, S., Rahman, Md.H.H., Seddiqui, Md.H.: Linked Open Data Representation of Historical Heritage of Bangladesh, Computer and Information Technology (ICCIT), 8–10 March 2014, pp. 242–248 (2014)
12. Suca, E.G., da Silva, F.S.C.: Sharing Cultural Heritage Information Using Linked Open Data at a Museum of Contemporary Art, CIDOC, 6–11 September 2014 (2014)
13. Skevakis, G., Makris, K., Arapi, P., Christodoulakis, S.: Elevating natural history museum's cultural collections to the linked data cloud. In: 3rd International Workshop on Semantic Digital Archives (2013)
14. Mouromtsev, D., Haase, P., Cherny, E., Pavlov, D., Andreev, A., Spiridonova, A.: Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing, The Semantic Web Latest Advances and New Domains, 21 May 2015, pp. 637–651 (2015)
15. Damova, M., Dannells, D.: Reasonable view of linked data for cultural heritage. In: Third International Conference on Software, Services and Semantic Technologies (2011)

16. Matsumura, F., Kamura, T., Kato, F., Kobayashi, I., Takahashi, T., Ueda, H., Ohmukai, I., Takeda, H.: LODAC Museum: integration and utilization of linked open data in museums. In: The 26th Annual Conference of the Japanese Society for Artificial Intelligence (2012)
17. Buranarach, M., Thein, Y.M., Supnithi, T.: A community-driven approach to development of an ontology-based application management framework. In: Semantic Technology Lecture Notes in Computer Science, vol. 7774, pp. 306–312 (2013)
18. The ontology-based application management (OAM) framework. http://lst.nectec.or.th/oam/

RDF Dataset Management Framework for Data.go.th

Pattama Krataithong[1,2(✉)], Marut Buranarach[1], and Thepchai Supnithi[1]

[1] Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center (NECTEC), Pathum Thani, Thailand
{pattama.kra, marut.bur, thepchai.sup}@nectec.or.th
[2] Faculty of Science and Technology, Department of Computer Science, Thammasat University, Pathum Thani, Thailand

Abstract. Recently, Thailand has started initiating the Thailand open government data project that continuously triggers an increment in the number of open datasets. Open data is valuable when the data is reused, shared and integrated. Converting the existing datasets to the RDF format can increase the values of these datasets. In this paper, we present the architecture and processes for RDF dataset management for Data.go.th based on OAM Framework which supports the entire processes: RDF data publishing, and data querying. Our approach is different from other LOD platforms in that users do not require the knowledge of RDF and SPARQL. Our platform would facilitate data publishing and querying process for novice users and make it easier to use. This framework provides a common ontology-based search interface and RESTFul APIs constructed automatically for the datasets when they are published. With the provided services for the datasets, it can simplify the user's tasks in publishing datasets and create applications for the datasets. In consuming the RDF data, we implemented a sample mash-up application which accessed the published weather and reservoir datasets from the Data.go.th website via RESTful APIs.

Keywords: Dataset management process · Open government data
Open data platform

1 Introduction

Open Government Data initiatives are widely adopted in many countries in order to increase the transparency, public engagement and the channel of communication between government and their citizens. Each country has a different policy or strategy to publish their datasets. There are three main open data strategies including democratic representation, service provision and law enforcement [1]. For instance, the United States focuses on transparency strategy to increase democratic accountability and efficiency in government. The United Kingdom has promoted publishing open government data (OGD) datasets by law enforcement. UK government has a clear policy to make datasets in machine-readable formats and published under an open license [2, 3].

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_4

As of July 2015, there are over 150,000 datasets on Data.gov and over 25,000 datasets on Data.gov.uk. Although there are lots of datasets in several formats which are published on the web, only small proportion of these datasets are actually used [4]. One of the reasons is that most of datasets are published in various formats (e.g. XLS, CSV, HTML, PDF). Thus, it can be difficult for application developers to process varieties of file formats and spend more time to manage the updates of these datasets.

In Thailand, the government has promoted the initiative by issuing policies and regulations related to open data publishing and try to educate government sectors to understand about open government data. Data.go.th is the national open government data portal of Thailand, which started to publish several datasets in 2014. As of July 2015, there are several hundred datasets available on this portal in various formats (e.g. XLS, CSV, PDF) and users can download raw datasets from this web site. Data.go.th also promotes applications and innovations that are created based on the content published on the portal.

Based on the 5-star open data model [4], Resource Description Framework (RDF) is a standard data format that can support Linked open data. Consuming RDF data is usually achieved by querying via an SPARQL endpoint. SPARQL is a query language for RDF. RDF data is typically stored as linked structure as a "graph data" and represented in URI. Therefore, a developer who want to use the SPARQL endpoint must have the knowledge about SPARQL and RDF. Our work proposes that Web API is an easier way for retrieving RDF-based open data. There are several advantages including:

1. Data as a service—developers who do not have background in RDF and SPARQL can query a dataset via a RESTFul API service.
2. Standard data format—developers do not need to study a new data format, the query results will be returned in the standard JSON format.

In this paper, we develop a framework and a set of tools for data management based on RDF. The framework focuses on transforming the existing datasets into the RDF format and providing convenient way to access the published RDF datasets. Specifically, users who want to access the published datasets do not require knowledge about the underlying data formats or query language, i.e., RDF and SPARQL query. They can access the data in simple-to-use format via RESTFul APIs. We propose an architecture and processes for transforming and publishing RDF datasets from some existing datasets on Data.go.th and using the published data in applications. The framework is extended from the Ontology Application Management (OAM) framework [5]. The framework was used to support the entire process including RDF data publishing and consumption. To demonstrate an actual use case, we developed a demonstration application that mashed up two datasets from Data.go.th via the provided RESTFul APIs. Our approach is different from other LOD platforms in that users do not require the knowledge of RDF and SPARQL. This would facilitate the data publishing and querying process for typical users that can make it easier to use.

This paper is structured as follows. Section 2 introduces the overview about data catalogs and dataset management portal for open data and background of OAM Framework. Section 3 describes the architecture, design of the APIs and processes for converting and publishing the datasets. Section 4 describes a sample application to

demonstrate use of the published RDF datasets. Section 5 provides a summary and discussion of some future directions.

2 Background

2.1 Data Catalogs and Dataset Management Systems for Linked Open Data

Public sector data and information are mostly scattered across many different websites and portals and are often collected in various formats. Many countries create online data catalogs as the central data repository where datasets can be found and accessed [6]. CKAN is the open data catalogs which allow data owners to upload the collections of data and provide information about the data. There are many countries using CKAN for storing open government datasets, such as the US's data.gov, the UK's data.gov.uk, and the European Union's publicdata.eu.

The varieties of file formats of raw datasets are often inconvenient for some users or developers to use. Specifically, using proprietary formats can reduce the dataset accessibility. Based on the 5-star open data principle [4], improving quality of datasets for open data includes data conversion to an open standard format like RDF that provides better support for data sharing, reuse, integration and data querying. Therefore, RDF dataset management system should be provided to support RDF data publishing and provide the data querying service to users who want to access the published data. Unlike open data catalogs, the features of dataset management system for RDF-based open data should cover the processes for data conversion, data publishing and data querying.

There are several dataset management systems for Linked open data. Each system provides different tools for data conversion/creation, data publishing and data query/access. TWC LOD [2] is a portal for linked open government data supporting data conversion, publishing, and querying. They use open source tool named Csv2rdf4lod[1] to convert CSV files into Linked Open Government Data (LOGD) RDF files and then use OpenLink Virtuoso[2] for loading all metadata and converted files into their SPARQL endpoint. DaPaaS Platform [7] aims to provide a Data-and Platform-as-a-Service environment that support data importing/transforming, data publishing, and data querying. They use Linked Data manufacturing tool named Grafter[3] for converting tabular data to RDF format and publish the converted data on DaPaaS. Users can query the published data via SPARQL endpoint [8]. St. Petersburg's LOD [9] is a project for publishing Russian open government data. In this project, the Information Workbench tool[4] is developed to support data storing, dataset converting to RDF format, providing SPARQL endpoint to access the data, and visualizing data in various formats (charts, tables, graphs, etc.). LGD Publishing

[1] Csv2rdf4lod (http://data-gov.tw.rpi.edu/wiki/Csv2rdf4lod).

[2] OpenLink Virtuoso (http://virtuoso.openlinksw.com/).

[3] Grafter (http://grafter.org/).

[4] Information Workbench (http://www.fluidops.com/en/portfolio/information_workbench/).

Pipeline [10] is not a dataset management portal. The system provides the self-service approach to create Linked Government Data (LGD). This approach provides a publishing pipeline for users to convert a raw government data to LGD by themselves. The pipeline was based on the Google Refine (currently named "OpenRefine[5]").

Most LOD platforms focused on converting raw datasets to RDF to provide support for linked open data and to provide a data querying service via SPARQL endpoint. Thus, publishing and using linked open data normally requires knowledge about RDF and SPARQL query. Our approach is different from other LOD platforms in that the users do not require the knowledge of RDF and SPARQL. This would facilitate the data publishing and querying process for novice users that can make it easier to use. To simplify the task of consuming the RDF data, we design and develop the RESTFul APIs as a data querying service on top of SPARQL endpoint. Our RDF dataset management system is built based on the OAM Framework, which is a platform that simplifies development of ontology-based semantic web application [10]. We believe that the process simplification can help to promote more publication and usage of Linked open data.

2.2 Ontology Application Management (OAM) Framework

The Ontology Application Management (OAM) framework is a java-based web-application development platform which helps users to build a semantic web application with less programming skill required. The underlying technology of OAM is Apache Jena, D2RQ and RDF data storage [5]. OAM includes three main modules follows as:

- Database-to-Ontology Mapping provides a user interface for mapping between an existing relational database schema and ontology file (OWL). This process helps users who not have a programming skill in mapping and converting relational database to RDF format.
- Ontology-based Search Engine provides a Form-based SPARQL data querying service for users to query each dataset by defining search conditions.
- Recommendation Rule System provides a simplified interface for rule management. Users can define a condition of rules that do not require knowledge of the rule syntax of reasoning engine.

In developing an RDF-based platform for the Thailand open data project, we adopted two primary features of OAM. First, the RDF data publishing function of OAM was extended to support multiple dataset services. The multi-dataset support is necessary for building a scalable open data portal. Second, we extended the ontology-based search engine by implementing RESTFul APIs as a dataset querying service on top of SPARQL endpoint. The design of the APIs includes the basic SPARQL querying operations such as filter, aggregate, order by, limit and offset. Figure 1 shows an example of using OAM for database-ontology mapping and publishing an RDF dataset from an existing weather statistics dataset.

[5] OpenRefine (http://openrefine.org/).

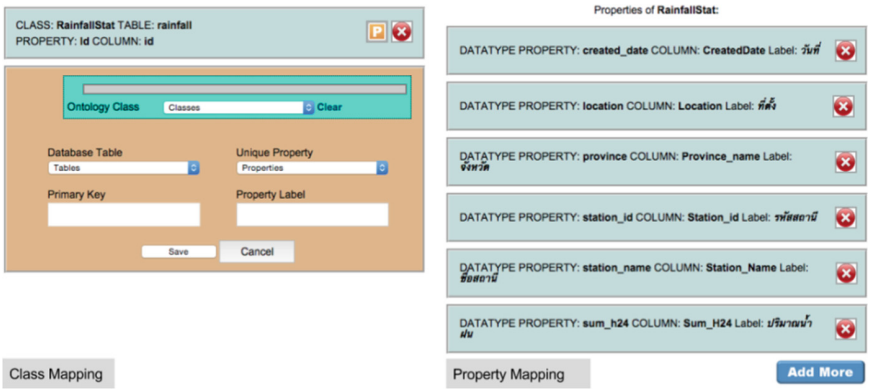


Fig. 1. An example of using OAM for database-ontology mapping and publishing an RDF dataset from an existing weather statistics dataset

3 RDF Dataset Management Processes for Data.go.th

Management of RDF Dataset is done after the raw datasets are published on the Data.go.th website. Specifically, data publishing workflow have two processes: raw dataset publishing on Data.go.th portal and RDF dataset publishing on a separate portal (http://demo-api.data.go.th). The data owners can publish their datasets on the data portal Data.go.th in various file formats. Basic functionality of this portal is to provide the interface for configuring metadata, uploading, and searching the datasets. Afterwards, the data owner has an option to import some selected datasets in tabular format, e.g. CSV or XLS, for RDF data transformation and publishing on the RDF dataset portal. The RDF dataset publishing has to be done separately because only well-structured datasets can be successfully transformed to RDF data. Thus, the data owner must be trained to prepare the datasets in proper forms, e.g. putting column names in the top rows, putting data in successive rows. Figure 2 shows the dataset publishing workflow.

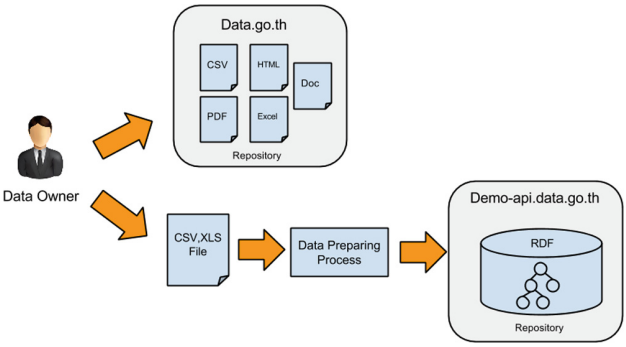


Fig. 2. Dataset publishing workflow

3.1 RDF Dataset Management Processes

RDF dataset management on the Demo-api.data.go.th portal has two main processes: publishing and querying, as shown in Fig. 3.

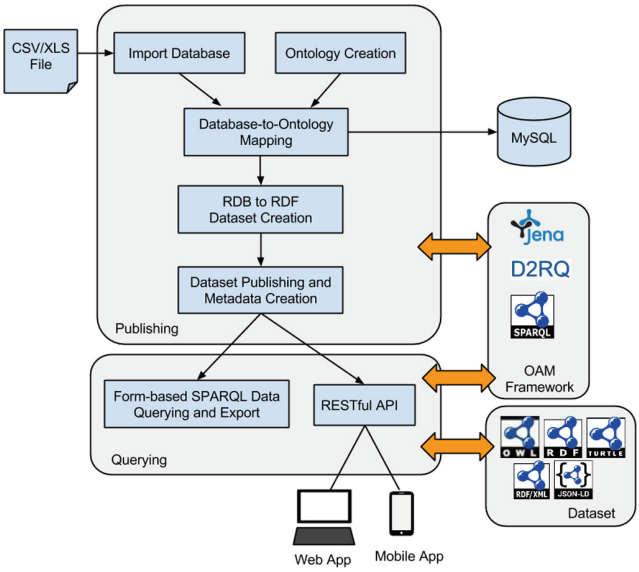


Fig. 3. Overview of RDF dataset management processes

(1) **Publishing process**. The steps for preparing and converting data to the RDF format are listed below:

- *Data preparing*: Users need to prepare database and OWL ontology. First, user can fetch a raw dataset in tabular format from Data.go.th and import the tabular files to relational database. Second, creating OWL ontology which defines the dataset schema.
- *Data mapping*: Using OAM framework, users can define database schema-to-ontology mapping. OAM provides the interface for schema mapping and vocabulary mapping between ontology file (OWL) and a relational database schema. After the mapping process, user can convert a dataset to RDF through interface. The converted dataset is ready to publish.
- *Dataset Publishing*: Once the RDF files were successfully created. The RDF files will be published on the Demo-api.data.go.th portal.

(2) **Querying process**. This steps for publishing and querying the RDF datasets are listed below:

- *Form-based SPARQL data querying*: Ontology-based search engine of the OAM framework provides the form-based SPARQL data querying, which hides the syntax of SPARQL query. Users can choose any RDF dataset for data

querying. Search interface for each dataset is formed based on ontology of the dataset. The search results are presented in table form and users can also export the results in JSON and CSV format.

- *RESTFul APIs*: For each dataset, the data querying service is automatically provided as RESTFul API by means of the OAM framework. Application developers can query each dataset via APIs and the returned search results are provided in JSON format.

3.2 Design of Dataset APIs

Demo-api.data.go.th provides the data as a service through RESTFul APIs for every dataset which was converted in RDF format and published on this portal. Figure 4 shows the dataset provided as a service through RESTFul API.

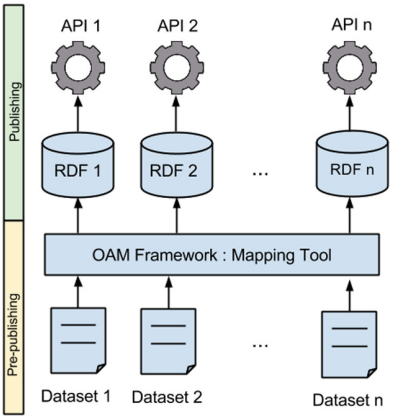


Fig. 4. Layers of Dataset Service API at Demo-api.data.go.th

In querying data, we design RESTFul APIs to cover most of the basic operations of SPARQL query such as filter, aggregation functions, order by, limit and offset. All datasets have the same patterns of querying APIs. Thus, it can help users to reduce a learning curve. In addition, the ontology-based search interface also provides a direct access to access the querying API as an alternative for writing the APIs directly. They can copy the API querying URL to use in their applications as shown in Fig. 5.

The APIs are provided for three main functions: getting all dataset names, getting description of the dataset and data querying by condition based on ontology. Examples of APIs of each function are provided as follows:

- Get all dataset name:
  http://demo-api.data.go.th/searching/api/dataset/get_name
- Get description of the dataset such as concept, property, and operators:
  http://demo-api.data.go.th/searching/api/dataset/get_info?dsname={datasetname}

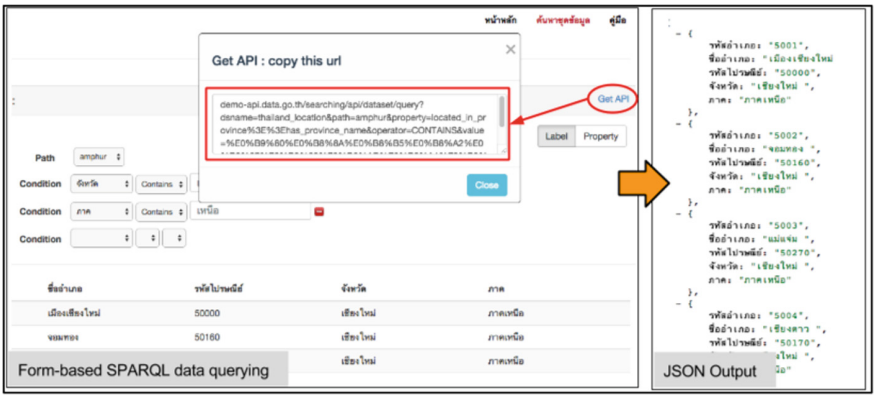


Fig. 5. Access to the API via the ontology-based search interface

- Querying by conditions:
  http://demo-api.data.go.th/searching/api/dataset/query?dsname={datasetname}&path={concept}&property={property}&operator={operator}&value={keywordorconcept}&limit=100&offset=0

The parameters in querying API are based on the structure of the ontology of each dataset. There are two types of property with the supported querying operators as follows:

- Datatype property: This type relates things to literal (e.g., String, Integer, Float etc.). Operators of literal value for querying API include 'CONTAINS' for string value and =, >, <, >=, <= for number value comparison.
- Object property: This type relates individuals to individuals which refer to some particular class. Operator of querying API for object property is 'IS-A' operator, which supports 'Subclass-of' inferencing.

3.3 SPARQL Query Translation and Result Structure of APIs

In this section, we provide details and examples of SPARQL query translation for the dataset APIs. SPARQL queries and the results of each API function are described as follows.

1. Get all dataset names: This API returns the list of dataset names; both short names and full names of datasets. The short name of a dataset is used to reference the dataset in other APIs.
   - URL: http://demo-api.data.go.th/searching/api/dataset/get_name
   - Result: list of dataset names (Fig. 6)
2. Get description of the dataset: This API returns description of the dataset including concept, property, and operators as shown in Fig. 7.

```
{
  expense_stat: "ค่าใช้จ่ายเฉลี่ยต่อเดือนต่อครัวเรือน จำแนกตามภาค และจังหวัด พ.ศ. 2545 - 2556",
  tambon: "ข้อมูลพิกัด LAT/LONG ที่ตั้งตำบล",
  income_stat: "รายได้เฉลี่ยต่อเดือนต่อครัวเรือน จำแนกตามภาค และจังหวัด พ.ศ. 2539 - 2556",
  asean_law: "ชุดข้อมูลกฎหมายของประเทศในกลุ่มอาเซียน",
  dam_stat: "ชุดข้อมูลสภาพน้ำในอ่างเก็บน้ำขนาดใหญ่ทั่วประเทศย้อนหลัง ๓ ปี พ.ศ. 2554 - 2557",
  city_bus: "ชุดข้อมูลสายรถเมล์และป้ายรถเมล์",
```


Fig. 6. Example JSON result for list of dataset names

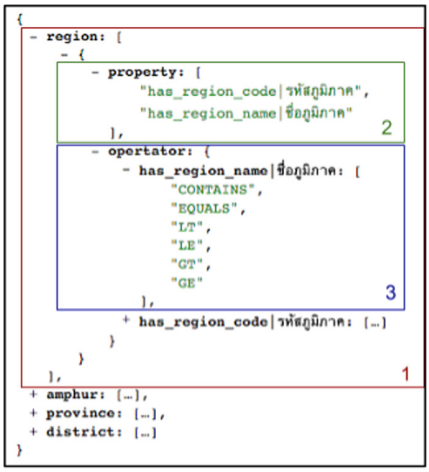


Fig. 7. Example of JSON result for description of the dataset

- URL: http://demo-api.data.go.th/searching/api/dataset/info?dsname=thailand_location
- Result: there are three parts in the description of the dataset as follows
  (1) Concept: the scope of concept or class of ontology.
  (2) Property: all properties of the concept that include list of property names and labels.
  (3) Operators: operators of literal value for querying API include 'CONTAINS' for string value and 'EQUALS' (=), 'GT' (>), 'LT' (<), 'GE' (>=), 'LE' (<=) for number value comparison. Operator of querying API for object property is 'IS-A' operator.

3. Querying by conditions: This API allows users to submit the parameters for querying by conditions. Each query condition consists of property, operator and value. The property and operator are selected from those returned by the API to obtain description of the dataset.

For example, a sample query to get the names of all districts located in the northern region ("ภาคเหนือ") that contain the string "Chiang" ("เชียง") in province names from the "Thailand Location" dataset can be formed as (Fig. 8):

```
http://demo-api.data.go.th/searching/api/dataset/query?dsname=thailand_location&
path=amphur&property=located_in_province>>located_in_region>>has_region_name&o
perator=CONTAINS&value=ภาคเหนือ&property=located_in_province>>has_province_
name&operator=CONTAINS&value=เชียง&limit=100&offset=0
```

Fig. 8. Example API for querying a dataset by search conditions

- Pattern of SPARQL query: in the above example, the API for querying by search condition using "AND" condition to supports more than one condition. There are two steps for query translation including getting IDs for all subjects and then using these IDs to query property value data as shown in Figs. 9 and 10.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?x ?id
WHERE{
?x <http://data.go.th/thai_location#located_in_province> ?a01 .
?a01 <http://data.go.th/thai_location#located_in_region> ?a02 .
?a02 <http://data.go.th/thai_location#has_region_name> ?a03 .
FILTER (regex(?a03, 'ภาคเหนือ' , 'i' ))
?x <http://data.go.th/thai_location#located_in_province> ?a11 .
?a11 <http://data.go.th/thai_location#has_province_name> ?a12 .
FILTER (regex(?a12, 'เชียง' , 'i' ))
?x <http://data.go.th/thai_location#has_amphur_code> ?id .
?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.go.
th/thai_location#amphur> .
}ORDER BY ASC(xsd:integer(?id)) LIMIT 100 OFFSET 0
```

Condition1

Condition2

Fig. 9. Example of SPARQL query for getting IDs of all results

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?x ?has_amphur_name ?has_postal_code ?located_in_province4 ?
located_in_province6
WHERE {
?x <http://data.go.th/thai_location#has_amphur_code> ?pkey .
FILTER( ?pkey = '5001' || ?pkey = 5001 || ?pkey = '5002' || ?pkey = 5002 || ?pkey = '5003' ||
?pkey = 5003 || ?pkey = '5004' || ?pkey = 5004 || ?pkey = '5005' || ?pkey = 5005 || ?pkey =
'5006' || ?pkey = 5006 || ?pkey = '5007' || ?pkey = 5007 ||...).
?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.go.
th/thai_location#amphur> .                                                    (1)
OPTIONAL{ ?x <http://data.go.th/thai_location#has_amphur_name> ?has_amphur_name}
OPTIONAL { ?x <http://data.go.th/thai_location#has_postal_code> ?has_postal_code }
OPTIONAL { ?x <http://data.go.th/thai_location#located_in_province> ?6a1 .
?6a1 <http://data.go.th/thai_location#has_province_name> ?located_in_province4 . }
OPTIONAL { ?x <http://data.go.th/thai_location#located_in_province> ?8a1 .
?8a1 <http://data.go.th/thai_location#located_in_region> ?9a2 .
?9a2 <http://data.go.th/thai_location#has_region_name> ?located_in_province6 . } .   (2)
}ORDER BY ASC(xsd:integer(?pkey))
```

```
- {
    รหัสอำเภอ: "5001",
    ภาค: "ภาคเหนือ",
    รหัสไปรษณีย์: "50000",
    ชื่ออำเภอ: "เมืองเชียงใหม่ ",
    จังหวัด: "เชียงใหม่ "
  },
  {
    รหัสอำเภอ: "5002",
    ภาค: "ภาคเหนือ",
    รหัสไปรษณีย์: "50160",
    ชื่ออำเภอ: "จอมทอง ",
    จังหวัด: "เชียงใหม่ "
  },
- {
    รหัสอำเภอ: "5003",
    ภาค: "ภาคเหนือ",
    รหัสไปรษณีย์: "50270",
    ชื่ออำเภอ: "แม่แจ่ม ",
JSON OUTPUT
```

Fig. 10. Example of SPARQL query to obtain all property values of the returned subjects

In Fig. 10, (1) specifies conditions of "FILTER" that use IDs returned from SPARQL query shown in Fig. 9 and (2) adds "OPTIONAL" clause in getting each property value.

4 Data Mashup

In this section, we present an implementation of dataset mashup in an application prototype, which consumes the data from the weather and reservoir statistics datasets via the provided APIs. In this application, the water volume levels in some selected dams are visualized together with the rainfall statistics of the same provinces. This application visualizes statistics data in various forms such as in google map, bar chart and table to allow the user to see the relations between rainfalls and water volume in reservoir. In consuming the data, we requested RESTFul APIs from Demo-api.data.go.th for two datasets: dam levels statistics (2011–2014) and rainfall statistics (2014). We manually obtained location information (latitude, longitude) of dams from Wikipedia. Figure 11 demonstrates the relationships between the user interface, APIs and data sources.

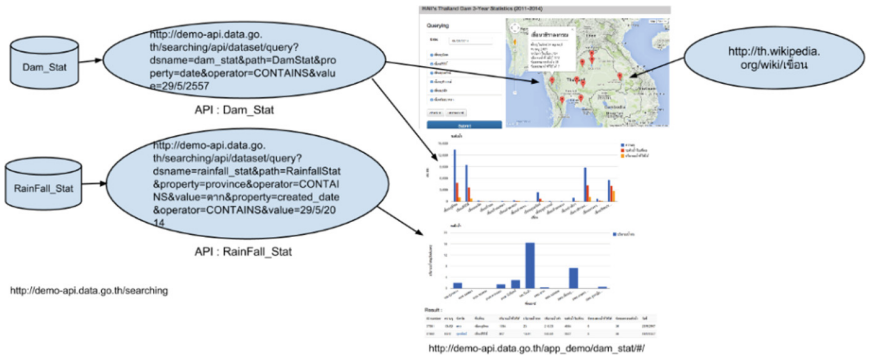


Fig. 11. Sample application demonstrating the dataset mash-up

In terms of application development, accessing open data via RESTFul APIs is more convenient and more flexible comparing to manually collecting the full dataset files from Data.go.th. The query results from RESTFul APIs are in the more ready-to-use format. Developers also do not need to manually collect the data files. In terms of data mash-up, access to multiple datasets can be done in a uniform way through the RESTFul APIs, which allow the developers to programmatically combine the data from different datasets more conveniently.

5 Conclusion

In this paper, we present an RDF dataset management framework and system for publishing RDF datasets on Data.go.th based on the OAM framework. This system provides services such as database-to-ontology mapping and conversion of relational database to RDF/XML format. After publishing an RDF dataset on the portal, the system provides the form-based SPARQL and RESTFul APIs data querying services for each dataset automatically. Users who want to use the published data do not need to write SPARQL queries. With the uniform query pattern of the provided APIs, it can help users to reduce time in learning and developing an application from open government datasets comparing to using the dataset files. To demonstrate consuming the open data APIs, we implemented a sample mash-up application which utilized dataset querying APIs of multiple datasets from Demo-api.data.go.th. Our future work will extend the OAM framework to support namespace management of the datasets for better support of linked open data and to simplify the process of publishing RDF datasets on the portal.

Acknowledgments. This project was funded by the Electronic Government Agency (EGA) and the National Science and Technology Development Agency (NSTDA), Thailand.

References

1. Huijboom, N., Van Den Broek, T.: Open data: an international comparison of strategies. Eur. J. ePract. **12**, 1–13 (2011)
2. Ding, L., Lebo, T., Erickson, J.S., Difranzo, D., Williams, G.T., Li, X., Michaelis, J., Graves, A., Zheng, J.G., Shangguan, Z., Flores, J., McGuinness, D.L., Hendler, J.A.: TWC LOGD: a portal for linked open government data ecosystems. J. Web Semant. **9**, 325–333 (2011)
3. Shadbolt, N., O'Hara, K., Berners-lee, T., Gibbins, N., Glaser, H., Hall, W., Schraefel, M.C.: Linked open government data: Lessons from data.gov.uk. IEEE Intell. Syst. **27**, 16–24 (2012)
4. 5 Star Open Data. http://5stardata.info/
5. Buranarach, M., Thein, Y.M., Supnithi, T.: A community-driven approach to development of an ontology-based application management framework. In: Lecture Notes Computing Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7774, pp. 306–312 (2013)
6. About—PublicData.eu. http://publicdata.eu/about
7. DaPaaS|DaPaaS—a data-and-platform-as-a-service approach to efficient open data publication and consumption. http://project.dapaas.eu/
8. Kim, S., Kim, S., Berlocher, I., Lee, T.: RDF based linked open data management as a DaaS Platform LODaaS. In: The First International Conference on Big Data, Small Data, Linked Data and Open Data (2015)
9. Mouromtsev, D.I., Vlasov, V.V., Parkhimovich, O.V., Galkin, M., Knyazev, V.S.: Development of the St. Petersburg's linked open data site using information workbench. In: 2013 14th Conference of Open Innovations Association (FRUCT), Espoo, pp. 77–82 (2013)
10. Maali, F., Cyganiak, R., Peristeras, V.: A publishing pipeline for linked government data. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7295, pp. 778–792 (2012)

Probabilistic Argumentation for Service Restoration in Power Distribution Networks

Nguyen Duy Hung(✉), Santi Kaisaard, and Saung Hnin Pwint Oo

Sirindhorn International Institute of Technology, Pathum Thani, Thailand
hung.nd.siit@gmail.com, santi.kaisa@gmail.com, hnin.pwint172004@gmail.com

Abstract. Service restoration in power distribution systems aims at finding plans to restore power considering multiple objectives and constraints. In this paper, we use Probabilistic Argumentation to integrate diverse knowledge sources about alternative plans. We then harness the automated probabilistic and logical reasoning of Probabilistic Argumentation to evaluate the provability degree that a restoration plan under consideration ensures desired objectives and constraints. The key advantage of our approach is perhaps its capability to produce arguments/proofs for/against restoration plans together with probabilities of their acceptances, in presence of possible incompleteness and uncertainty of the knowledge about such plans.

Keywords: Service restoration · Power distribution · Argumentation

1 Introduction

Fault in power distribution system is unpredictable, and often results in a power outage. Service restoration aims at finding restoration plans—series of switching operations—to restore power for the outage area. Alternative plans need to be constructed, then compared according to how their executions will realize desired objectives and constraints. So far many methods have been proposed to solve the service restoration problem, from different perspectives or interests, focusing on only one or several steps of the problem [1,2]. For example, in [3] heuristics representing the expertise of experienced operators are hard-coded in the proposed algorithm to construct alternative plans. In [4] Huang C.M. proposes to use Fuzzy cause-effect networks to deal with imprecise linguistic terms occurring in such heuristics. In [5] the authors highlighted the concept of relative performance index in evaluating and ranking alternative plans. Optimization techniques like Ant-Colony Optimization are applied when there is a large number of plans to consider [6–8]. Contributing to this line of research, in this paper, assuming that alternative plans have been identified, we focus on evaluating these plans taking into account desired objectives and constraints. We view this evaluation as consisting of two steps in an overarching argumentation-based model of service restoration shown in Fig. 1: (1) constructing a Probabilistic Argumentation

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_5

framework $\mathcal{P}$ integrating diverse knowledge bases about the plan; (2) reasoning within $\mathcal{P}$ to compute the degree of provability that the plan ensures desired objectives and constraints. The key advantage of our approach is perhaps its capability to produce arguments/proofs for/against restoration plans together with probabilities of their acceptances, in presence of possible incompleteness and uncertainty of the knowledge about such plans.

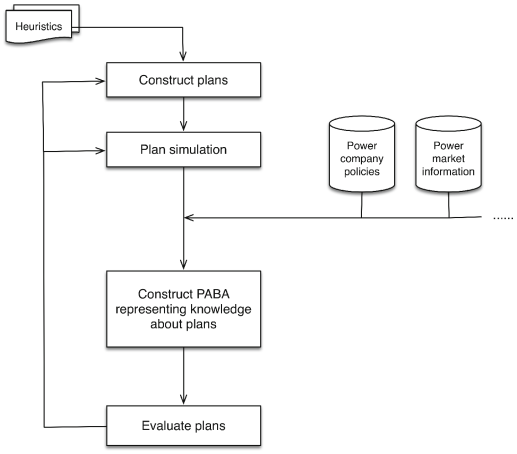


Fig. 1. An argumentation-based view of service restoration problem

The organization of this paper is as follows: Sect. 2 describes the current practices at a power company to deal with the service restoration problem, stressing the importance of the knowledge about plans from plan simulations. Section 3 recalls Probabilistic Assumption-based Argumentation frameworks (PABA) from [9] and presents Bayesian PABA frameworks—a simple class of PABA frameworks yet expressive enough for our purpose. Section 4 structures PABA frameworks to address the service restoration problem. Section 5 presents a prototype implementation and Sect. 6 concludes.

2 Simulation-Based Service Restoration

To get a handle on current practices in service restoration, we did a case study at a power company.[1] Figure 2 shows a power distribution system under its management which consists 3 substations, 5 feeders, 26 loads, and 19 switches. Power from substations flow through feeders to be consumed at loads. Loads supported by the same feeder form a zone. Different zones connect with others via switches. Often the spare capacities of support feeders and the total loads of zones are observed frequently by SCADA systems. Loads often vary during

[1] Provincial Electricity Authority of Thailand.

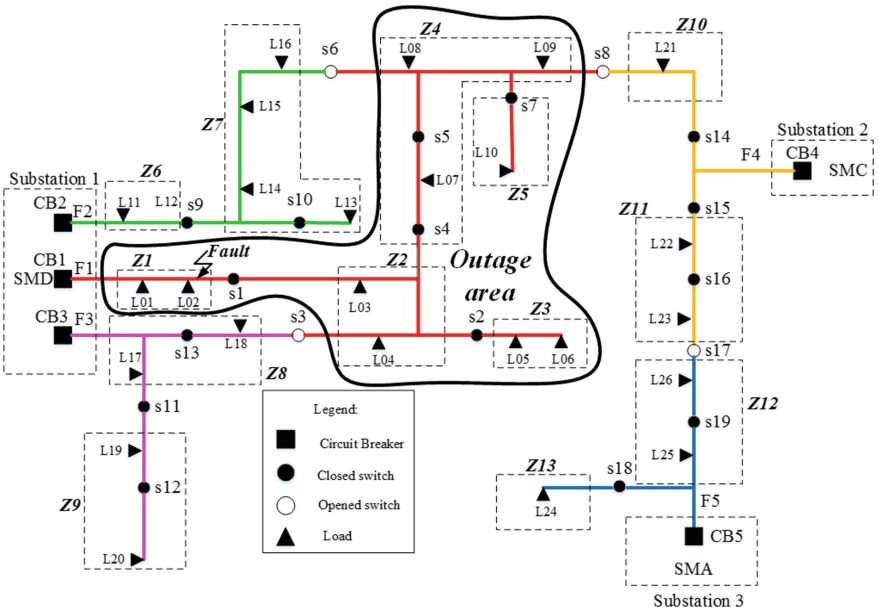


Fig. 2. A power distribution network of provincial electricity authority, Thailand

Table 1. Prefault feeder and zone load data (SCADA's observations)

Light load case	Feeder [MW]	F1	F2	F3	F4	F5		
		3.18	5.78	3.78	5.84	2.89		
	Zone [MW]	Z1	Z2	Z3	Z4	Z5	Z6	Z7
		0.22	0.22	1.53	0.57	0.64	0.29	5.49
		Z8	Z9	Z10	Z11	Z12	Z13	
		1.89	1.89	0.64	5.20	1.45	1.45	
Peak load case	Feeder[MW]	F1	F2	F3	F4	F5		
		4.98	7.35	5.05	9.39	5.32		
	Zone [MW]	Z1	Z2	Z3	Z4	Z5	Z6	Z7
		0.35	0.35	2.39	0.90	1.00	0.37	6.98
		Z8	Z9	Z10	Z11	Z12	Z13	
		2.53	2.53	1.03	8.36	2.66	2.66	

time of day. Table 1 exemplifies SCADA's observations at a light-load period and a peak-load period. Figure 3 shows SCADA's observations for the loads within 24 h.

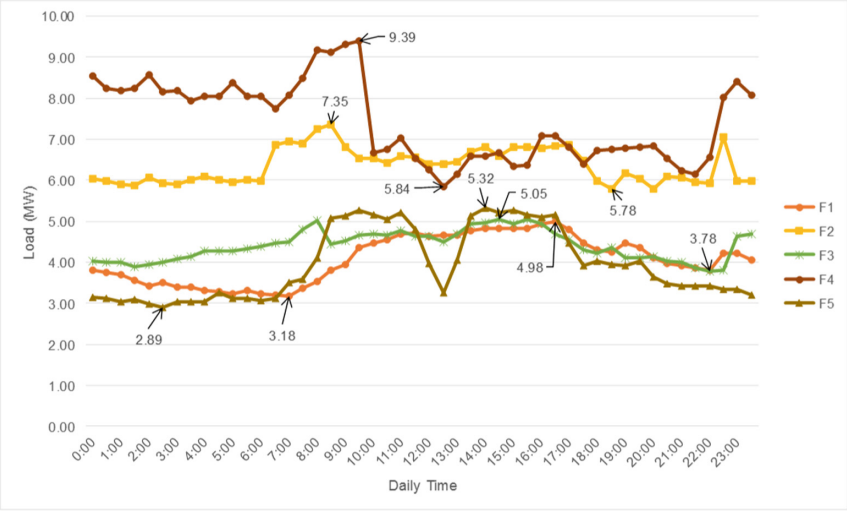


Fig. 3. SCADA's observations for the loads within 24 h

Now suppose that a fault occurs at $Z1$ in feeder 1. The circuit breaker $CB1$ will trip automatically to isolate the fault, causing a power outage in all zones from $Z1$ to $Z5$. Assume that the operator has opened switch $S1$ to isolate the fault. Now he needs to find a plan to restore power for all loads in the outage area, which consists of zones from $Z2$ to $Z5$ (see Fig. 2). The first step is to construct possible plans. Here by experience the operator often applies several heuristics, for example: H1 (*Group restoration*): If supporting feeders have enough spare capacity for the entire outage area, then close a normally opened switch between these feeders and the outage area; H2 (*Zone restoration*): If the supporting feeder does not have enough spare capacity, then close a normally opened switch to transfer some load of the outage area to a lateral feeder; H3 (*Load transfer*): The capacity of a supporting feeder can be increased by transferring some part of its load to other feeders. One application of a heuristic gives rise to one switching operation, hence to construct a plan, the operator may need to apply different heuristics multiple times, as illustrated by Table 2. Once the operator has constructed possible plans, he needs to evaluate how each plan satisfies predefined objectives and constraints. Table 3 below lists common objectives and constraints.

Intuitively a plan falsifying a constraint will be considered as totally unfit; and the more a plan minimizes important objective functions, the greater its degree of "fitness" will be. Here mathematically an objective (resp. constraint)

Table 2. Construction of possible plans by applying heuristics

Plans	Switching operations	Applied heuristics
# 1	Close S8	H1
# 2	Close S3	H1
# 3	Open S4, Close S6, Close S3	H2, H1, H1
...	...	...

is a real-valued (resp. boolean-valued) function taking two kinds of input parameters: (1) *system parameters* are those intrinsic to the power distribution system and often they can be computed from a physical description of the system, and (2) *environment parameters* are those extrinsic to the power distribution system and often they come from the environment. For example, all input parameters of objective function f_1 (determining the amount of unrestored load) are system parameters, while for C_3 (feeder line currents must be in range), I_js are system parameters, however I_{max} is an environment parameter since it is specified by the power company. This classification is not meant to be clear-cut, but often it makes sense, and more importantly it suggests that we can obtain approximations of system parameters by physically simulating a plan. Indeed in our previous work [11], we have demonstrated the idea by specifying power distribution systems within DIgSILENT PowerFactory [12]—an industrial-strength power system simulation software. We then develop a module which reads system parameters from DIgSILENT PowerFactory, environment parameters from other knowledge bases (e.g. power company's policies) and produces as output values of objective functions and constraints. Table 4 shows the outputs of the module for Plan #1 (Close S8): in a light-load period, $f_1 = 0.220$ and C_3 is fulfilled; however, in a peak-load period, $f_1 = 0.350$ and C_3 is unfulfilled. Recall that $C_3 \triangleq I_j \leq I_{max}$ where I_js are system input parameters. Hence the values of these I_js are provided by DIgSILENT, as illustrated by a graphical user interface of DIgSILENT in Fig. 4.

So far many methods have been proposed to solve the service restoration problem, from different perspectives or interests, focusing on only one or several steps of the problem (plan construction, plan evaluation or weighting/ranking). However, in practice power restoration is still a manual process responsible by experienced operators. This is perhaps proposed systems are not general or flexible enough to be applied in actual practice of service restoration, which is often decided on case-by-case basis.

3 Probabilistic Assumption-Based Argumentation

In general, knowledge about a plan consists of a probabilistic (quantitative) part (containing, for example the degree of belief of the operator about the reliability of the simulation or about the current load status (light or heavy) of the power

Table 3. Common objectives and constraints

Objectives	Input parameters	Information sources
Minimize the unrestored load $f_1 \triangleq \sum_{i=1}^{N_{bus}} L_i k_i$ (unit: MW)	N_{bus}: the number of restored buses	DIgSILENT
	L_i : the load at the ith bus	
	k_i : the status at the ith bus	
Minimize the operation time $f_2 \triangleq \sum_{i=1}^{N_s} (OT_{Ai} + OT_{Mi})$ (unit: second)	N_s: the number of switching operations	DIgSILENT
	OT_{Ai} : the operation time of a remote-controlled switch	Operator experience
	OT_{Mi} : the operation time of a manually-controlled switch	
Minimize the power loss $f_3 \triangleq \sum_{j=1}^{N_b} (\|I_j\|^2 R_j)$ (unit: MWh)	N_b: the number of branches	DIgSILENT
	I_j: the current of the jth branch	
	R_j: the resistance of the jth branch	
Constraints	**Input parameters**	**Information sources**
Maintain the radial network: C_1	-	DIgSILENT
Bus voltage in range $C_2 \triangleq V_{min} \leq V_j \leq V_{max}$	V_j: the voltage at jth bus	DIgSILENT
	V_{min}: minimal standard voltage	Power company's policy [10]
	V_{max}: maximal standard voltage at jth bus	
Feeder line currents in range $C_3 \triangleq I_j \leq I_{max}$	I_j: the current at the jth line	DIgSILENT
	I_{max}: maximally allowed current	Power company's policy [10]

Table 4. Simulation results

Objectives and constraints	Plan #1 in light load	Plan #1 in peak load
f_1	0.220	0.350
f_2	1,200	1,200
f_3	0.2301	0.4575
C_1	fulfilled	fulfilled
C_2	fulfilled	fulfilled
C_3	fulfilled	unfulfilled

network) and a logical (qualitative) part (containing, for example the logical rules to determine the overall degree of "fitness" of the plan). We could represent the probabilistic part by a Bayesian net and the logical part by a separate logical theory. The disadvantage of this approach is that we need to perform probabilistic reasoning (within the probabilistic part using various inference procedures for Bayesian nets) and logical reasoning (within the logical part using logical inference procedures) separately, then combine the results manually. To automate also this combination, we have to use a more general knowledge representation model that allows both probabilistic reasoning and logical reasoning. For this we propose to use probabilistic assumption-based argumentation (PABA) recently proposed in [9].

The logical part of an PABA framework is represented by an assumption-based argumentation framework (ABA) [13].

Definition 1. Assuming a language $\mathcal{L}$, an ABA framework is defined as a triple $\mathcal{F} = (\mathcal{R}, \mathcal{A}, \bar{})$ where $\mathcal{R}$ is a set of inference rules of the form $l_0 \leftarrow l_1, \ldots, l_n$ (for $n \geq 0$), $\mathcal{A} \subseteq \mathcal{L}$ is a set of assumptions, and $\bar{}$ is a (total) one-to-one mapping from $\mathcal{A}$ into $\mathcal{L}$, where $\overline{x}$ is referred to as the *contrary* of x. Assumptions in $\mathcal{A}$ do not appear in the heads of rules in $\mathcal{R}$.

- A *(backward) deduction* of a conclusion π supported by a set of premises Q is a sequence of sets $S_1, S_2, \ldots, S_n$ where $S_i \subseteq \mathcal{L}$, $S_1 = \{\pi\}$, $S_m = Q$, and for every i, where σ is the selected proposition in S_i: $\sigma \notin Q$ and $S_{i+1} = S_i \backslash \{\sigma\} \cup body(r)$ for some inference rule $r \in \mathcal{R}$ with $head(r) = \sigma$.
- An argument for $\pi \in \mathcal{L}$ supported by a set of assumptions Q is a (backward) deduction δ from π to Q and denoted by (Q, δ, π). An argument (Q, δ, π) attacks an argument (Q', δ', π') if π is the contrary of some assumption in Q'. For simplicity, we often refer to an argument (Q, δ, π) by (Q, π) if there is no possibility for mistake.
- A set of assumptions A_1 attacks a set of assumptions A_2 iff an argument supported by a subset of A_1 attacks an argument supported by a subset of A_2.

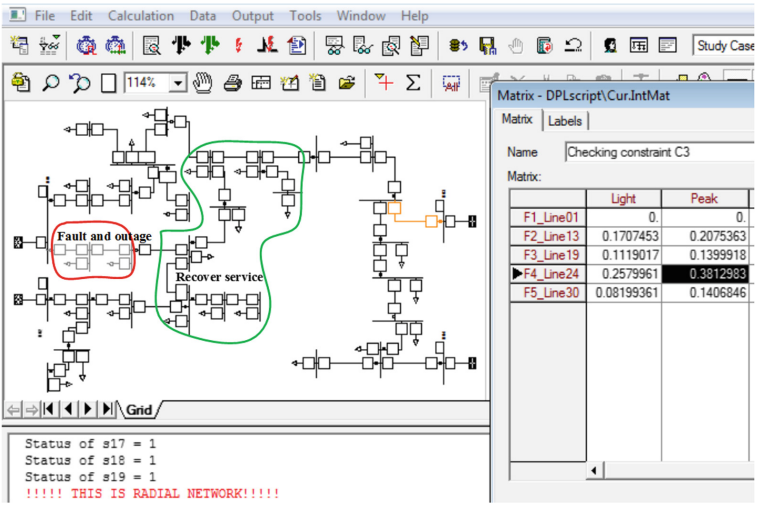


Fig. 4. Simulation of plan #1 by DIgSILENT

Example 1. Consider ABA $\mathcal{F} = (\mathcal{R}, \mathcal{A}, \overline{\ })$ where $\mathcal{R} = \{\neg do \leftarrow \neg C_3;\ \neg do \leftarrow \neg C_2;\ \ C_3 \leftarrow do\}$, $\mathcal{A} = \{do, \neg C_3, \neg C_2\}$ with $\overline{do} = \neg do, \overline{\neg C_3} = C_3, \overline{\neg C_2} = C_2$ which may represent a service restoration situation with two constraints C_2 and C_3. Argument for the plan under consideration (denoted by do) is simply $Arg_0 = (\{do\}, do)$. Arguments $Arg_1 = (\{\neg C_3\}, \neg do)$ and $Arg_2 = (\{\neg C_2\}, \neg do)$ say that the plan under consideration should be given up if it does not satisfy either constraint. Using the rule $C_3 \leftarrow do$ we can construct another argument $Arg_3 = (\{do\}, C_3)$ which attacks Arg_1. However we can not construct any argument that attacks Arg_2.

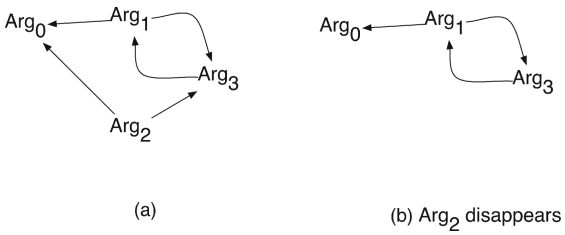


Fig. 5. Abstract argumentation frameworks

An ABA framework $\mathcal{F}$ generates an abstract argumentation framework [14] consisting of arguments and attacks that can be constructed from $\mathcal{F}$. An abstract argumentation framework could be represented graphically by a graph where

nodes represent arguments and edges represent attacks, as illustrated by Fig. 5a for the ABA of Example 1.

With arguments and attacks defined, all argumentation semantics of abstract argumentation [14] can be applied for an ABA $\mathcal{F}$, as follows.

- A set of assumptions is *admissible* iff it does not attack itself and attacks any $A \subseteq \mathcal{A}$ that attacks it.
- A proposition π is said to be an (admissible) *consequence* denoted ABA $\mathcal{F} \vdash \pi$ iff π is supported by some admissible set of assumptions.

Example 2. (Cont. Ex 1) $\neg do$ (but not do) is a consequence of the given ABA $\mathcal{F}$. However, if we consider only constraint C_3 (by removing the second rule $\neg do \leftarrow \neg C_2$), then do becomes a consequence in the new ABA because Arg_2 can be no longer constructed (see Fig. 5b).

The probabilistic part of an PABA framework is represented by a set of probabilistic assumptions and probabilistic rules as follows.

Definition 2. A probabilistic assumption-based argumentation (PABA) framework $\mathcal{P}$ is a triple $(\mathcal{A}_p, \mathcal{R}_p, \mathcal{F})$ satisfying the following properties

1. $\mathcal{A}_p$ is a set of **probabilistic assumptions** where
   - Elements of $\mathcal{A}_p \cup \neg\mathcal{A}_p$, where $\neg\mathcal{A}_p = \{\neg p \mid p \in \mathcal{A}_p\}$ and $\neg$ is the classical negation operator, are called **probabilistic literals**.
   - A **possible world** of $\mathcal{P}$ is a maximal (wrt set inclusion) consistent subset of $\mathcal{A}_p \cup \neg\mathcal{A}_p$.
2. $\mathcal{R}_p$ is a set of probabilistic rules of the form

$$[\alpha : x] \leftarrow \delta_1, \ldots, \delta_n \quad n \geq 0, x \in [0, 1]$$

   where proposition $[\alpha : x]$, called a **probabilistic proposition**, represents that the probability of probabilistic literal α is x.
3. $\mathcal{F} = (\mathcal{R}, \mathcal{A}, \overline{\ })$ is an ABA framework.

Example 3. (Cont. Ex 1) In practice of service restoration, the operator often has to predict the effects of a plan by, for example physical simulation. Suppose that for the plan under consideration, the simulation says that C_3 is satisfied only if the power system is in a light load period. To represent this, one may change the rule $C_3 \leftarrow do$ to $C_3 \leftarrow do, p_{lightLoad}$ where $p_{lightLoad}$ is a probabilistic assumption.[2] So the PABA framework representing the problem is $\mathcal{P} = (\mathcal{A}_p, \mathcal{R}_p, \mathcal{F})$ where $\mathcal{F}$ contains $\neg do \leftarrow \neg C_2$; $\neg do \leftarrow \neg C_3$; and $C_3 \leftarrow do, p_{lightLoad}$; $\mathcal{A}_p = \{p_{lightLoad}\}$; and $\mathcal{R}_p$ may contain $[p_{lightLoad} : 0.6] \leftarrow$ and $[\neg p_{lightLoad} : 0.4] \leftarrow$ which together say that the probability that the system is in a light load period is 0.6

[2] This new rule still does not capture the issue, which is addressed in the next section, that the simulation may or may not be reliable.

For each possible world ω of $PABA\ \mathcal{P} = (\mathcal{A}_p, \mathcal{R}_p, \mathcal{F})$, let's define $ABA\ \mathcal{P}_\omega \triangleq (\mathcal{R} \cup \mathcal{R}_p \cup \{p \leftarrow | \ p \in \omega\}, \mathcal{A}, \overline{\ })$, the instantiation of $PABA\ \mathcal{P}$ by the truths that ω asserts. $ABA\ \mathcal{P}_\omega$ could be also called the $PABA\ \mathcal{P}$ conditional to ω.

For instance, in Ex 3, $\mathcal{W} = \{\{p_{lightLoad}\}, \{\neg p_{lightLoad}\}\}$. So for $\omega_1 = \{p_{lightLoad}\}$, $ABA\ \mathcal{P}_{\omega_1}$ is obtained from $\mathcal{F}$ by adding two rules: $[p_{lightLoad} : 0.6] \leftarrow$ and $p_{lightLoad} \leftarrow$.

Intuitively probabilistic part defined by $\mathcal{A}_p$ and $\mathcal{R}_p$ intends to generate a probability distribution over the set of possible worlds $\mathcal{W}$. To ensure the coherence of this distribution, the authors in [9] have introduced some conditions on $\mathcal{R}_p$, which are skipped because of lack of space. In this paper we use a simple class of PABAs, called *Bayesian* PABAs that satisfy all conditions in [9] (i.e. Bayesian PABAs always generate a coherent probability distribution). In Bayesian PABAs, $\mathcal{A}_p$ and $\mathcal{R}_p$ represents a Bayesian Network by having each rule $r \in \mathcal{R}_p$ represents one entry of a conditional distribution table. Let $P_{BN}(\omega)$ denote the probability of possible world ω according to the Bayesian Network. The probability of a proposition π being accep with $\mathcal{P}_\omega$ is defined as follows;

$$Prob(\pi) \triangleq \sum_{\omega \in \mathcal{W}: ABA\ \mathcal{P}_\omega \vdash \pi} P_{BN}(\omega)$$

Example 4. (Cont. Ex 3) Let's compute $Prob(do)$. There are two possible worlds in $\mathcal{W}$: $\omega_1 = \{p_{lightload}\}$ and $\omega_2 = \{\neg p_{lightload}\}$. It is easy to see that $ABA\ \mathcal{P}_{\omega_1} \vdash do$ but it is not the case that $ABA\ \mathcal{P}_{\omega_2} \vdash do$. Hence $Prob(do) = P_{BN}(\omega_1) = 0.6$.

4 Probabilistic Assumption-Based Argumentation Framework for Service Restoration

In this section we specify a typical structure of PABA for a service restoration problem. Assuming a set $\mathcal{C} = \{C_1, C_2, C_3, \dots\}$ of constraints, $\mathcal{O} = \{f_1, f_2, f_3, f_4, \dots\}$ of objective functions where for each $f_i \in \mathcal{O}$ there are numbers f_i^{min}, f_i^{max} representing the minimal and maximal values of f_i. The knowledge about a particular plan is represented by such a PABA $\mathcal{P} = (\mathcal{A}_p, \mathcal{R}_p, (\mathcal{R}, \mathcal{A}, \overline{\ }))$ that:

- The underlying language $\mathcal{L}$ contains: a set $\{C_1, C_2, C_3, \dots, \neg C_1, \neg C_2, \neg C_3, \dots\}$ of propositions representing the satisfaction/dissatisfaction of the constraints; a set $\{f = r \mid f \in \mathcal{O}, r \text{ is a number}\}$ of propositions representing predicted values of objective functions; and a set $\{F_1, F_2, F_3, F_4, \dots\}$ of propositions representing that objective functions $f_1, f_2, f_3, f_4, \dots$ are successfully minimized (if f_i is close enough to f_i^{min}, then F_i is true).
- $\mathcal{A}$ contains the following assumptions
  - $\neg C_1, \neg C_2, \neg C_3, \dots$, where $\overline{\neg C_i} = C_i$ representing that all constraints can be considered as unfulfilled unless there are evidences on the contrary.

  - $\neg F_1, \neg F_2, \neg F_3, \neg F_4, \ldots$, where $\overline{\neg F_i} = F_i$ representing that if there is no evidence that F_i is fulfilled, then $\neg F_i$ is assumed.
  - do, where $\overline{do} = \neg do$, representing the selection of the plan.
- $\mathcal{A}_p$ contains the following probabilistic assumptions
  - $p_{peakLoad}, p_{lightLoad}, \ldots$ representing uncertain environment parameters.
  - $p_{reliableSimulationOf-C_i}$ (resp. $p_{reliableSimulationOf-f_i}$) representing the reliability of the simulation in predicting constraint C_i (resp. objective function f_i).
  - $p_{importanceOf\_f_i}$ representing the importance of objective function f_i.
- $\mathcal{R}$ contains subsets R_s and R_g
  - R_s represents the results of simulating the plan, containing

$$C_i \leftarrow do, L_1, \ldots, L_n, p_{reliableSimulationOf-C_i}$$

    where $C_i \in \mathcal{C}$ is a constraint, $L_1, \ldots, L_n$ are environment atoms (e.g. $p_{peakLoad}$, $p_{lightLoad}$) when the simulation says that executing the plan in environment $L_1, \ldots L_n$ fulfills constraint C_i; and containing

$$f_i = r \leftarrow do, L_1, \ldots, L_n, p_{reliableSimulationOf-f_i}$$

    when the simulation says that executing the plan in environment $L_1, \ldots L_n$ causes objective function f_i to acquire value r.
  - R_g represents how constraints and objective functions are aggregated to determine the overall fitness of the plan, so containing

$$\neg do \leftarrow \neg C_i$$

    representing that if constraint $C_i \in \mathcal{C}$ is not fulfilled, then the plan should be dropped; and containing

$$\neg do \leftarrow \neg F_i, p_{importanceOf\_f_i}$$

    representing that a good plan should minimize important objective functions; and containing

$$F_i \leftarrow f_i = r, r \approx f_i^{min}$$

    representing that F_i is close to true if the value of f_i is close to f_i^{min}.[3]

The evaluation of the plan under consideration amounts to computing $Prob(do)$ which could be interpreted as the degree of provability that the plan ensures declared objectives and constraints.

Our model is intuitive in the following manner: if there is no information about a constraint, say C_i, then $\neg C_i$ is assumed and hence the argument using $\neg do \leftarrow \neg C_i$ is attacked by no arguments in any possible world. Thus

[3] We could use a probabilistic assumption with default probability $1 - \frac{r - f_i^{min}}{f_i^{max} - f_i^{min}}$ to substitute the approximation $r \approx f_i^{min}$.

$Prob(do) = 0$. Similarly, if there is no information about an objective function, say f_i, then $\neg F_i$ is assumed. However, this does not mean that $Prob(do) = 0$ because an argument for $\neg do$ using the rule $\neg do \leftarrow \neg F_i, p_{importanceOf\_f_i}$ would need probabilistic assumption $p_{importanceOf\_f_i}$ whose truth value depends on a possible word. As this probabilistic assumption measures the importance of f_i, $Prob(do)$ will be reduced by an amount proportional to this importance.

5 Demonstration

We have implemented the proposed method using PENGINE, an engine for PABA we develop in a separate work. PENGINE is layered on top of Prolog and Python with libpgm,[4] as illustrated by the following architectural diagram (Fig. 6).

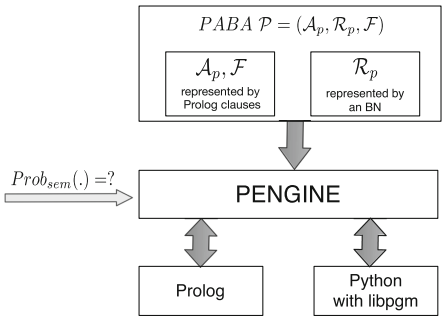


Fig. 6. PENGINE architectural diagram

To receive a Bayesian $PABA\ \mathcal{P} = (\mathcal{A}_p, \mathcal{R}_p, \mathcal{F})$, PENGINE provides two editors. The first receives $\mathcal{R}_p$ component in the form of a Bayesian network in JSON format according to libpgm. Figure 7 shows partially the content of this editor and Bayesian network for an instance of our service restoration problem.

The second editor of PENGINE receives components $\mathcal{F}$ and $\mathcal{A}_p$ represented by a set of Prolog clauses where `iRule(..., [...])` declares an inference rule in $\mathcal{F}$; `iNas([...])` lists assumptions in $\mathcal{F}$; `contr(...)` maps an assumption to its contrary; and `iPas([...])` lists elements of $\mathcal{A}_p$. Figure 8 shows how we define components $\mathcal{F}$ and $\mathcal{A}_p$ for an instance of the service restoration problem with one constraint (Feeder line current must be in range) and one objective (minimize the unrestored load). Figure 9 illustrates how one loads an PABA framework representing the knowledge about a restoration plan into PENGINE and then asks the engine to compute the provability degree that the plan ensures declared objectives and constraints.

[4] A Python BN library by CyberPoint Labs (see http://pythonhosted.org/libpgm).

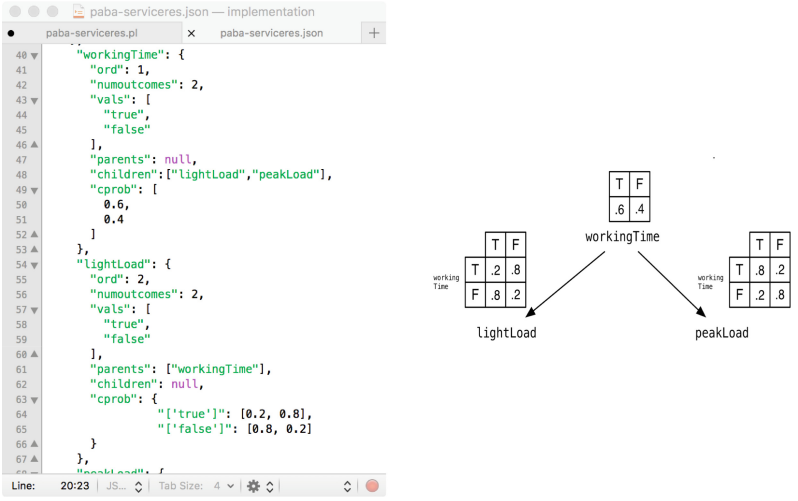


Fig. 7. $\mathcal{R}_p$ is by a Bayesian network

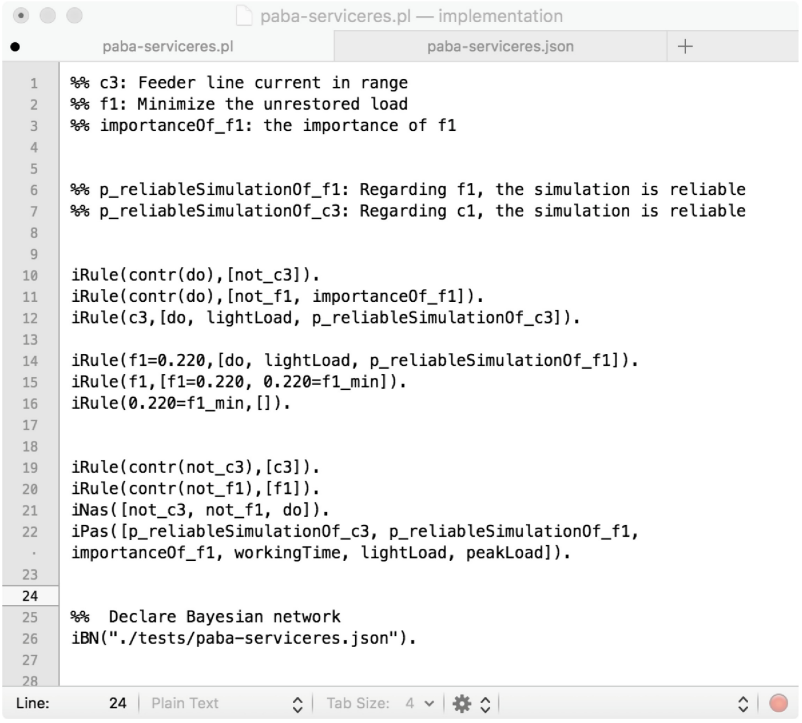


Fig. 8. Defining $\mathcal{F}$ and $\mathcal{A}_p$

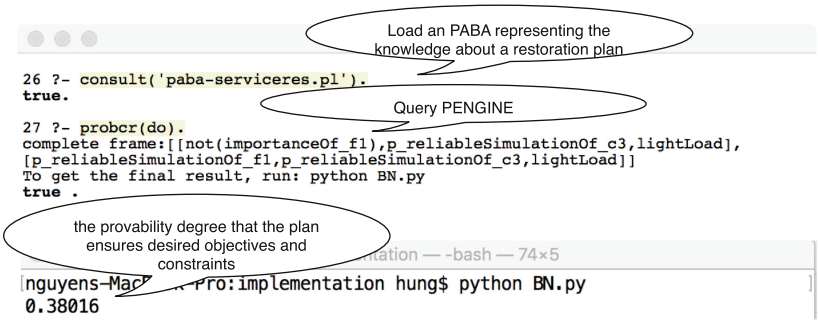


Fig. 9. Evaluating a restoration plan

6 Conclusions and Discussion

Despite a large number of restoration systems already proposed (e.g. [1–3,5]), in practice power restoration is still a manual process responsible by experienced operators. One reason that makes it hard to put proposed systems into practice is that operators constantly find some source of information important for a restoration task at hand, but ignored by the proposed systems. In this paper, we propose to use an Probabilistic Argumentation Framework to combine different information sources about a plan, then harnessing the automated probabilistic and logical reasoning within Probabilistic Argumentation to compute the degree of provability that the plan, if executed, shall ensures desired objectives and constraints. Such degrees can be used to compare different plans, however in this paper we do not focus on the construction of possible plans, as well as efficient ways to construct knowledge about them (e.g. by reusing some knowledge among similar plans). Techniques to deal with (big) spaces of plans have been studied elsewhere, using different methods, such as Genetic Algorithm in [7,8], Ant Colony Optimization in [6]. The relationship between the degree of provability for a plan presented in this paper, and other degrees measuring plans' goodness proposed in the literature, for example the relative performance index in [5] or those in [4,15] using Fuzzy Logic, is a future work.

Acknowledgements. This work was funded by: (1) SIIT Young Researcher Grant under Contract No SIIT-2014-YRG1; (2) Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Thammasat University.

References

1. Kumar, Y., Das, B., Sharma, J.: Multiobjective, multiconstraint service restoration of electric power distribution system with priority customers. IEEE Trans. Power Delivery, **23**(1), 261–270, Jan 2008

2. Zidan, A., El-Saadany, E.F.: A cooperative multiagent framework for self-healing mechanisms in distribution systems. IEEE Trans. Smart Grid **3**(3), 1525–1539 (2012)
3. Hsu, Y.-Y., Huang, M.M., Kuo, H.-C., Peng, S.K., Chang, C.W., Chang, K.J., Yu, H.S., Chow, C.E., Kuo, R.T.: Distribution system service restoration using a heuristic search approach. In: Proceedings of the 1991 IEEE Power Engineering Society Transmission and Distribution Conference, pp. 639–645, Sept (1991)
4. Huang, C.-M.: Multiobjective service restoration of distribution systems using fuzzy cause-effect networks. In: IEEE Transactions on Power Systems, vol. 18, no. 2, pp. 867–874. IEEE (2003)
5. Chen, W.-H.: Quantitative decision-making model for distribution system restoration. IEEE Trans. Power Syst. **25**(1), 313–321 (2010)
6. Karn, R., Kumar, Y., Agnihotri, G.: ACO based method for service restoration in distribution system using fuzzy M.O.M.C approach. Int. J. Soft Comput. Eng. **1**(4), September 2011
7. Kumar, Y., Das, B., Sharma, J.: Service restoration in distribution system using non-dominated sorting genetic algorithm. Electr. Power Syst. Res. **76**(9), 768–777 (2006)
8. Moazami, E., et al.: Optimal penalty method in distribution service restoration using genetic algorithm. In: 2013 IEEE 7th International Power Engineering and Optimization Conference (PEOCO). IEEE (2013)
9. Dung, P.M., Thang, P.M.: Towards (probabilistic) argumentation for jury-based dispute resolution. In: Computational Models of Argument: Proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8–10, pp. 171–182 (2010)
10. Provincial electricity authority, Thailand
11. Kaisaard, S., Hung, N.D.: A simulation-based framework for distribution service restoration. In: 7th International Conference on Mechanical and Electrical Technology (2015)
12. Digsilent powerfactory
13. Dung, P.M., Mancarella, P., Toni, F.: Computing ideal skeptical argumentation. Artif. Intell. **171**(10–15), 642–674 (2007)
14. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. Artif. Intell. **77**(2), 321–357 (1995)
15. Kaewmanee, J., Sirisumrannukul, S.: Multiobjective service restoration in distribution system using fuzzy decision algorithm and node-depth encoding. In: 2011 8th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 893–896. IEEE (2011)

Experimental Teaching of Collaborative Knowledge Creation Using Analogue Games

Tomomi Kaneko[1(✉)], Ryoju Hamada[2], and Masahiro Hiji[3]

[1] Hokkaido University of Science Junior College, Sapporo, Japan
kaneko@hus.ac.jp
[2] Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani, Thailand
hamada@siit.tu.ac.th
[3] Accounting School, Tohoku University, Sendai, Japan
hiji@econ.tohoku.ac.jp

Abstract. Collaborative knowledge creation is a key factor of success in the business world. However, it is difficult to teach it using a traditional lecture style at a university. To resolve this difficulty, the authors apply BASE business games, a participation-type education technique using analogue games, to teaching collaborative knowledge creation. As described herein, the authors used the SCC game and the SCC2 game, which are BASE business games, at the School of Management Technology (MT) of the Sirindhorn International Institute of Technology (SIIT), Thammasat University in 2014 and 2015. The authors analyzed student responses to a questionnaire from a collaborative knowledge creation perspective. Results show that this teaching method enhances their knowledge into meaningful experiences and encourages students to experience elementary collaborative knowledge creation with others.

Keywords: Analogue games · BASE business games
Knowledge integration · Collaborative knowledge creation
Supply chain collaboration (SCC)

1 Introduction

Students study diligently and acquire much knowledge through university lectures. This knowledge has high specialty and difficulty, but it is difficult to establish connections meaningfully. In the real business world, specialized knowledge alone is insufficient to produce or find business solutions. Students must learn how to integrate this knowledge meaningfully and achieve innovation with collaborative knowledge creation if they want to succeed in a business environment. Nevertheless, it is difficult to teach those points to students during a short time at the university level using conventional teaching methods alone. Therefore, the authors attempt to apply BASE business games to the teaching of collaborative knowledge creation. The authors have been developing tangible board games, called BASE business games, since 2007. The games have been played by more than 1000 students in Japan and Thailand. In terms of having students grasp a business outline, the authors believe that an analogue business

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_6

game is a respectable method. BASE business games have three major features. First, BASE business games are pure board games. Instead of computers, a tangible company board, cards, and market boards are represented on a tabletop. All activities are visible to other participants. Secondly, because of physical limits and time limitations, analogue games are inadequate to describe actual businesses. The authors waived some details and simplified actual businesses. Thereby, the authors were able to create a familiar game for beginners. Finally, BASE business games are applied for the development of future engineers in business or entrepreneurship courses as a part of a lecture. Engineers need not know everything about business, but they must learn management and accounting skills to a certain degree. The basic mode of playing BASE business games is simple. Two or three participants constitute a team and play roles as CEOs and CFOs in their respective companies, competing with other players' companies.

The authors can confirm some experiences related to using gaming simulations in education. Reed [1] describes one of the oldest manufactured business games. Po and Deng [2] report a recent example of a supply chain management game. Regarding software engineering, Zapata [3] attempted to compile software development games. Hamada et al. [4] produced SKG, which is a card game, to learn skills for the management of human resources, process, cash, and others in IT industry. The authors can confirm some experience related to using analogue games for education. Ichikawa [5] reported that a policy-making game was commonly used by military and government personnel before World War II in Japan. Nowadays, the Beer Game, which is still commonly used to teach supply chain management, is available in tabletop and computer versions. Of course, the tabletop version is still popular. Kikkawa [6] presents "Crossroad" as a card game to teach people how to confront a national disaster.

2 BASE Business Games at SIIT Lecture in 2014

The authors collaborated with SIIT from August 2010 to ascertain the global and general efficiency of BASE business games. SIIT provides students of the "Engineering Management (EM) Course" with lectures related to basic accounting, finance, human resource management, supply chain management, etc. Although students study diligently, their knowledge is not meaningfully connected. SIIT strives for students to unite their knowledge horizontally and to seek tools of meeting SIIT's object. Therefore, BASE business games have been adopted as a conclusion of their education. The authors created two types of BASE business games in accordance with SIIT's request: "Supply Chain Collaboration Game (SCC game)" and "Supply Chain Collaboration 2 Game (SCC2 game)." These business games require that students form teams and operate a mock-up company. They can understand and manage cash accounts, fixed costs, operations, inventory control, strategy, risks, and finally, supply chain collaboration as an extended concept of supply chain management experimentally. The authors also assume that the operation of a company, even if it is a mock-up company, causes them collaborative knowledge creation.

Figure 1 presents an SCC game outline. Players operate Smartphone Manufacturers and create smartphones comprising a motherboard and a display. Each motherboard

and display has six levels of quality. The price of each motherboard and display is fixed with the level of quality. Therefore, players must devote careful consideration to which combination represents the best choice for the market situation and other companies'

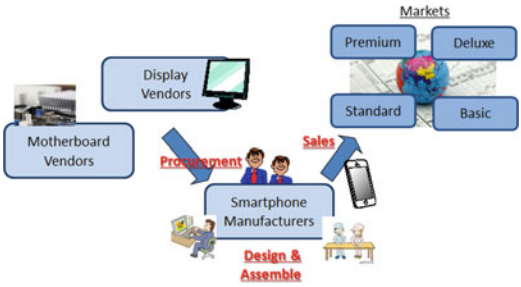


Fig. 1. Outline of the SCC game

strategies.

Figure 2 presents an SCC2 game outline. It is more complex than the SCC game. In the SCC2 game, players separate and form three companies: Motherboard Vendors, Display Vendors, and Smartphone Manufacturers. Each company assembles motherboards, displays, and smartphones, as in the SCC game. A salient difference between the SCC2 game and the SCC game is that Motherboard Vendors and Display Vendors are subcontractors of Smartphone Manufacturers. Thereby, Smartphone Manufacturers must negotiate with Motherboard Vendors and Display Vendors on prices and many motherboards and displays. All companies must negotiate with awareness of their own company's cash flow. This is an outstanding feature of the SCC2 game.

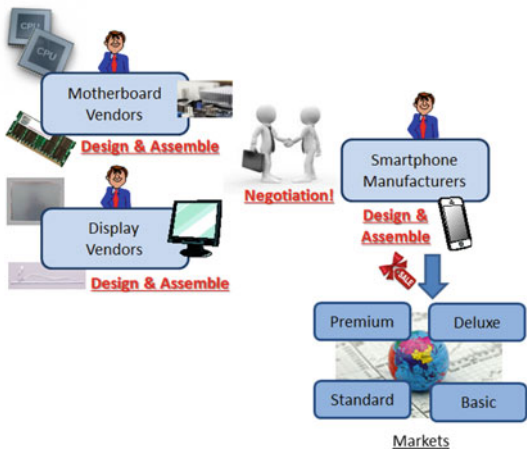


Fig. 2. Outline of the SCC2 game

During August 15–September 12, 2014, the authors conducted a lecture for 43 students: "Entrepreneurship for IT Business Development." All were senior students of the undergraduate Engineering Management (EM) Course. They already understood supply chain management to a certain degree. However, they had no ideas related to supply chain collaboration. The SCC game and SCC2 game were used in most lectures. At the beginning of the lecture, they were concerned about the adjunct Japanese associate professor and the contents of this lecture. However, as they came to understand the game rules, they were able to enjoy lectures more. Figure 3 presents photographs of this lecture.

Fig. 3. Photographs of the lecture

3 Discussion

To verify the game's effectiveness, the authors administered a self-evaluation questionnaire to students before and after the lecture. Table 1 presents results of their knowledge usage. These questions are single questions. Results indicate that they have already studied "Accounting", "Inventory Control" and "Marketing Research" because they are students of the Engineering Management course. After the lecture, the three ideas above and "Supply Chain Management" worked well in these games.

Table 1. Usage of their knowledge

	Have you ever studied the following ideas?		Do you think this knowledge and skills helped you play SCC & SCC2?	
	Yes	No	Yes	No
Supply chain management	21	18	35	4
Accounting	39	0	38	1
Inventory control	37	2	38	1
Human resource management	18	21	26	13
Marketing research	34	5	33	6

The authors assume that they understood these ideas as experiences. However, the result for "Human Resource Management" is low because these games have no mechanism for teaching "Human Resource Management."

Table 2 presents the reported degree of connection their knowledge and skills horizontally in the student mind. This questionnaire uses a Likert scale, an ordinal scale, to record responses. The average results of the question can be compared because the same students gave responses at different times. The average value of the question response increased 0.92. These results indicate that their reported knowledge and skills are connected well horizontally. Changes of each student are shown in the right table. This table shows that students showed low recognition of their own knowledge connection before the lecture, but the results improved significantly after the lecture, e.g. "from 2 to 5" for 3 students and "from 3 to 5" for 8 students. The score of only one student dropped "from 5 to 4", but the grade of the student was very high (B+). The authors assume that the student recognizes a weak horizontal connection of knowledge and skills through the lecture. Then the student marked a lower score. Overall results indicate that this teaching method is useful for experimental learning of knowledge integration and that it is especially useful for students exhibiting a low horizontal connection.

The authors also administered another questionnaire using a Likert scale for assessing the understanding of management knowledge of the students before and after the lecture. Comparison enabled the authors to evaluate the game impact. Responses for the following questions were given on a five-point scale.

Table 2. Degree of connection their knowledge and skill horizontally in your mind

Do you think this knowledge and skills are connected horizontally in your mind?		
Before	After	
0	0	1. Strongly no
6	0	2. No
13	2	3. Neutral
15	19	4. Yes
5	18	5. Strongly yes
3.49	4.41	Average

		After				
		1	2	3	4	5
Before	1	0	0	0	0	0
	2	0	0	0	3	3
	3	0	0	2	3	8
	4	0	0	0	11	4
	5	0	0	0	1	4

1. Strongly do not understand
2. Do not understand

3. Neutral
4. Understand
5. Strongly understand

The average responses to questions are presented in Table 3. Overall results show that the understanding of management knowledge was generally improved. Outstanding improvement can be recognized for Question 8 by a mean value increase of 1.21, Question 9 by a mean value increase of 1.02, and Question 11 by a mean value increase of 1.01. It is difficult to teach the contents of these questions using traditional lectures. The SCC games are designed to teach the importance of continuous collaboration with suppliers and inventory control as an experience. The authors confirm that SCC games satisfy the lecture course requirements.

Figure 4 presents comparison between before and after results on the understanding of management knowledge in 2014 and 2015. Overall results show that results of both years exhibit a similar tendency and understanding of management knowledge was generally improved.

Table 3. Impact of the games on the understanding of management knowledge

Questions	Before		After		Before/after difference
	Mean	SD	Mean	SD	
1. What would a company do when cash becomes short?	3.32	0.66	4.13	0.61	0.81
2. Why does a good company sometimes become bankrupt suddenly?	3.03	0.98	4.00	0.51	0.97
3. Why do most companies borrow money from a bank?	3.87	0.53	4.21	0.57	0.34
4. How does a company reduce manufacturing costs?	3.45	0.55	4.00	0.76	0.55
5. What would a company do when they produced too many goods?	3.50	0.65	4.36	0.74	0.86
6. Why is risk management important?	3.32	0.75	4.08	0.62	0.75
7. How is a break-even point calculated?	3.16	0.72	3.92	0.77	0.77
8. What does supply chain collaboration have as factors other than price?	2.71	0.65	3.92	0.58	1.21
9. Why is a continuous relationship with suppliers important to sustain a company?	3.29	0.73	4.31	0.57	1.02
10. Why do companies establish a business strategy?	3.24	0.54	4.05	0.76	0.81
11. Concept of inventory control	3.22	0.77	4.23	0.78	1.01
12. Concept of production planning	3.32	0.62	4.23	0.87	0.91
13. Concept of human resource development	3.21	0.62	3.77	0.87	0.56

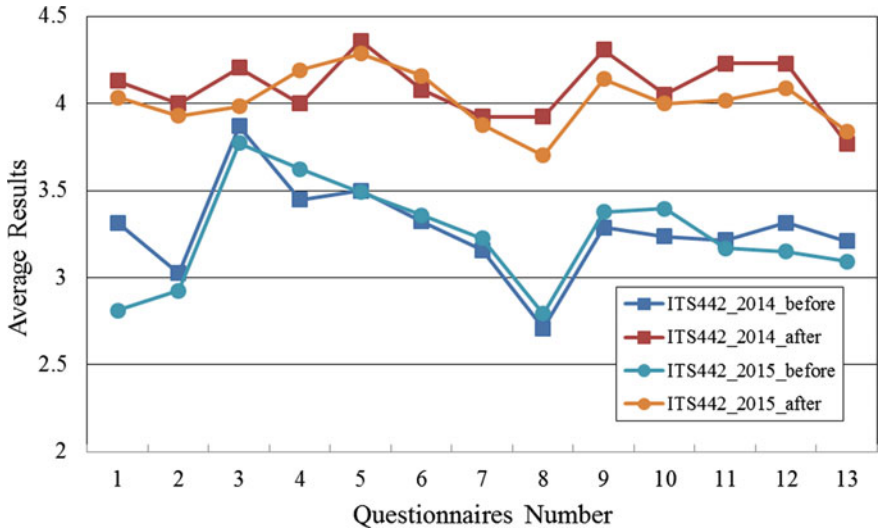


Fig. 4. Comparison between before and after results on the understanding of management knowledge in 2014 and 2015

Most outstanding improvement can be found for Questionnaire Number 8. The authors analyze details related to Questionnaire Number 8. Table 4 presents changes before and after results for each student on Questionnaire Number 8 in 2014 and 2015. A large number of results in 2014 are "2 + 2 = 4," which represent "result before the lecture + differences = result after the lecture," for 10 students and "3 + 1 = 4" for 10 students. However, the large number of results in 2015 is "3 + 1 = 4" of 13 students. Results show that the students who were unable to understand the issue before the lecture became able to understand it.

Table 4. Detail analysis of student answer on Question 8 in 2014 and 2015

2014	After						2015	After					
	1	2	3	4	5			1	2	3	4	5	
Before	1	0	0	0	0	0	Before	1	0	0	4	2	1
	2	0	0	1	10	2		2	0	3	4	3	0
	3	0	0	6	10	2		3	0	0	6	13	2
	4	0	0	1	2	1		4	0	0	3	3	3
	5	0	0	0	0	0		5	0	0	0	0	1

4 Conclusions

This paper presented the authors used the SCC game and the SCC2 game at the lecture of SIIT in 2014 and 2015 and considered the responses to questionnaires by students from a collaborative knowledge creation perspective.

(1) The SCC game and SCC2 games, which are designed for learning supply chain collaboration, are useful tools for teaching and learning. Through game play, students learn the importance of supply chain collaboration and acquire a holistic view of management.
(2) After the lecture, students can connect their knowledge and skills horizontally. The authors confirm that this teaching method encourages students to understand collaborative knowledge creation as experience.
(3) Detailed analyses of Questionnaire Number 8 indicate that this teaching method can facilitate the comprehension of students: they were unable to understand it before the lecture, but became able to understand it.

References

1. Kyle Reed, S.: The use of production scheduling simulation in a production planning and control course. Simul. Games Exp. Learn. Tech. **1**, 123–126 (1974)
2. Po, C., Deng, W.: Simulating processes: an application in supply chain management. Dev. Bus. Simul. Exp. Learn. **37**, 62–67 (2010)
3. Jaramillo, C.M.Z.: Teaching software development by means of a classroom game: the software development game. Dev. Bus. Simul. Exp. Learn. **36**, 156–164 (2009)
4. Hamada, R., Hiji, M., Kaneko, T.: Development of software engineering business board game. Dev. Bus. Simul. Exp. Learn. **41**, 292–299 (2014)
5. Ichikawa, A.: Early Japanese policy games. In: Back to the Future of Gaming, pp. 68–79. W. Bertelsmann Verlag, Bielefeld (2014)
6. Kikkawa, T.: Roles of play: the implications of roles in games. In: Proceedings of the 45th Conference of International Simulation and Gaming Association, pp. 268–275 (2014)

Ontology-Based Framework to Support Recommendation Rule Management Using Spreadsheet

Marut Buranarach[1(✉)], Thanyalak Rattanasawad[2], and Taneth Ruangrajitpakorn[1,3]

[1] Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center (NECTEC), Pathum Thani, Thailand
{marut.bur, taneth.rua}@nectec.or.th
[2] Faculty of Engineering, Department of Computer Engineering, Chulalongkorn University, Bangkok, Thailand
thanyalak.rattanasawad@gmail.com
[3] Faculty of Science and Technology, Department of Computer Science, Thammasat University, Pathum Thani, Thailand

Abstract. Recommender system development for the Semantic Web data typically requires ontology, rules and rule-based inference engine to be applied over the RDF data. Rule creation is a critical process in supporting knowledge base building and typically must rely on knowledge engineers. This paper describes an ontology-based framework to support knowledge engineers and experts in managing recommendation rules using a spreadsheet application by using their own vocabulary instead of ontology terms. Decision table, ontology, vocabulary mapping configuration, and recommendation rule template are core components of the framework. The framework was successfully validated in a large-scale project using a case study of excise duty recommender system. Our work differs from existing systems in that it is designed for the Semantic Web data and uses a flexible recommendation rule template. Our framework also emphasizes the need for knowledge engineers and experts to use their own vocabulary in the decision tables. Thus, it hides the complexity both in terms of syntax and terminology, which would make rule management task more user-friendly.

Keywords: Rule editor · Recommender system · Decision table
Knowledge engineering

1 Introduction

Recommender system is a type of system that generates meaningful recommendations to support user's decision. Recommender system development for the Semantic Web data typically requires ontology, rules and rule-based inference engine to be applied over the RDF data. Rule creation is a critical process in supporting knowledge base building and typically must rely on knowledge engineers and experts. However,

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_7

existing rule editors are normally too complex for users who have no background in logics.

There are at least three issues that make rule editors difficult for the users. Firstly, rule syntax is typically complex and requires some background in logics. In addition, there are various rule languages and syntax supported by different rule-based systems. Although many rule editors can hide the rule syntax, the users still need some background in rule language structure, e.g. predicates, variables, etc., in managing the rules. Secondly, the users often prefer to encode the rules in their own vocabulary rather than using the knowledge base vocabulary, e.g. ontology terms. Specifically, ontology and rules may be independently developed by different knowledge engineers or experts. Thus, allowing knowledge engineers or experts to create rules in their own terms can improve user-friendliness of the editor and reduce their effort in learning the knowledge base terminology. Thirdly, knowledge engineers or experts typically need to learn new application and user interfaces in rule creation and maintenance. By allowing them to use tools that they are already familiar in managing rules, such as office applications, would reduce their learning curve.

In this paper, we propose an ontology-based framework for generating recommendation rules from spreadsheets. Using this framework, knowledge engineers or experts can manage recommendation rules using a spreadsheet application and using their own vocabulary instead of ontology terms. The framework supports rule generation in four steps: (1) knowledge engineer or expert creates decision tables in spreadsheets, (2) knowledge engineer or expert provides mapping of user vocabulary to ontology vocabulary, (3) rule translation system creates rules in the OAM (Ontology Application Management) recommendation rule template format [2] (4) the system serializes rules in the template form to appropriate rule syntax for the inference engine. The recommendation rule template is a generalized rule expression that allows the created rules to be independent of rule language and syntax of the underlying rule-based systems.

The paper is organized as follows. Section 2 describes some existing approaches in designing rule editors. Section 3 describes design and implementation of an ontology-based framework for rule management using spreadsheets. Section 4 presents a use case scenario of the framework in a large-scale project supporting an excise duty recommender system. Section 5 provides conclusion and some future work.

2 Background

2.1 Rule Editors

A knowledge-based system is a system that employs human knowledge captured in a computer to solve problems that ordinarily require human expertise [5]. Rule-based knowledge is a typical representation of domain expert knowledge. In creating rule-based knowledge, knowledge engineers, who may lack the domain knowledge, and domain experts, who may lack background in logics, are normally involved. The first creation step is typically knowledge elicitation process, which knowledge engineer creates dialog with domain expert to elicit the relevant knowledge required for the

knowledge base. The second step normally involved knowledge engineer using rule management tools in coding the knowledge into the language supported by the knowledge-based system.

Although most rule management tools, i.e., rule editors, are designed to be used by knowledge engineers, several research [4, 6] have addressed the need for the tools that would also allow domain experts to directly manage the rule creation and maintenance process. There are mainly four approaches in designing rule editors to simplify rule creation and maintenance process for the users: controlled language-based, dialog-based, graph-based editors and decision tables.

- **Controlled language-based rule editor**. Controlled language (CL) is close to Natural language (NL) interface. Users can enter sentences in a language that is similar to English. The input texts are then converted to the rule language and syntax used by the knowledge-based system. Examples of editor in this category include ROO [4], ACEView [8], etc. An advantage of this approach is that the editor has a simple text-input interface. However, it is normally difficult to obtain good rule translation accuracy because of variety in users' rule writing styles.
- **Dialog-based rule editor**. This type of editor gathers the user input via form fields. The user fills the rule condition criteria via dialog-based or questionnaire-based interface. Examples of editors in this category includes domain independent framework such as KANAL [6] and domain dependent tools such as those supporting clinical reminder applications [11]. This approach has advantage that the interface has fixed information structure and is easier to control user's behavior. However, the user requires some learning effort in using these tools.
- **Graph-based or visual rule editor**. This type allows the users to formulate rules using graphical user interface such as graph-based modeling. Some tools also include visualized inference results. Examples of editor in this category include [7, 9]. This approach has an advantage that the relationships between entities and rules are clearly illustrated and can be easily adjusted from the interface. Domain experts still require some learning effort in using these tools.
- **Decision tables**. This type allows the user to define rules in decision table form. Each row normally represents one rule. Each column represents each condition criteria and value. The decision tables can be stored in CSV or Excel format and created using spreadsheet applications. Examples of editors in this category include Open Rules [10] and Drools [7]. This approach has an advantage that domain experts can use the tools that they are already familiar with in creating and maintaining rules. Domain experts still require some learning effort in creating decision tables.

2.2 Related Work

Our work focuses only on the decision table approach in rule management. Open Rules [10] and Drools [7] are two systems that use the decision table approach in rule editing. The created rules can be transformed to the declarative or programming language to be used in business rule engines and applications. Our work differs from both systems in that we propose an ontology-based framework that is designed for the Semantic Web

data, i.e. RDF(S) and OWL. Our work also designs a flexible recommendation rule template that can be serialized to different rule language syntax. In addition, our approach emphasizes the need for domain experts to use their own vocabulary in the decision tables. Thus, it aims to hide the complexity of the underlying rule-based systems both in terms of syntax and terminology, which would make rule management task more user-friendly to the users.

3 Ontology-Based Framework for Recommendation Rule Management Using Spreadsheets

The decision table approach using spreadsheet was adopted in our framework due to the following advantages.

- **Bulk rule creation**. Although rule editor is effective in terms of constraints, it is not efficient for the user when creating a large number of rules. With the decision table approach, the user can efficiently create a large number of rules using the spreadsheet application, e.g. using the copy/paste functions.
- **Rule updating and maintenance.** The decision tables stored in spreadsheet files are portable and interchangeable. The user can update the rules without requiring additional software.
- **Ruleset management.** Managing rulesets is important in rule management. A knowledge base may consist of multiple rulesets, i.e. subjects. In addition, a ruleset can have multiple sub-rulesets, i.e. sub-subjects. A spreadsheet document can represent a ruleset while a worksheet within the document can represent a sub-ruleset. This allows each ruleset and sub-ruleset to be independently managed before they are merged into a knowledge base.

The processes and system components of the ontology-based framework for recommendation rule management using spreadsheets are described as follows.

3.1 Rule Management Process

Figure 1 shows the layered architecture of the process. The rule management process consists of the following steps: creating decision tables, mapping decision tables with ontology, rule template generation, and rule syntax serialization.

- **Creating decision tables**. In creating decision tables in spreadsheet, each decision table contains rules of the same schema. Each attribute and criteria is defined in one column header. Table cells contain values of the respective table headers. Number of rows in the table corresponds to the number of rules. The attributes and criteria values are defined in user's own vocabularies.
- **Mapping decision tables with ontology**. The mapping configuration is typically created to map between the user vocabularies in the decision tables and ontology terms. The condition attributes and criteria values are mapped with ontology properties and classes respectively. The mapping is needed for both criteria attribute-to-property and criteria value-to-class-or-literal mappings.

- **Rule template generation**. Based on the decision tables, ontology and mapping configuration, the rule translator system generates rules according to a recommendation rule template format [2], which is a generalized rule framework that is independent of the underlying rule-based reasoners.
- **Rule syntax serialization**. The rules in the template form can be serialized into different rule syntax required by different systems.

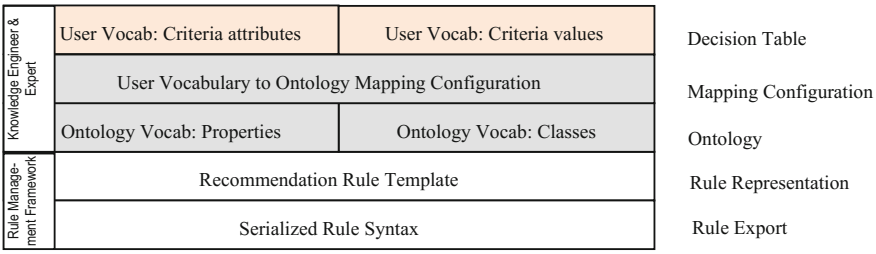


Fig. 1. Layered architecture of the ontology-based rule management process using spreadsheet

3.2 OAM Recommendation Rule Template

The Ontology Application Management (OAM) Framework [2] is an ontology application framework that focuses on simplifying creation and management of recommendation rules based on a recommendation template. The recommendation template consists of three main components: "recommendation", "recommend-of" and "recommend-to" resources, as shown in Fig. 2. A recommendation rule is created in a two-step process: create recommendation and link recommendation. Creating recommendation will create an instance of a recommendation container class, e.g. "Promotion" where the user can define conditions of class instances to be the recommendations, e.g. "Product". Linking recommendation allows the user to define conditions of class instances to recommendation receivers, e.g. "Customer". The framework facilitates the user to create such business logics that hides complexity of the rule syntax to be processed by reasoning engine.

In this framework, recommendation rule can be created based on a recommendation template, whose structure can be summarized as follows.

- **Recommendation rule:** A recommendation rule consists of rule name, two condition sets for matching each part of recommendation rules: recommendations and recommendation receivers, and a property of the receiver class for receiving the recommendations.
- **Condition set:** A condition set consists of condition set name, matching class, a class of the individuals to be matched, and conditions.
- **Condition:** A condition for matching individuals. It consists of a property chain, an operator, and a value object.
- **Property Chain:** A property chain is an ordered series of one or more object-type properties. Each property links to an individual of the object of a triple recursively like joining many tables in relational database.

- **Operator:** An operator for matching or comparing between the condition's value object and a triple's object. The operators supported are comparison operator (=, >, <, >=, <=, contains, rdf: type)
- **Object:** Object of a condition is in three types: literal value, URI, and property chain node value. A literal value is an RDF literal with a specified data-type. A property chain node is a node that refers to value of another property chain.

To exemplify the format of the rule template, we make an example of a recommendation "Recommend the products which have discount rate more than 10%, and have price more than 10 dollars but less than 15 dollars, to the customers which have bought products from the store more than 20 times".

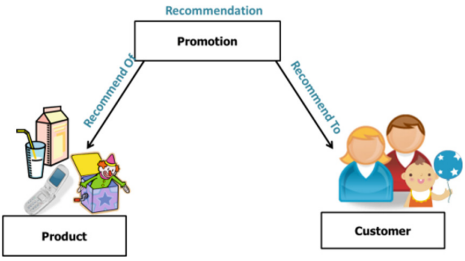


Fig. 2. Components of the OAM recommendation template

```
"ClearancePromotion": {
  recOfRule: "DiscountedProduct",
  recToRule: "FrequentCustomer",
  recProperty: "suggestedProducts" }
```

Fig. 3. Elements of recommendation rule and linked condition sets

```
"DiscountedProduct": {
matchingClass: "http://ex.org/Product", conditions: [
{ propertyChain: ["http://ex.org/discountRate"], operator: ">", object: ["float","10.00"]},
{ propertyChain: ["http://ex.org/price"], operator: ">", object: ["float", "10.00"]},
{ propertyChain: ["http://ex.org/price"], operator: "<", object: ["float", "15.00"]]}

"FrequentCustomer": {
matchingClass: "http://ex.org/Customer",
conditions: [{ propertyChain: ["http://ex.org/boughtRecord", "http://ex.org/boughtTimes"],
operator: ">", object: ["int", "20"]]}
```

Fig. 4. Example of matching condition sets

Figure 3 illustrates elements of the recommendation rule and linked condition sets in JSON format. Condition sets are illustrated in Fig. 4. Rules in decision tables can be translated to this template form. The recommendation rule template can be serialized to a rule language to be processed by rule-based inference engines.

4 Case Study: Excise Duty Recommender System

This section describes a case study of implementing the framework in an excise duty recommender system developed for the Thailand Excise Department. Excise duty is a type of tax charged on certain products and services that are considered unnecessary or luxurious. The Excise duty recommender system [3] aims to reduce users' effort in product-to-duty classification task and reduce errors in tax payment. The framework was adopted in this system to simplify management of recommendation rules for different excise products. The spreadsheet approach has supported bulk rule creation, rule updating and maintenance and ruleset management.

4.1 System Architecture

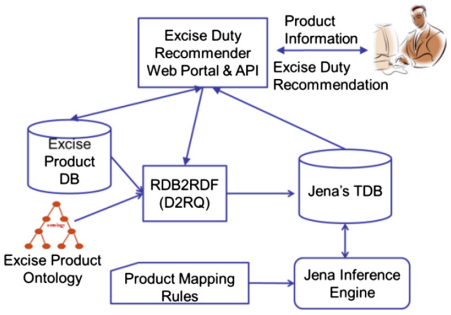


Fig. 5. System architecture of the excise duty recommender system

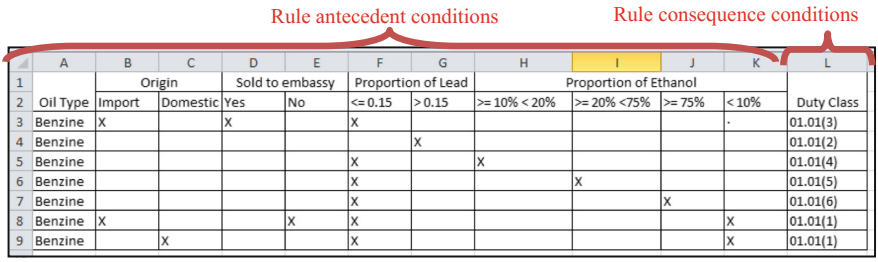
The excise duty recommender system was developed using the Semantic Web-based system architecture. Figure 5 shows the system architecture of the excise duty recommender system. The system was implemented using the Ontology Application Management (OAM) framework. Information about the user' products was obtained from the Web-based data input form, which were then stored in an excise product database. The OAM framework facilitated mapping between the excise product database and the excise product ontology. The framework adopted the D2RQ system [1] in creating RDF data from relational database. The RDF data were stored in the Jena's TDB triple store and were processed by Jena inference engine based on the product-to-duty recommendation rules. The recommendation results were stored in the triple store and can be retrieved using the OAM API.

4.2 Recommendation Rule Management in Spreadsheets

The ontology-based rule management framework using spreadsheets was adopted in this system to simplify management of recommendation rules for different excise products. There is the total of 22 excise products. Each product differs in terms of the number of product properties and the number of rules. Overall, there are over 70

product properties and over 250 recommendation rules in total. This section exemplifies the rule management process in this project.

4.2.1 Creating Decision Tables



	A	B	C	D	E	F	G	H	I	J	K	L
1		Origin		Sold to embassy		Proportion of Lead		Proportion of Ethanol				
2	Oil Type	Import	Domestic	Yes	No	<= 0.15	> 0.15	>= 10% < 20%	>= 20% <75%	>= 75%	< 10%	Duty Class
3	Benzine	X		X		X					·	01.01(3)
4	Benzine						X					01.01(2)
5	Benzine					X		X				01.01(4)
6	Benzine					X			X			01.01(5)
7	Benzine					X				X		01.01(6)
8	Benzine	X			X	X					X	01.01(1)
9	Benzine		X			X					X	01.01(1)

Fig. 6. Example of creating product-to-duty recommendation rules for petroleum products

Rules for each product are stored and maintained in each spreadsheet file. Example of a decision table for the petroleum products is shown in Fig. 6. In this example, the duty class of a petroleum product is determined by the type, origin, selling purpose, proportion of Lead and Ethanol of the product. An example rule may be defined as when a petroleum product has the Benzine type and contains the proportion of Lead of less than 15% and the proportion of Ethanol between 10–20%, the product is classified into the '01.01(4)' duty class.

4.2.2 Mapping Decision Tables with Ontology

Figure 7 exemplifies mapping rule conditions in spreadsheet with the Petroleum product ontology. A configuration user interface allows mapping spreadsheet criteria conditions with properties and classes in the ontology. In this project, the rule creation and condition-to-ontology mapping tasks were performed by knowledge engineers.

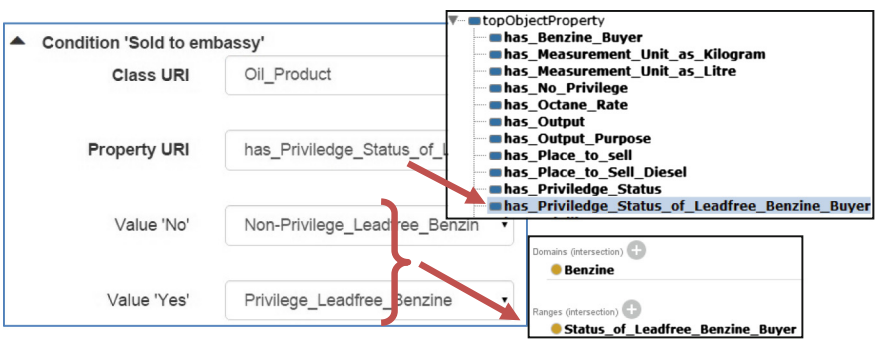


Fig. 7. Example of mapping spreadsheet rule conditions with the petroleum product ontology

4.2.3 Rule Template Generation and Rule Syntax Serialization

Figure 8 shows an output of the rule translation system which translates the decision table to rules in the template format and subsequently in Jena's syntax. All rules in the decision tables were successfully transformed and used by the excise duty recommender system.

Fig. 8. Example of rules in the template format and in Jena's syntax for petroleum products

5 Conclusion

This paper describes an ontology-based approach to support rule management by knowledge engineers and experts using spreadsheets. Our approach allows the users to use their own vocabulary and existing tool, i.e. spreadsheet application, in managing recommendation rules. Our work differs from existing systems in that it is designed for the Semantic Web data and uses a flexible recommendation rule template. Design of the rule template promotes interoperability with different rule-based reasoners that adopt the same ontology. The framework was successfully validated in a large-scale project in developing an excise duty recommender system, where bulk rule creation and update support are required.

Our future work aims to conduct a usability study and to assess accuracy of rule management performed by users with different background. Other supported functions such as rule consistency checking and rule compression are also planned.

Acknowledgments. This work is supported by the Excise Department and the National Science and Technology Development Agency (NSTDA), Thailand.

References

1. Bizer, C., Seaborne, A.: D2RQ-treating non-RDF databases as virtual RDF graphs. In: Poster at the 3rd International Semantic Web Conference (ISWC2004) (2004)
2. Buranarach, M., et al.: OAM: an ontology application management framework for simplifying ontology-based semantic web application development. Int. J. Softw. Eng. Knowl. Eng. **26**(01), 115–145 (2016)
3. Buranarach, M., Ruangrajitpakorn, T., Anutariya, C., Wuwongse, V.: Ontology design approaches for development of an excise duty recommender system. In: Kawtrakul, A., Laurent, D., Spyratos, N., Tanaka, Y. (eds.) Information Search, Integration, and Personalization. CCIS, vol. 421, pp. 119–127. Springer International Publishing (2014)
4. Dimitrova, V., Denaux, R., Hart, G., Dolbear, C., Holt, I., Cohn, A.: Involving domain experts in authoring OWL ontologies. In: Sheth, A., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) The Semantic Web—ISWC 2008. LNCS, vol. 5318, pp. 1–16. Springer, Berlin, Heidelberg (2008)
5. Giarratano, J., Riley, G.: Expert Systems: Principles and Programming. Course Technology, Boston, MA (2004)
6. Håkansson, A., Widmark, A., Edman, A.: The KANAL-system: an intelligent rule editor supporting the knowledge acquisition process. In: Swedish Artificial Intelligence Society Annual Workshop 2000 (2000)
7. Kaczor, K., Nalepa, G., Łysik, Ł., Kluza, K.: Visual design of drools rule bases using the XTT2 method. In: Katarzyniak, R., Chiu, T.-F., Hong, C.-F., Nguyen, N. (eds.) Semantic Methods for Knowledge Management and Communication. SCI, vol. 381, pp. 57–66. Springer, Berlin, Heidelberg (2011)
8. Kaljurand, K.: ACE view—an ontology and rule editor based on attempto controlled english. In: Dolbear, C., Ruttenberg, A., Sattler, U. (eds.) OWLED. CEUR-WS.org (2008)
9. Nowak, M., Bak, J., Jedrzejek, C.: Graph-based rule editor. In: Aït-Kaci, H., Hu, Y.-J, Nalepa, G.J., Palmirani, M., Roman, D. (eds.) Proceedings of the RuleML2012@ECAI Challenge, at the 6th International Symposium on Rules. CEUR-WS.org (2012)
10. OpenRules: Open Rules: Rule Editors. http://openrules.com/ruleeditors.htm
11. Regier, R., Gurjar, R., Rocha, R.A.: A clinical rule editor in an electronic medical record setting: development, design, and implementation. In: AMIA Annual Symposium Proceedings. pp. 537–541 (2009)

Factuality Classification Using Multi-facets Based on Elementary Discourse Units for News Articles

Khaing Swe Wynn[1(✉)] and Sasiporn Usanavasin[2]

[1] School of Information, Computer and Communication Technology (ICT), Thammasat University, Khlong Nung, Thailand
khaineshwewynn@gmail.com
[2] Sirindhorn International Institute of Technology (SIIT), Thammasat University, Khlong Nung, Thailand
sasiporn.us@siit.tu.ac.th

Abstract. Factuality classification is used for classifying information based on degrees of certainty. It has been actively used in different applications including information extraction, textual entailment, finding semantic uncertainty and certainty, or fact extraction. In this paper, we propose an approach to improve factuality classification by analyzing information in Elementary Discourse Units (EDUs) and their relations. We use news articles as our case study since it contains information that has various degrees of certainty or factuality values (i.e., information about certain events or uncertain information from factual and opinionated information). In this work, we use five sets of facets for factuality classification, which are (1) Epistemic Modality set, (2) Subjectivity Type set, (3) Rhetorical Structure Theory (RST) set, (4) Semantic Implicative and Factive Patterns set and (5) Weasel Words set. Unlike previous works on factuality classification, we use multiple facets of EDU to examine certainty and unambiguity level of information. We performed experiments based on news articles in FactBank corpus. We evaluated our method by comparing with several state-of-the-art factuality classification techniques and the results clearly show that our method can improve accuracy in terms of precision, recall and F1-measure as 94.1%, 93.9% and 93.9%, respectively.

Keywords: Factuality classification · Subjectivity · Implicative and factive patterns · Weasel words

1 Introduction

A huge volume of news articles in various domains present information about situations of events in the real world. They present both factual and opinionated information about those events that happened, are happening and are going to

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_8

happen in the real world. Since all we perceive come from news articles, these articles are increasingly becoming very useful among information seekers to know what are happening in the real world. Also, online news articles enable fact seekers to find facts in which events actually happened in the reality and which statements are assured.

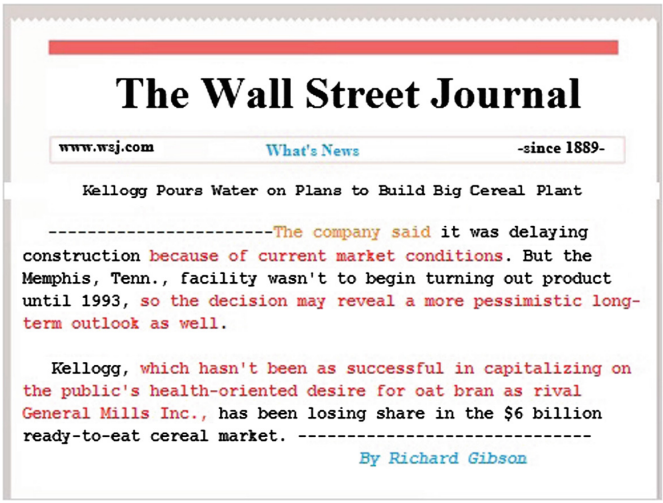
There are various and growing information about events in different sources (e.g., news articles, web blogs, posts, etc.) and a lot of such information in news documents has different certainty levels. Thus, factuality classification of event information becomes essential as one of the techniques in exploring different certain information levels. Regardless of sources trustworthy, different kinds of information levels that readers perceive come from authors' or sources' committed belief and assertion about propositions or events. The different levels of information are certainty(CT), probability(PR), possibility(PS) and underspecification(Uu) about propositions or events. Traditional text analysis and epistemic modality analysis techniques cannot be directly applied to the event information in online news articles and thus need to be adopted to address the factuality of event information because the writers or speakers present information about events using rhetorical relations among discourses.

When a lot of event information is presented in news articles, many writers report about events in a form of using multiple discourse units and relating those units into one sentence. Describing events by using ambiguous or vague words in news articles may reduce facuality degrees of event. Therefore, investigating factuality levels of events using lexical and syntactic based approaches may not be sufficient. We study that EDU level based facets such as rhetorical relations among EDUs, weasel words in EDUs, lexical semantics of subjectivity types can help to tackle this issue. For example, the sentence "The construction would resume when market conditions warrant it"contains a temporal relation between the nucleus of sentence ("The construction would resume") and its satellite ("when market conditions warrant it"). The presence of this relation helps to classify semantically factuality information about events "resume" and "warrant". Based on rhetorical relation between two EDUs and their contents in each EDU, we classify semantic factuality of each event in EDUs.

In this paper, we propose a factuality classification using multi-facets approach that is performed based on elementary discourse units (EDUs) extracted from a news article. We use news articles from FactBank corpus [23] as our experiments. Fig. 1 shows part of news from the Wall Street Journal. Table 1 shows a list of EDUs that are taken from the news article as shown in Fig. 1. We combine epistemic modality with semantic factive and implicative patterns, weasel clues, subjectivity types, and rhetorical relation facets for factuality classification to enhance semantic analysis of factual information.

2 Related Works

In recent years, there has been increasing focus on factuality analysis that is one of the classification techniques of information level about certainty or uncertainty


The Wall Street Journal

www.wsj.com What's News -since 1889-

Kellogg Pours Water on Plans to Build Big Cereal Plant

----------The company said it was delaying construction because of current market conditions. But the Memphis, Tenn., facility wasn't to begin turning out product until 1993, so the decision may reveal a more pessimistic long-term outlook as well.

Kellogg, which hasn't been as successful in capitalizing on the public's health-oriented desire for oat bran as rival General Mills Inc., has been losing share in the $6 billion ready-to-eat cereal market. ----------

By Richard Gibson

Fig. 1. News of the Wall Street Journal

Table 1. List of EDUs for news of Wall Street Journal

EDU no	Sent:	EDU	Factuality values
E1	S1	The company said	CT
E2	S1	It was delaying construction	CT
E3	S1	Because of current market conditions	CT
E4	S2	But the Memphis Tenn. Facility wasn't to begin...until 1993	Uu
E5	S2	So the decision may reveal..... as well	PS
E6	S3	Which hasn't been as successful.....Mills Inc	CT
E7	S3	Kellogg has been losing share.....cereal market	CT

of events. Factuality on events depends on the certain information of events and many previous systems classified factuality values or degrees of certainty by means of linguistic structure or interaction of linguistic cues.

We first provide a brief survey of previous works on factuality classification in different domains. Factuality classification is of the utmost important technique to distinguish factual events from non-factual events to improve information extraction. The first seminal works [2,17] related to committed belief

tagging annotated propositions with committed belief, uncommitted belief and non-applicable. Sauri and Pustejovsky [23] implemented FactBank corpus annotated with event factuality and built De-Facto system [21,22,24] to classify events into certainty, possibility, probability and under-specified information by exploiting polarity, epistemic modality, event indexing predicates and syntactic constructions and evaluated the system's performance with gold standard baseline method. Marneffe and Manning [11,12] explored veridicality assessment on events by applying context and world knowledge and extended FactBank corpus. Vlachos and Riedel [28] defined task and constructed dataset for fact checking by finding the assessment of the truthfulness of a claim.

Secondly, we focused on surveying about speculations and hedge cues of events to detect certainty and uncertainty. Detecting speculation or hedge cues is the utmost essential technique to distinguish semantic uncertainty from certainty for factuality classification. Previous researches detected speculative and hedge cues by exploiting different means of machine learning techniques. Medlock and Briscoe [13,14] identified hedges in biomedical literature using semi-supervised machine learning approach. Agarwal and Yu [1] detected not only hedge cues but also their boundary in biomedical text to extract uncertain events.Kilicoglu and Bergler [4] proposed a method to recognize speculative language using lexical surface and syntactic patterns. Their CRF-based model, a supervised machine learning algorithm, performed statistically better than the baseline systems. Moncecchi and Minel [15] improved detection of speculative cues by using linguistic knowledge.

Next, other works have performed detection of semantic uncertainty and certainty by applying different machine learning approaches. Li and Gao [9] performed semantic uncertainty detection by learning hedge cues using hidden markov model and compared with prior results on the CoNLL-2010 Shared Task 1 using Wikipedia and biomedical datasets. Their systems outperformed the best result of the CoNLL-2010 systems. Szarvas [26] focused on semantic uncertainty detection model across different domains and genres. The differences between [26] and other works are that they performed a uniform sub-categorization of semantic uncertainties and applied domain adaptation techniques between target and source domains to adapt a new domain based on it. For detection of certainty or assertion, Velupillai [27] developed a lexicon for pyConTextSwe in clinical texts and integrated hedge cues from four lexicons. They classified four assertion classes and binary classes to pyConTextSwe and obtained promising results for all classes by using final developed lexicon.

Our work is different from these previous works by adding advanced facets with traditional facets to boost certainty and unambiguity of events. Our work explores semantic uncertainty and ambiguity by using rhetorical relation among EDUs, subjectivity types, semantic implicative and factive patterns, and weasel words facets combined with epistemic modality facet.

3 Multi-facets for EDU Level Factuality Classification

We classify event factuality into four classes based on certainty levels which are certainty(CT), probablity(PR), possibility(PS), and under-specification(Uu). We define certainty that events are absolutely certain that they happened or are happening, probability that events will happen in the future and they have moderate certainty, possibility that events may happen or exist and they are low in certainty level, and under-specification that events are lack of knowledge or lack of clarity and so they are uncertainty. We take into account the following traditional and advanced facets to build our classifier.

1. **Epistemic Modality Set:** Based on previous works [16,19,20], we use the epistemic modality set as a ground truth since epistemic modality is related to degree of certainty or factual status of event information. Epistemic modality has composed of four facets: modality adjectives, adverbs, verbs, and modality type. These are binary attributes based on the appearance of modality verbs, adjectives, adverbs and four attributes for modality type in EDUs which are certainty(CT), probablity(PR), possibility(PS) and under-specification(Uu). We denote modality type based on modality verbs, adjectives or verbs.
2. **Subjectivity Type Set:** The subjectivity set is a good indicator for factuality classification and it is composed of a subjectivity type facet in our study. For this set, we encode subjectivity or non-subjectivity types of events by checking subjective clues [29]. Subjectivity(Sub) types are private states [29,30] including speculations, beliefs, sentiment, evaluations, goals, judgments, knowledge, report, etc. whereas the rests of events are defined as non-subjectivity (Non-sub). Based on these subjective clues, we encode subjectivity types to learn factuality event or non-factuality event in EDUs.
3. **RST Set:** We use RST [10] that can split texts into segments (EDUs) that are rhetorically related to one another. In RST, each EDU can be either nucleus or satellite, and nucleus is more significant than satellite as it provides main information with respect to understanding and interpreting the text. We use Sentence level PArsing of DiscoursE (SPADE) tool [25] that provides RST trees for individual sentences and 14 rhetorical relations for EDUs. They are 1. Attribution, 2. Backgroud, 3. Cause, 4. Comparison, 5. Condition, 6. Contrast, 7. Consequence, 8. Elaboration, 9. Enablement, 10. Evaluation, 11. Embedded Elaboration, 12. Explanation, 13. Joint, and 14. Temporal. The default relation for nucleus to satellite is Span. In RST set, we take into account eight facets in our study (see Table 2) that encode facet values for appearance of one type of RST relations among EDUs in a given sentence.
4. **Weasel Words Set:** Weasel words [3] are assumed to be unsupported attribution by Wikipedia because weasel words offer the vague or ambiguous claim for the event and they violate the factual information into ambiguous information. Weasel words could be divided into three categories:
   - Numerically vague expressions (e.g. “some people, experts, many, etc.”)
   - Use of the passive voice to avoid specifying an authority (e.g. “it is said, it is regarded, etc.”)

Table 2. List of five sets and their facets

Set	Facets	Types of Facet value
Epistemic modality	Modality type	CT, PR, PS, Uu
	Modality adjectives	Binary
	Modality adverbs	Binary
	Modality verbs	Binary
Subjectivity	Subjectivity type	Sub types and Non-sub
RST	IsNucluesorSatellite	Binary
	RST relation type	14 relations in RST set
	TempNucluesIsNotCT	Binary
	TempNucluesSubIsIntenorExp	Binary
	AttrSatelliteBeliefExpGoalIntenSpec	Binary
	AttrSatelliteModalityCT	Binary
	ExpNucluesisUnCT	Binary
	CondSatellite	Binary
SIFP	IsImplicativeorFactive	Binary
	haveAncestorComplement	Binary
	haveAncestorComplementImpl	Binary
	SyntacticCategory	8 patterns in SIFP set
	SemanticMeaning	6 types in SIFP set
Weasel words	Number of weasel words	Integer
	Normalized weasel words	Normalized values between 0 and 1

- Adverbs that weaken (e.g. "often, probably, etc.")

We use two facets for weasel words in our work that are the number of weasel words and its normalized value that appear in each EDU.

5. **Semantic Implicative and Factive Pattern (SIFP) Set:** Factive constructions involve presuppositions and implicative constructions [5–8] involve entailment about the factuality of the event and presuppositions. Semantically, factive and implicative patterns can be concluded into one statement and they can be used to determine factuality values of events. In case these two facets do not cover in EDUs, the default values for patterns and meanings will be set to null. For example, (a) has factive verb "acknowledged" and it presupposes that the report is correct, and (b) has implicative verb "forgot" and it entails that she did not reschedule the meeting and presupposes that she intended to reschedule it.

(a) They acknowledged the report to be *correct*.
(b) Kim forgot to *reschedule* the meeting.

In our work, we encode the binary values for appearance of implicative and factive construction, and its related two facets in each EDU, their syntactic categories, and their semantic meanings to highlight the factuality of events.We exploited eight syntactic patterns and six semantic meanings for implicative and factive patterns according to [8].

Table 2 summarizes lists of five sets and their facets for factuality classification and Table 3 shows the experimental results of multi-facets for EDUs in Table 1.

4 Experiment and Design

In order to test EDU level factuality classification, a collection of labeled events from FactBank [23] is applied in our work. The FactBank corpus contains news from different sources and is built on TimeBank [18]. The annotations include both respective sources and factuality classification about the events. In our work, we ignore the role of sources and we only emphasize factuality of events based on content properties in EDUs. We focus on four levels of factuality information: certainty, probability, possibility, and under-specified classes. For EDU segmentation, we applied SPADE tool [25] and 14 relations among EDUs. The rhetorical relation among sentences is out of our scope and we emphasize only discourse relations in each sentence. Based on epistemic modality, we constructed ground truth for EDU level factuality classification: each unit is examined and classified as probable, possible or under-specified if the unit has at least modality adjectives, adverbs, or verbs. As a result of this process, our ground truth is composed of 1759 units: 1461 units judged as CT, 176 units as PR, 48 units as PS, and 74 units as Uu.

4.1 Comparison with Baseline Approach

We evaluate the performance of our classification approach with standard baseline methods. For baseline facet, we use epistemic modality as a ground truth for EDU Level factuality classification. In this work, we exploit one rule (OneR) factuality classifier as our baseline method. We performed experiments by using all of our proposed content based facets such as epistemic modality combined with subjectivity types, RST relations, semantic categorization of implicative and factive patterns, and weasel words.

4.2 Experimental Setting

We used Weka data mining toolkit with default settings to conduct our experiments. We conducted various supervised learning algorithms to perform our classification experiments. We experimented with Naive Bayes, Support Vector Machines(SVM), Logistic regression, Bagging, Boosting and Decision Trees. Based on our proposed features, the logistic regression outperformed all the

Table 3. Experiment results of multi-facets for EDUs in Table 1. The RST relation types are abbreviated as follows: Attribution (Attr), Explanation (Exp), Evaluation (Eva), Embedded Elaboration (Embed) and Span (Span)

Facets	EDU						
	E1	E2	E3	E4	E5	E6	E7
Modality type	CT	CT	CT	Uu	PS	CT	CT
Modality adjectives	0	0	0	0	0	0	0
Modality adverbs	0	0	0	0	0	0	0
Modality verbs	0	0	0	1	1	0	0
Subjectivity type	Report	Non-sub	Non-sub	Non-sub	Non-sub	Non-sub	Non-sub
IsNucluesorSatellite	0	1	0	1	0	0	1
RST relation type	Attr	Span	Exp	Span	Eva	Embed	Span
TempNucluesIsNotCT	0	0	0	0	0	0	0
TempNucluesSubIsIntenorExp	0	0	0	0	0	0	0
AttrSatelliteBeliefExpGoalIntenSpec	0	0	0	0	0	0	0
AttrSatelliteModalityCT	0	0	0	0	0	0	0
ExpNucluesisUnCT	0	0	0	0	0	0	0
CondSatellite	0	0	0	0	0	0	0
IsImplicativeorFactive	0	0	0	0	0	0	0
HaveAncestorComplement	0	0	0	0	0	0	0
HaveAncestorComplementImpl	0	0	0	0	0	0	0
SyntacticCategory	null	null	null	null	null	null	null
SemanticMeaning	null	null	null	null	null	null	null
Number of Weasel Words	0	0	0	0	0	0	0
Normalized Weasel words	0	0	0	0	0	0	0

other classifiers. To evaluate the performance of our classifiers, we used weighted-average precision, recall and F1-measure. We set 10-fold cross validation in our experiments. We compared precision, recall and F1-measure with performance of the baseline method. The ground truth that classifies all units in the majority class has a weighted-average precision, recall and F1-measure of 86.6%, 86.7% and 84.7% respectively for FactBank Corpus.

4.3 Classification Results

Table 4 reports the results of factuality classification obtained from different combination of facets with epistemic modality. We conducted a total of 20 experiments based on five facet groups combined for data set. We denote epistemic modality as Mod, subjectivity type as Sub, RST as Rhe, semantic implicative and factive pattern as ImpFact and weasel words as Wea. From the results, we note that all facets set, Mod+Subj+Rhe+ImpFact set, and Mod+Subj+Rhe set perform better than other facets.

Table 4. Classification performance of different combined facet sets. Numbers in bold correspond to the best performance

No	Combination of facet sets	Precision	Recall	F1
1	**Mod (Ground Truth)**	**0.866**	**0.867**	**0.847**
2	Sub	0.74	0.761	0.686
3	Rhe	0.752	0.792	0.738
4	ImpFact	0.608	0.734	0.63
5	Wea	0.618	0.734	0.632
6	Mod+Subj	0.898	0.899	0.883
7	**Mod+Rhe**	**0.899**	**0.902**	**0.896**
8	Mod+ImpFact	0.868	0.874	0.858
9	Mod+Wea	0.869	0.873	0.858
10	**Mod+Sub+Rhe**	**0.93**	**0.93**	**0.928**
11	Mod+Sub+ImpFact	0.901	0.903	0.891
12	Mod+Sub+Wea	0.899	0.903	0.891
13	Mod+Rhe+ImpFact	0.905	0.909	0.905
14	Mod+Rhe+Wea	0.905	0.908	0.904
15	Mod+ImpFact+Wea	0.873	0.879	0.867
16	Mod+Sub+Rhe+ImpFact	**0.938**	**0.936**	**0.935**
17	Mod+Sub+Rhe+Wea	0.937	**0.936**	**0.935**
18	Mod+Rhe+ImpFact+Wea	0.911	0.915	0.911
19	Mod+Sub+ImpFact+Wea	0.901	0.906	0.896
20	**Mod+Sub+Rhe+ImpFact+Wea**	**0.941**	**0.939**	**0.939**

We perform classification experiment using one combination of facet sets at a time. For different unary facets, the ground truth epistemic modality is a strong indicator of its factuality classification. RST relation is the second best performing facet outperforming the subjectivity type, weasel words, and semantic implicative patterns. Moreover, weasel words show better results than the use of semantic factive and implicative patterns. This does not mean that semantic factive and implicative pattern is not a good indicator of factuality. This means that in case we experiment many data sets that use semantic factive and implicative patterns, we can conclude that those patterns can be a good indicator of factuality.

Moreover, we performed preliminary analysis and it shows that RST relations are good indicators of factuality. Among 14 types of EDU relations, we discover that 19% of all EDUs have attribution relation and 90% of those attribution relations EDUs are classified as certainty whereas 7% is probable and possible and 3% is under-specified. We discover that attribution relation is dependent relation between satellite and nucleus. The second most significant relation in EDUs is embedded relation as the writers gave additional information for entities. In embedded relation, 78% of elaborating units has certainty level.

Among these relations, we note that condition, purpose and enablement relations totally give under-specified information to the events whereas some relations are marginal. In this example "Led by its oat-based Cheerios line, General Mills has gained an estimated 2% share so far this year", the satellite "Led by its oat based Cheerios line" is connected to the nucleus "General Mills has gained an estimated 2% share so far this year" with background relation. We investigate that background relation supports certainty information in all sentences and when these rhetorical relation facets are combined with other facet sets, the performance is improved. Moreover, subjectivity types show better performance than weasel words, and implicative and factive patterns.

Combining two facet sets based on modality called ground truth provides the better result in classification. Modality facet with rhetorical relation outperforms other binary facet sets. Similarly, combining of modality and subjectivity type can be judged as a good indicator for factuality. Next, we discover that Mod+Rhe+Sub performs better than any other ternary facet sets. Among 4-ary facet sets, Mod+Sub+Rhe+ImpFact set is the best attribute for factuality classification and it means that combination of epistemic modality, subjectivity type, rhetorical relation, and semantic implicative and factive patterns are useful indicators for factuality classification. Next, we analyze the classification using all of five facet sets and we observe that using five facet sets gives the best performance result for factuality classification.

We achieve overall accuracy of 93.01%, a precision of 94.1% and F1-measure of 93.9% on FactBank corpus. We further analyze the classification performance of our classifier by analyzing its performance for four classes. Table 5 reports precision, recall and F1-measure for four factuality classes. We study that the classification performance for certainty (CT) class is higher than others in the system. This can be concluded that there are more factual information than

Table 5. Classification performance of proposed model for four classes

	Precision	Recall	F1
CT	0.983	0.984	0.984
PR	0.809	0.911	0.857
PS	0.725	0.787	0.755
Uu	0.878	0.72	0.791
Weighted average	0.941	0.939	0.939

Table 6. Comparison with baseline approach

	$Macro - A$			$Micro - A$		
	Precision	Recall	F1	Precision	Recall	F1
Baseline (OneR) approach	0.705	0.86	0.775	0.867	0.867	0.867
Our proposed approach	0.85	0.848	0.848	0.939	0.939	0.939

non-factual information in news articles of FactBank corpus. In Table 6, we also study that classification performance of our proposed model outperforms the baseline method by means of macro-average and micro-average.

4.4 Facet Weights

We study the importance of individual facets by using information gain with respect to each class. The higher the weight value is, the more important it is to classify the factuality. Table 7 shows top 10 facets ranked by information gain ranking filter. According to the table, we can conclude that epistemic modality type is the strongest indicator to classify factuality of event. Epistemic modality is denoted based on modality adjectives, modality adverbs and modality verbs and this modality type is the most fundamental and important facet for factuality classification. The second most informative facet is RST relation which is also important to assess factuality class as all of the events in news articles are connected by using rhetorical relation and this type of relation becomes significant to classify the factuality of event.

The subjectivity type is ranked as the third position among 20 facets as this subjectivity type significantly classifies the factuality of events. For example, by analyzing the subjectivity type such as “belief and opinion”, “expectation”, “doubt”, “desire”, “goal”, etc. our proposed model vividly discriminate the events into non-factual classes. Another important facet in epistemic modality set is modality verb because it is frequently used to describe the events in news articles.

Table 7. List of top 10 facets with the highest w_i in the best classifier

Rank	w_i	Feature	Set
1	0.516	Modality type	Epistemic Modality
2	0.280	RST relation type	RST
3	0.173	Subjectivity type	Subjectivity
4	0.168	ModalityVerbs	Epistemic modality
5	0.047	IsNucleusorSatellite	RST
6	0.033	NumOfWeaselWords	Disambiguate
7	0.031	IsImplicativeorFactive	SIFP
8	0.028	NormalizedWeaselWords	Disambiguate
9	0.026	AttrSatelliteBeliefExpGoalIntenSpec	RST
10	0.01	TempNucluesIsNotCT	RST

As seen in Table 7, the facet types of Satellite and Nucleus are informative indicators for factuality classification of the whole data sets as related to rhetorical relation, checking EDUs as satellite or nucleus is very important to classify. Among top 10 facets out of 20, number of weasel words is ranked higher than semantic implicative and factive pattern facet. Next, related to RST relation and subjectivity type, checking whether attribution relation has subjectivity type of "belief and opinion", "intention", "speculation", "doubt", or "expectation", etc. is one of the promising informative facets. According to our experiment, we found that if the event in nucleus is not certain, the event in satellite is also determined as non-factual information. Therefore, the temporal relation in RST set can also be used to indicate factuality of events. Our results show that rhetorical relation, subjectivity type, and number of weasel words are good indicators for developing a new semantic factuality classifier to extract the facts in order to overcome the limitations of the lexical and syntactic techniques.

5 Conclusion and Future Works

In this paper, we proposed an approach for improving accuracy of factuality classification for news articles. In our approach, we use the supervised machine learning model for EDU levels and use multiple facets such as subjectivity types, RST, semantic categorization of implicative and factive patterns and weasel words combined with classical epistemic modality to analyze the degree of certainty of information. We evaluated our model by performing several experiments and had compared them with the standard supervised machine learning technique. The experimental results clearly show that by using the proposed facets, the accuracy of classification can be improved in terms of precision, recall and

F1-measure as 94.1%, 93.9% and 93.9%, respectively. For future work, we plan to explore other facets such as consistency and unambiguity among events in order to improve our classification technique.

Acknowledgement. This research was supported by the Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Intelligent Informatics and Service Innovation (IISI) and NRU grant at SIIT, Thammasat University.

References

1. Agarwal, S., Yu, H.: Detecting Hedge Cues and their scope in Biomedical Text with conditional random fields. J. of Bio. Info. **43**, 953–961 (2010)
2. Diab, M.T., Levin, L.: Committed belief annotation and tagging. In: Proceedings of the Third Linguistic Annotation Workshop, ACL-IJNLP 09 (2009)
3. Ganter, V., Strube, M.: Finding hedges by chasing weasels: hedge detection using wikipedia tags and shallow linguistic features. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pp.173-176 (2009)
4. Kilicoglu, H., Bergler, S.: Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. In: Current Trends in Biomedical Natural Language Processing, pp. 46–53, Ohio, USA (2008)
5. Karttunen, L.: Implicative verbs. Language **47**, 340–358 (1971)
6. Karttunen, L.: Simple and phrasal implicatives
7. Karttunen, L., Zaenen, A.: Veridicity. In: Annotating, Extracting and Reasoning about Time and Events (2005)
8. Karttunen, L.: Simple and phrasal implicatives. In: Proceedings of the Sixth International Workshop on Semantic Evaluation, pp. 124–131 (2012)
9. Li, X., Gao, W., Shavlik, J.W.: Detecting semantic uncertainty by learning hedge cues in sentences using an HMM. In: Proceedings of Workshop on Semantic Matching in Information Retrieval, pp. 30–37, Queensland, Australia (2014)
10. Mann, W.C., Thompson, S.A.: Rhetorical structure theory: toward a functional theory of text organization. Text **8**, 243–281 (1988)
11. Marneffe, M.C.D., Manning, C.D., Potts, C.: Veridicality and utterance understanding. In: Fifth IEEE International Conference on Semantic Computing, pp. 430–436 (2011)
12. Marneffe, M.C.D., Manning, C.D., Potts, C.: Did it happen? The pragmatic complexity of veridicality assessment. Com. Ling. **38**, 301–333 (2012)
13. Medlock, B., Briscoe, T.: Weakly supervised learning for hedge classification in scientific literature. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 992–999, Prague, Czech Republic (2007)
14. Medlock, B.: Exploring hedge identification in biomedical literature. J. of Bio. Info. **41**, 636–654 (2008)
15. Moncecchi, G., Minel, J.L., Wonserver, D.: Improving speculative language detection using linguistic knowledge. In: Proceedings of the ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics(EXProM-2012), pp. 37–46, Jeju, Republic of Korea (2012)
16. Morante, R., Sporleder, C.: Modality and negation: an introduction to the special issue. Com. Ling. **38**, 223–260 (2012)

17. Prabhakaran, V., Rambow, O., Diab, M.: Automatic committed belief tagging. In: Coling 2010: Poster Volume, pp. 1014–1022, Beijing, China (2010)
18. Pustejovsky, J., Hanks P., See, A., Sauri, R.: The timebank corpus. In: Proceedings of Corpus Linguistics, pp. 647–656 (2003)
19. Rubin, V.L.: Epistemic modality: from uncertainty to certainty in the context of information seeking as interactions with texts. Inf. Process. Manage. **46**, 533–540 (2010)
20. Sauri, R.: Determining modality and factuality for text entailment. In: Proceedings of the First IEEE International Conference on Semantic Computing, pp. 509–516. Irvine, California (2007)
21. Sauri, R., Pustejovsky, J.: From structure to interpretation: a double-layered annotation for event factuality, In:Prooceedings of the 2nd Linguistic Annotation Workshop. The Sixth International Conference on Language Resources and Evaluation, pp. 1–8 (2008)
22. Sauri, R.: A factuality profiler for eventualities in text. Ph. D. thesis, Brandeis University (2008)
23. Sauri, R., Pustejovsky, J.: FactBank : a corpus annotated with event factuality. Lang. Resour. Eval. **43**, 227–268 (2009)
24. Sauri, R., Pustejovsky, J.: Are you sure that this happened? Assessing the factuality degree of events in text. Comp. Ling. **38**, 261–299 (2012)
25. Soricut, R., Marcu, D.: Sentence level parsing using syntactic and lexical information. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, pp. 149–156 (2003)
26. Szarvas, G., Vincze, V., Farkas, R., Mora, G., Gurevych, I.: Cross-genre and cross domain detection of semantic uncertainty. Com. Ling. **38**, 335–368 (2012)
27. Velupillai, S., Skeppstedt, M., Kvist, M., Mwery, D.: Cue-based assertion classification for Swedish clinical text-developing a lexicon for pyConTextSwe. Artif. Intell. Med. **61**, 137–144 (2014)
28. Vlachos, A., Riedel, S.: Fact checking: task definition and dataset construction. In : ACL 2014 Workshop on Language Technologies and Computational Social Science, pp. 18–22 (2014)
29. Weibe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in Language. Lang. Resour. Eval. pp. 165–210 (2005)
30. Wiebe, J., Riloff, E.: Creating subjective and objective sentence classifiers from unannotated texts. In: CICLing05 Proceedings of the 6th international conference on Computational Linguistics and Intelligent Text Processing, pp. 486–497 (2005)

A Recommender of Transportation Planning in Campus Using Ontology

Suwit Somsuphaprungyos[1(✉)], Salin Boonbrahm[1], Poonpong Boonbrahm[1], and Taneth Ruangrajitpakorn[2,3]

[1] School of Informatics, Walailak University, Nakhon Si Thammarat, Thailand
{ss.rungyos, salil.boonbrahm, poonpong}@gmail.com
[2] Language and Semantic Technology Laboratory, Nation Electronics and Computer Technology Center (NECTEC), Pathum Thani, Thailand
taneth.rua@nectec.or.th
[3] Department of Computer Sciene, Thammasat University, Pathum Thani, Thailand

Abstract. This paper introduces intelligent services for efficient management in campus. Ontology is used as knowledge base in an expert system to yield reasonable solution. In a current state, a headcount of people in a building in map-based visualization and a recommendation of shuttle bus management according to student movement are developed. Instances of actual data from Walailak University were inputted for our test case scenario. The map-based visualization displays a headcount of students in buildings by their enrolled courses and a reservation of meeting room. A recommendation of shuttle bus management is a supporting service to calculate an amount of service shuttle busses and their route based on student movement. From the result of recommendation, the system returned a reasonable and viable solution for different scenarios.

Keywords: Smart campus · Intelligent service · Expert system
Visualization decision support system · Transportation plan

1 Introduction

Academic campus refers to the lecture theaters, laboratory, dormitories, football field, street and all other spaces in the university, college or school. A campus is a place where students, instructors, and academic employees gathered for academic and social activities. Large campus normally composed of many buildings which are designed for several activities. Activities on campus can be many, but the main purpose of it is for education activities. During study hours, students are gathered at many locations on campus and their destinations are different regarding to the courses they enrolled. Some may have to move from one building to another which may not in walking distance. On large campus, transportation service is the most important issue for facilitating students.

For smart campus, the issue of effective management and resource allocation method is the main concern. Knowing the number of people and their movement on campus should be well estimated for this purpose so that planning to facilitate all

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_9

activities can be done in advance. A sensor or hardware detector has been suggested in several researches [1–3] to monitor people on campus, but to install and maintenance such items is extremely costly especially for a large campus. Using that hardware is an effective and viable option, but not all campuses can afford such expenditure. Some researches [4, 5] suggest the use of Wi-Fi from cell phones for head counting, but this option is only available for the campuses which have Wi-Fi signal available for staffs, students and visitors. This approach has a good advantage on exploiting existing facilities on campus since registered Wi-Fi should nowadays be a common service in an academic area. However, the crucial downside is that the given headcount data is displayed only in real-time. Managing and allocating resource can hardly be arranged to serve a sudden change in number of people. These methods are good monitoring approach, but it needs to know the amount of people and their destination beforehand to plan an effective resource management.

To make it clear, let's take the student movement on campus as an example. For effective management for bus service, the headcount of students and relevant academic staffs in every area must be known. Their movement should be predictable accordingly to their class enrollment. Hence, the system is able to monitor each student's activities and such information can be used for other systems. Based on the course enrollment data, not only students' location and movement can be predicted, but also the instructors.

For this research, our main purpose is to create an infrastructure of a smart campus by focusing on a student enrolment data where we can get every class schedules including instructor for each class. An ontology tool is chosen to represent campus domain knowledge including person in campus, university facilities, activities in a university, and events. With ontology as knowledge base, relationships of things, people and actions are then defined semantically. The next step is developing visualizing information to indicate a student headcount in a campus building and a student movement regarding to their studying classes. Google map is applied in this work for better visualization. A service of transportation management and building management i.e. operating time and etc. using our implemented infrastructure are exemplified in this paper.

2 Background

There are many existing researches and works dedicating on creating a smart campus or smart university. Many definitions of a smart campus were given, but some do not match to this work intention as well as their scope. Hence, we review those researches and summarize them into Table 1.

From the comparison, we found that some explained their ideal plan to create a smart campus with latest technologies by enhancing the benefit of hardware exploitation [1, 6]. On the other hand, some applied rich information in a campus and implement a service based on it [5, 7]. The systems in the latter group are designed to serve as an information provider for user to manually analyze the information and make use of it.

Table 1. Related works in smart campus

Title	Description	Technology/tool	Capability
System architecture for a Smart University Building [1]	Detection from hardware and information integration by ontology	• Hardware • Wi-Fi, RFID, ZigBee, home automatic network, power line communication • Data technology • Ontology, semantic web	• Detect activity in campus • Collect data using ontology schema
A threshold based context change detection in pervasive environments: application to a smart campus [4]	Detection from hardware and information integration by application on server	• Wi-Fi • Sensor • Seamless connectivity server	• Integrate information from different devices
Ambient intelligence applications: introducing the campus framework [5]	Detection from hardware and information integration by ambient intelligence and multi agent system	• Hardware • Wi-Fi, SOAP, WSDL • Data technology • Ontology, agent communication language	• Integrate information from different devices
A discussion on the framework of smarter campus [6]	Cloud based digital content in learning and service in campus	• Hardware • IP v6 protocol, RFID in campus ID card, Wi-Fi • Data technology • Could computing plate-form	• Allow learning from cloud • Store data on cloud • Make use of hardware for service access
An ontology based framework for modeling movement on a smart campus [7]	Creating a model of people movement in campus	• Ontology • GIS	• Generate a visualizing model to display people movement

Both methods have their own advantage and profit in implementation. However, we prefer the approach of using existing information for creating an intelligent system since we aware that some universities cannot afford the investment and maintenance cost in practical. In this work, we introduce an intelligent system using general and accessible information of the campus. The system is designed to serve as a supportive decision making system using ontology-based reasoning to assist university staffs without applying an expensive hardware.

3 System Implementation

This work presents a framework of smart campus solely using information in university. The aim of this work is to utilize the information and creates a useful and effective solution to assist in campus management according to given environment and specification of each university. Ontology was selected as a system knowledge base since it has a potential in storing semantic meaning of the relevant concepts and less ambiguous machine understandable schema. In the processing layer, an inference engine is exploited as a crucial reasoner to return an intelligent solution according to technical and human logic. On top of the processing layer, applications such as visualization and reminder system are implemented for end-users' comfort usage. The overview of the proposed system is illustrated in Fig. 1.

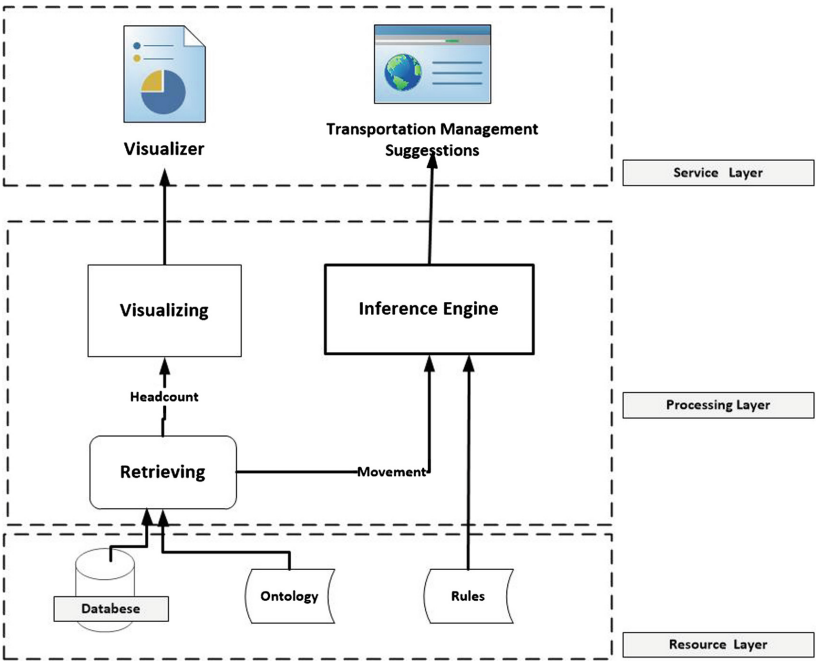


Fig. 1. An overview of the proposed framework

3.1 Campus Ontology

A campus ontology is designed to cover the concepts relevant to a university campus. This includes person, things, location, and function of those concepts in campus. This ontology is a manually developed heavy-weight ontology [8] regarding the concepts' semantic and their relation following Noy's Ontology Development guideline [9]. However, Hozo Ontology Editor [10] is chosen as a development tool because of its easy to understand and approve graphical representation among ontology developers.

In this work, some parts of the designed ontology are applied since the task of the presented system is a first phase of the smart campus project. Therefore, an explanation of the relevant concepts to this work is given in this paper, not the whole ontology. Some major concepts are demonstrated in Fig. 2.

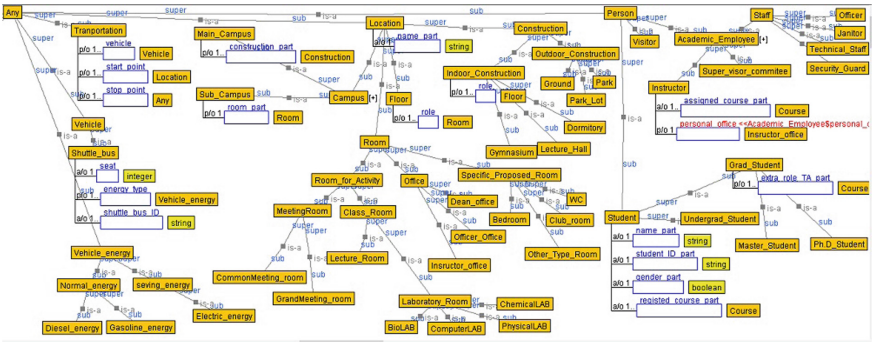


Fig. 2. Campus ontology

The major concepts of this ontology are construction type in campus, for example buildings, rooms, grounds, etc. People in campus also have a specific concept to fit their role and details. Another vital concept is things in campus such as instruments and available service vehicles. On the design, each concept was carefully crafted with rigorous meaning and semantically rich relations among them as we expect this ontology to be a knowledge representation of a smart campus conceptualization. In this ontology, the main class depends on a focused class that a smart service will use.

3.2 Instantiation

Data in this work are enrolment data which included student data, courses they enrolled, and class schedule, and data on university buildings and facilities. These data are slightly big and open data from a university database. In this work, OAM framework (Ontology Application Management) [11] is exploited to map the instance data to a campus ontology class. It has a visualizing interface for selecting data field in a database to one of uploaded ontology class or property. In the current state, data from Walailak University have been inputted and its students from three schools have been mapped in the system. The detail of the instance is shown in Table 2.

Table 2. Current instances mapped to campus ontology (accessed on 16th July 2015)

University	Faculty	Number of course	Number of student
Walailak University	School of Informatics	134	736
	School of Nursing	14	299
	School of Management	134	792

3.3 Visualization System for Decision Support

By the enrolment data, a summary of students' location based on day and time can be known. The query can also be done with the whole students or selected individuals. Aside from courses and their registered students, the room reservations (in case there is) can be selected to add up for query to show a more practical headcount of people in a building. A room reservation, such as instructor meeting and special group discussion, made by the university intranet system can only be counted toward the headcount since the system requires explicit information to calculate, not the uninformed usage of the room.

A visualization map was powered by Google Map API [12] since the ability of it to zoom-in or zoom-out has the usage in our tool. In zoom out visual, the sum of the area will be calculated to represent the headcount while headcount detail for each floor of the building will be displayed in a zoom-in mode. A captured UIs shown locations of a hundred randomized students is exemplified in Figs. 3 and 4.

Fig. 3. A result from a visualizing tool for headcount in campus from 100 students

Fig. 4. A zoomed version of Fig. 3 focusing on north building where most students locate

Figure 3 showed map with the random number of students from a School of Informatics and School of Nursing on Monday at 8–10 O'clock. From the map, the headcount (indicated in a balloon) shows that the majority of student (thirty-three

percentages of all students) has a class in Tai-Buri Building (north building in Fig. 3) while some scatter in other building. Note that fifteen students from all random students are not shown in the map because they do not have any class at the time and there is no room reservation at such time of the day. Figure 4 is the zoomed version of the Tai-Buri Building (north building in Fig. 3) where thirty-three students are plotted and extra eight headcounts from a room reserved for meeting. The headcount is given in more detail to each floor of the building.

3.4 Transportation Management Recommendation System

According to the students' movement, the system can recommend a plan for transportation within a campus by utilizing an inference engine and given preset rules as reasoner. Associated with the designed ontology and assigned instances, system returns the plan regarding to student headcount and their next class room. At first, the initial rules designed for cost efficiency and population sufficiency are installed as system default, but the additional rules are freely designable to match the campus policy and have the option to overrule the default rules. The production rule is composed with two components, triggering side and execution side, in a form as follow:

$$\texttt{IF} < \texttt{condition(s)} > \texttt{THEN} < \texttt{action} >$$

The triggering side is a condition(s) to match a query from users while the execution will be conducted in case the condition(s) are met. Sets of rule are separated according to a different service. As for the transportation management, classes in ontology involved in decision making include concepts of person amount in building, movement of persons, building location, and service vehicle. The designed rules in a pseudo description (for easy understanding) are demonstrated in Fig. 5.

```
IF <Person has Movement from Location#1 to Location#2>
      THEN <Service Vehicle Status = activate> AND <number of Service Vehicle =
(Person Amount ÷ Service Vehicle Seat)> AND <Service Vehicle has Initial Point =
Location#1 and Service Vehicle has Destination Point = Location#2>

IF <Service Vehicle Initial Point to Service Vehicle Destination Point#2 is a
SUBSET Service Vehicle Initial Point to Service Vehicle Destination Point#1> OR
<Service Vehicle Initial Point to Service Vehicle Destination Point#2 INTERSECT
Service Vehicle Initial Point to Service Vehicle Destination Point#1> AND <Service
Vehicle#1 has unused SEAT ≥ Person Amount at Service Vehicle Initial Point#2>
      THEN <Service Vehicle#2 Status = inactivate> AND <Service Vehicle#1 in use
for delivery for Service Vehicle Initial Point to Service Vehicle Destination
Point#2>

IF <available passage ≥ 2>
     THEN <COMPARE distance of passages> AND <GET distance with lowest distance
value>
```

Fig. 5. Examples of rule for Transportation management

From Fig. 5, the first rule is for setting up a location for picking students from building to another while considering amount of vehicle to use by calculating possible passenger and a vehicle seat amount. The second rule is the rule to prevent sending a shuttle bus in case of there are left over seats in the existing shuttle bus that the path is overlapped with another or having the same direction.

4 Result and Discussion

To test a potential of the service, we retrieved a movement of students from Friday schedule at around 2:00 PM from the database. The details of movement are shown in Table 3 and Fig. 6 is illustrated for indicating locations corresponding to detail in Table 3. From the data, we found that 457 students having classes at the time, and they spread out to all eight buildings.

Table 3. An amount and movement of 457 students on 2:00 PM, Friday

Building	Headcount before 2.00 PM	Movement of people to destination After 2.00 PM								
		A	B	C	D	E	F	G	H	Unspecified
A	67	23	0	0	24	10	10	0	0	0
B	52	2	0	13	17	0	0	0	15	5
C	30	0	0	0	0	20	10	0	0	0
D	45	0	0	7	0	20	18	0	0	0
E	99	12	15	0	22	35	0	0	10	5
F	75	10	11	17	0	9	28	0	0	0
G	89	0	0	0	0	6	22	59	2	0
H	0	0	0	0	0	0	0	0	0	0

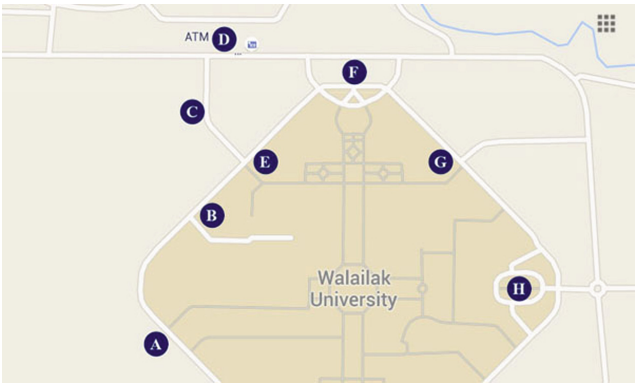


Fig. 6. An illustrated map to indicate buildings related to data in Table 3

We then asked the solution from the transportation management service based on the details in Table 3, and the criterion is that there are twenty units of a 30-seat shuttle bus. The service returns the solution shown in Table 4. From the overview of the result, the service returned a quality solution based on the given rules as eleven shuttle buses are employed for the issue. Nine routes are assigned. Additional route, route#0, was suggested for those who have unspecified destination.

Table 4. A solution from transportation management service by data from Table 3

Shuttle bus route ID	Amount of shuttle bus	Route
1	1	Building A→ D
2	1	Building A→E→F
3	1	Building B→C→D
4	1	Building B→E→F→H
5	1	Building D→C→E
6	1	Building D→F
7	1	Building F→E→C
8	1	Building F→B→A
9	1	Building G→F→E
0	2	All buildings and to cafeteria, dormitory, library, and entrance of university

From the result in Table 4, we found that the returned solution was not the same as a plan from human. Since university staffs concern on the starting point and tend to worry on student amount, they assign bus(es) in an extravagant manner. For example, starting point A has three routes assigned to go to Building D, E and F respectively. However, the system executed the plan based on the rule (rule#2 from Fig. 5) to merge the route in case there is an available seats in a bus and the routes have the same direction. This made a difference in a route number and a bus assignment between the human and system result. In details, we found that the map has a circular passage among building C, D, E and F. From any of those building, there are two available passages to each other. This issue brought much trouble to university staffs on planning a transportation service route and hence they assigned many routes from any building to another as C→D, C→E, C→F, D→C, D→E, D→F and so on to cover a possible confusion. In fact, the solution from the system can help to reduce an amount of bus usage, for instance, the system assigned route#5 and #6 from building D. Due to the merging rule, two buses were exploited and it suggested to merge the route into D→C→E instead of D→C→F since the distance of C→E is shorter than C→F based on given rule#3 from Fig. 5.

5 Conclusion

This work presents a framework of intelligent services to serve a management in campus using ontology as knowledge base. Campus ontology was designed to capture relevant concepts in campus including building, people, and things and semantic relation among them. Instances from actual data of Walailak University were mapped to the ontology to approve ontology coverage and correctness. Currently, two intelligent services, which are visualizing headcount in the building and a recommendation of service vehicle management, have been implemented based on the designed ontology. The visualizing tool powered by Google map API is developed for monitoring crowd in an area for supporting security management. A recommendation of service vehicle management is an automatic decision making system facilitating a complex management of shuttle bus route design to fit student movement. From testing, the recommendation service impressively returned a reasonable suggestion according to resource allocation.

To improve a quality of transportation management service, more rules will be designed based on other scenarios for better coverage. Moreover, more intelligent services, such as room management and automatic advisor appointment, will be implemented for expediting studying life of students and staffs in a campus.

Acknowledgements. This research was supported by Rajamangala University of Technology Suvarnabhumhi. Authors would like to thank Walailak University for providing students' data, courses and building information for our test study.

References

1. Stavropoulos, T.G., et al.: System architecture for a Smart University Building. In: ICANN'10 Intelligent Environmental Monitoring, Modeling and Management Systems for better QoL Workshop, Tessalaniki (2010)
2. Bueno-Delgado, V.M., et al.: The Smart University experience: a NFC-based ubiquitous environment. In: 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. https://doi.org/10.1109/IMIS.2012.110
3. Aqeel-ur-Rehman, et al.: Building a Smart University using RFID technology. In: 2008 International Conference on Computer Science and Software Engineering. https://doi.org/10.1109/CSSE.2008.1528
4. Khabou, N., et al.: A threshold based context change detection in pervasive environments: application to a smart campus. In: 5th International Conference on Ambient Systems, Network and Technologies (ANT-2014). Proc. Comput. Sci. **32**, 461–468 (2014)
5. Seghrouchni, A., et al.: Ambient intelligence application: introducing the campus framework. In: 13th IEEE International Conference on Engineering of Complex Computer Systems. https://doi.org/10.1109/ICECCS.2008.31
6. Chuling, L., Zanfu, X., Ping, P.: A discussion on the framework of smart campus. In: Third International Symposium on Intelligence Information Technology Application. https://doi.org/10.1109/IITA.2009.208
7. Fan, J., Stewart, K.: An ontology-based framework of modeling movement on a smart campus. In: Analysis of Movement Data, GIScience 2014 Workshop—Vienna, Austria

8. Mizoguchi, K.: Tutorial on ontological engineering—Part 1: Introduction to ontological engineering. In: Mizoguchi, R. (ed.) New Generation Computing, vol. 21, no. 4, pp. 365–384. OhmSha&Springer (2003)
9. Noy, N.F., McGuinness, D.L.: Ontology development 101: a guide to creating your first ontology. In: Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880
10. Hozo ontology editor. http://www.hozo.jp
11. Buranarach, M., Thein, Y.M., Supanithi, T.: OAM: an ontology application management framework for simplifying ontology-based semantic web application development. Int. J. Softw. Eng. Knowl. Eng.
12. Google Map API. https://developers.google.com/maps/documentation/javascript/3.exp/reference

TLex+: A Hybrid Method Using Conditional Random Fields and Dictionaries for Thai Word Segmentation

Sarawoot Kongyoung(✉), Anocha Rugchatjaroen, and Krit Kosawat

National Electronics and Computer Technology Center, National Science and Technology Development Agency, Pathum Thani, Thailand
{sarawoot.kon, anocha.rug, krit.kos}@nectec.or.th

Abstract. This paper proposes a new hybrid method of using Conditional Random Fields (CRFs) integrated with three dictionaries for Thai word segmentation. Based on TLex (Thai Lexeme Analyser), a pre-processing phase with an unambiguous list dictionary is added to deal with long expressions and long named entities (NE). Following this, the rest of the text is sent to the original TLex system, based on CRFs, to be segmented into words. Next, another dictionary is applied in a post-processing phase to check the number of unknown words in each of the top scored alternative segmentations from TLex in order to choose the best one. Finally, another NE dictionary is employed to merge each segmented named entity into one word. The results show that this hybrid method can improve the precision, recall and F-measure of TLex from 93.63%, 94.91%, and 94.27% to 97.64%, 97.37%, and 97.50%, respectively to become the most accurate Thai word segmentation system presently available.

Keywords: TLex+ · Hybrid method · CRFs · Dictionary
Word segmentation · Expression · Named entity · Unknown word · Thai

1 Introduction

Thai is the official language of Thailand. Based on the alphabetical system with its own characters, it writes from left to right without spaces between words. Thai word segmentation is then one of the most classic and indispensable issues in Thai language processing. For example, "ฉันเล่นฟุตบอล" should be segmented to "ฉัน(I)|เล่น(play)|ฟุตบอล (football)|" before it is sent to the next process, whether it be translation, speech synthesis, data mining, etc. However, there exists many ambiguous cases where more than one segmentation is possible; for example, "ตากลม" can be segmented to "ตาก(expose to)|ลม(wind)|" or "ตา(eye)|กลม(round)|" and both are meaningful, depending on their context. As a consequence, many Thai word segmentation systems have been developed to solve these problems for more than 30 years.

1.1 Rule-Based Methods

In the beginning, when computers were not as powerful, rule-based methods were normally used that were based on Thai syllable composing rules [1–3]. These methods

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_10

were simple and small but usually covered monosyllabic words only. However, many Thai words are multisyllabic.

1.2 Dictionary-Based Methods

Dictionary-based methods were then introduced [4]. These methods compared text-to-segment with words in their dictionaries. Accuracy was obviously greatly improved upon, however, the method still had problems with ambiguous cases and unknown words. As a result, "Longest Word Mapping" [5] was proposed to help solve the ambiguous cases. It was a greedy algorithm that compared characters from left to right and chose the longest strings mapped with ones in the dictionary. This method was later improved upon with "Back Tracking" [6]; i.e. if the longest words could not successfully segment text up to the end of the strings, then went back to each previous successful point and tried with a shorter word instead. On the other hand, "Maximal Matching" [7] preferred segmenting all possibilities first and then chose the one with the less number of words and unknown words. That meant it still preferred longest words but not necessary on the left side of the strings.

1.3 Machine Learning-Based Methods

With the introduction of more powerful computers, Thai word segmentation systems have moved towards machine learning-based methods [8, 9]. These methods learn from very large Thai words pre-segmented training corpora and create their own rules, based on statistical models, of how to segment the Thai character strings into words. These methods can deal very well with ambiguous cases because they consider their contexts too, and can even segment unknown words quite efficiently. However, some cases, such as compound words and named entities, are exceptions and need to be treated separately.

1.4 Hybrid Methods

Since every method has its advantages and disadvantages, many hybrid methods were introduced in an attempt to gain the benefit from their combined strengths. These included rule and dictionary-based hybrid methods [10, 11], dictionary and machine learning-based hybrid methods [12], and rule and machine learning-based hybrid methods [13].

1.5 Benchmark for Enhancing the Standard of Thai Language Processing (BEST)

However, although there were many proposed Thai word segmentation systems in place, it was very difficult to compare them because they did not use the same segmentation patterns and scoring criteria. Therefore, in 2009, National Electronics and Computer Technology Center (NECTEC) organised two activities based around Thai word segmentation: "BEST 2009: Thai Word Segmentation Software Contest" [14], and "InterBEST 2009 Thai Word Segmentation: an International Episode". At these

events, all participants had to develop their own system using the provided training data and their results were evaluated with the same test set and the same scoring program, thus allowing comparisons with all the submitted algorithms.

It was noted that the best three results [13, 15], (including Thai Lexeme Analyser (TLex) [16]) from the five submitted algorithms all used Conditional Random Fields (CRFs) as their learning machine. Therefore, it could be assumed from these results that this was the preferable method for Thai word segmentation at present.

1.6 Proposed Method

In this paper, we would like to improve TLex's capacity to deal with long expressions, named entities, and unknown words by proposing a new hybrid model, called TLex+ 1.0. This will use Conditional Random Fields integrated with three dictionaries in the pre- and post-processing phases. The results will be evaluated with the same test set and the same scoring program used by InterBEST 2009 to allow benchmarking with the previous algorithms.

The organization of this paper is as follows: Section 2 shows the problems of the original TLex system and proposed solutions. Section 3 presents the system diagram of our proposed algorithm. Section 4 provides experimental design and results. Section 5 discusses the findings. And finally, Sect. 6 provides our conclusion and proposals of future work.

2 Problems and Proposed Solutions

Although TLex is a good system, it still has the following problems:

2.1 Problem with Long Expressions

We noticed that TLex and indeed all other CRFs models cannot segment long expressions efficiently. This is because these expressions usually have low occurrence but are composed of many high occurrence short words. So, statistically, CRFs always segment them into many small words, which is normally incorrect. For example, "มีเงินมีทอง" is an expression which means "rich". Therefore, it should be segmented as one word: "มีเงินมีทอง‖". However, TLex always segments it into four words:

Table 1 Words and their occurrences in InterBEST training corpus

Word	Occurrence
มีเงินมีทอง (rich)	1
มี (have)	117,973
เงิน (money)	10,067
ทอง (gold)	1,078

"มี(have)|เงิน(money)|มี(have)|ทอง(gold)|" because of the way they occur in the training corpus, as shown in Table 1.

To solve this problem, a dictionary of unambiguous long expressions was created as a forced list, meaning that they can only be segmented in this way. The criteria of choosing words in the forced list are: (1) at least seven characters in length, for the reason that (from our experiments) short words tend to create more ambiguous cases than solve them; (2) low occurrence expressions, composed of high occurrence short words, for the same reasons as explained above.

This dictionary is applied in the pre-processing phase before using CRFs. Therefore, expressions already treated will not be sent to CRFs.

2.2 Problem with Unknown Words

After considering the five best results of segmentation from CRFs, TLex normally chooses the best score, which is not always the correct one. Because CRFs at times segment words that are either too short or too long, this then can create unexpectedly unknown words, as shown in Table 2.

Table 2 Top five segmentations from TLex and their numbers of unknown words

Segmentation	Score	Number of unknown words (in bold)
เชิดชูปรีชาสามารถ|ของ|ผู้นำ|ที่|พา|ชาติไทย|รอดพ้น|ลัทธิ|อาณานิคม	0.1821235245	1
เชิดชู|ปรีชาสามารถ|ของ|ผู้นำ|ที่|พา|ชาติไทย|รอดพ้น|ลัทธิ|อาณานิคม	0.1267063462	0
เชิดชูปรีชาสามารถ|ของ|ผู้นำ|ที่|พา|ชาติ|ไทย|รอดพ้น|ลัทธิ|อาณานิคม	0.07324614639	1
เชิดชูปรีชาสามารถ|ของ|ผู้นำ|ที่|**พาชาติไทย**|รอดพ้น|ลัทธิ|อาณานิคม	0.06758710722	2
เชิดชูปรีชาสามารถ|ของ|ผู้นำ|ที่|**พาชาติไทย**|รอดพ้น|ลัทธิ|อาณานิคม	0.05812950031	2

To improve upon this issue, it is better to apply another dictionary to check the number of unknown words from each alternative and choose the one with the fewest unknowns instead. This method adapts the idea of counting result units in Maximal Matching [7]. The only difference is that this research counts the number of unknown words instead of segmented words. The counted number is used as a cost for choosing the best candidate from CRFs top results. The cost function is called $f_c(d_i)$ when d_i is the ith candidate in the top list and the best result can be found by equation (1).

$$result = \arg\min_i f_c(d_i) \tag{1}$$

From the above example, despite the score from CRFs, the second segmentation is selected instead.

2.3 Problem with Named Entities

The named entities in InterBEST Corpora included person, location and organization names. They were usually proper nouns but could be composed of many common

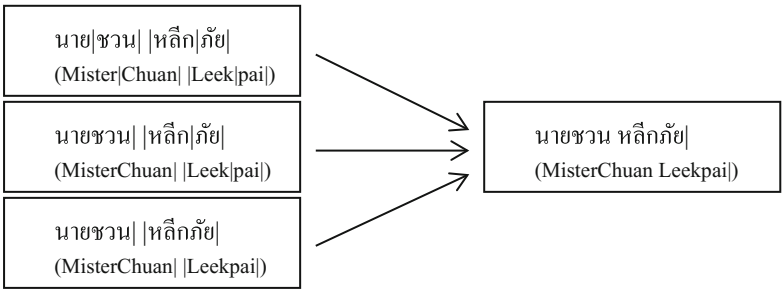


Fig. 1 NE merging

nouns too. However, according to the criteria of InterBEST, NEs should be segmented as one word. Therefore, for the same reason as in 2.1, we cannot rely only on CRFs but must also be wary of using the forced list too because the NEs are not always unambiguous.

To mitigate against this problem, we divided NEs into two lists: unambiguous NEs which were put into the forced list of 2.1; and a second list containing the rest which were compiled to a TRIE [17] to be compared with the end results after 2.2. If they were matched, small components of an NE would be merged into one word. The advantage of using TRIE over string matching, used in the original TLex, is that it is small, fast and does not necessarily require a 100% match; i.e. only the boundary of the string must be correct, no matter how the string is segmented inside. Fig. 1 shows an example of three different segmentation patterns of the person's name "Mister Chuan Leekpai", that can be merged into the same one NE by using this TRIE.

3 TLex+ 1.0 System Diagram

The proposed system, called TLex+ 1.0, aims to improve the quality of CRFs-based Thai word segmentations. Dictionary-based methods have a distinct advantage. They force word boundaries to be cut after using a given word list, which can then be used to enhance CRFs-based performance if they can be combined together.

After several attempts in scrutinizing the cooperation, the final proposed system architecture is shown in Fig. 2. The system works sequentially starting from:

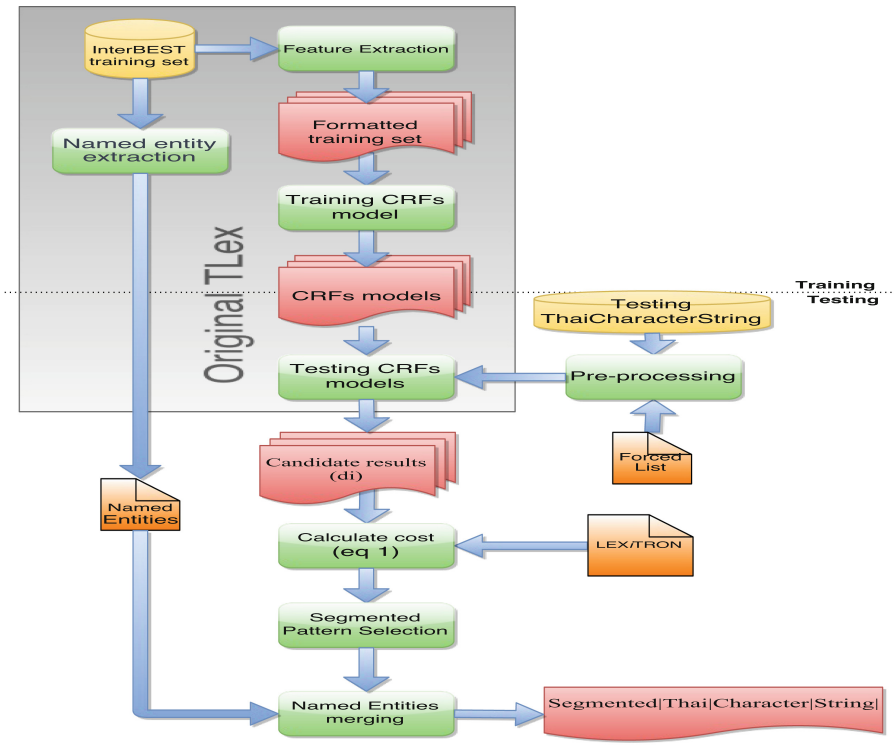


Fig. 2 TLex+ 1.0 system diagram

- Training CRFs models using character features, extracted from the InterBEST training data set.
- The creation of a forced list (3,897 entries), including unambiguous long expressions and named entities.
- The creation of another NE list (52,135 entries) extracted from the training data set.

To test the system, the steps were as follow:

- The input character strings are sent to the pre-processing system, where the forced list is applied to segment unambiguous expressions and named entities.
- The rest of the text is sent to the CRFs model to be segmented into small words. This should result in many possible segmented candidates.
- From the top five candidates sorted by CRFs' scores, the LEX*i*TRON dictionary (60,251 entries) [18] will be applied to check the number of unknowns and calculate cost.
- The candidate that has the lowest cost is selected as a result.
- The result is passed to the named entity-merging process.
- Finally, the answer is given after completing the merging of all possible named entities found.

4 Experimental Design and Results

The experiments were conducted over the protocol of InterBEST 2009, which contained approximately seven million words in the training set from eight genres: Article, Buddhism, Encyclopaedia, Law, News, Novel, Talk, and Wikipedia; while the test set contained approximately 500,000 words from 12 genres: previous genres plus NSC report, Old document, Royal news, and TV news.

The evaluations were presented in Precision (P), Recall (R) and F-measure ($F1$), calculated by these equations:

$$P = \frac{Number\ of\ Correct\ Words}{Number\ of\ Output\ Words} \times 100\% \tag{2}$$

$$R = \frac{Number\ of\ Correct\ Words}{Number\ of\ Reference\ Words} \times 100\% \tag{3}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

There were eight experiments conducted in this research to show the accuracies of pure CRFs (Fig. 3), CRFs with the forced list (Fig. 4), CRFs with the cost calculation (Fig. 5), CRFs with the new NE merging (Fig. 6), CRFs with the forced list and the cost calculation (Fig. 7), CRFs with the forced list and the new NE merging (Fig. 8), CRFs with the cost calculation and the new NE merging (Fig. 9), and finally CRFs with everything (or TLex+1.0 in Fig. 10). These experiments aimed to show the improvements with each additional module, and when combined together (Fig. 11).

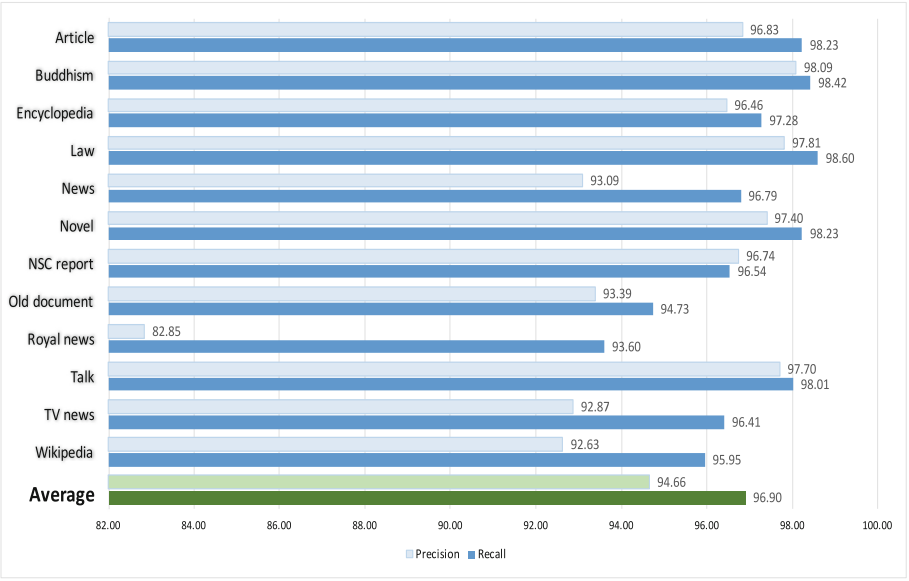


Fig. 3 Precision and recall of pure CRFs

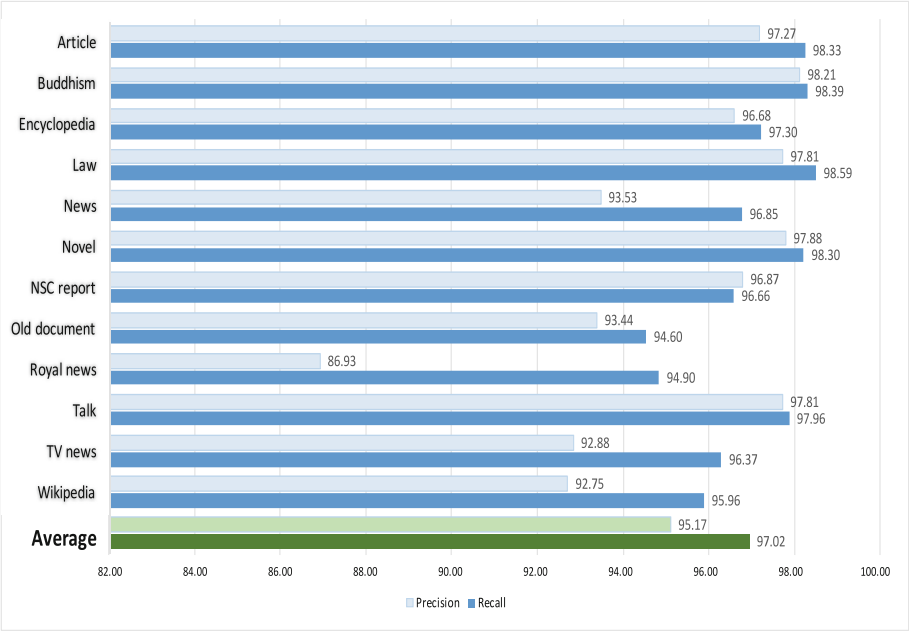


Fig. 4 Precision and recall of CRFs + forced list

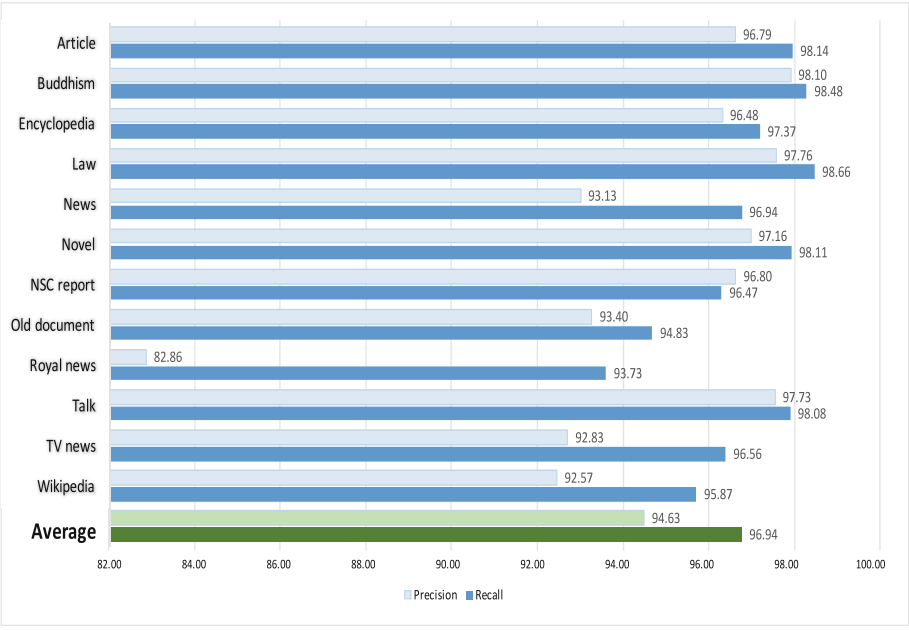


Fig. 5 Precision and recall of CRFs + cost calculation

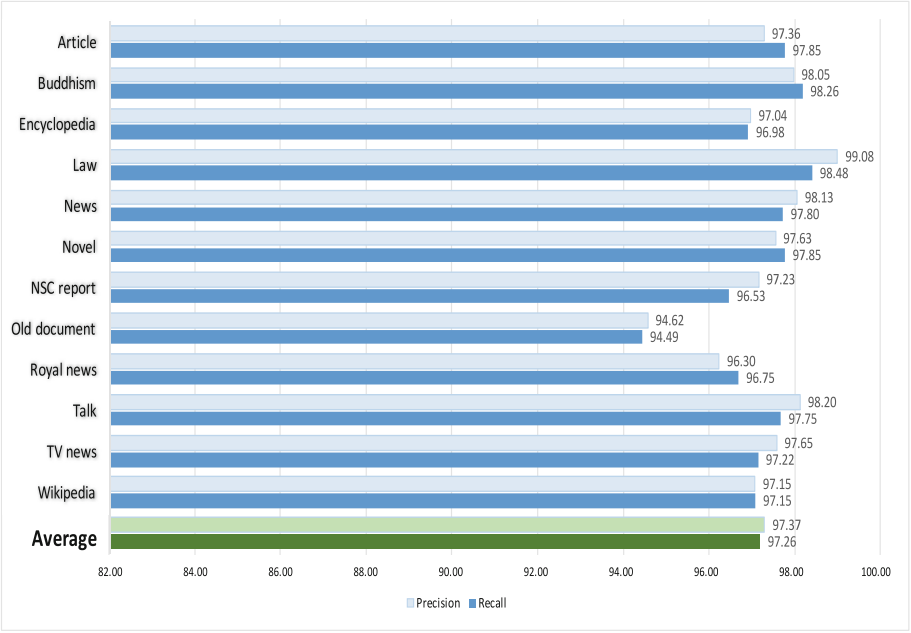


Fig. 6 Precision and recall of CRFs + new NE merging

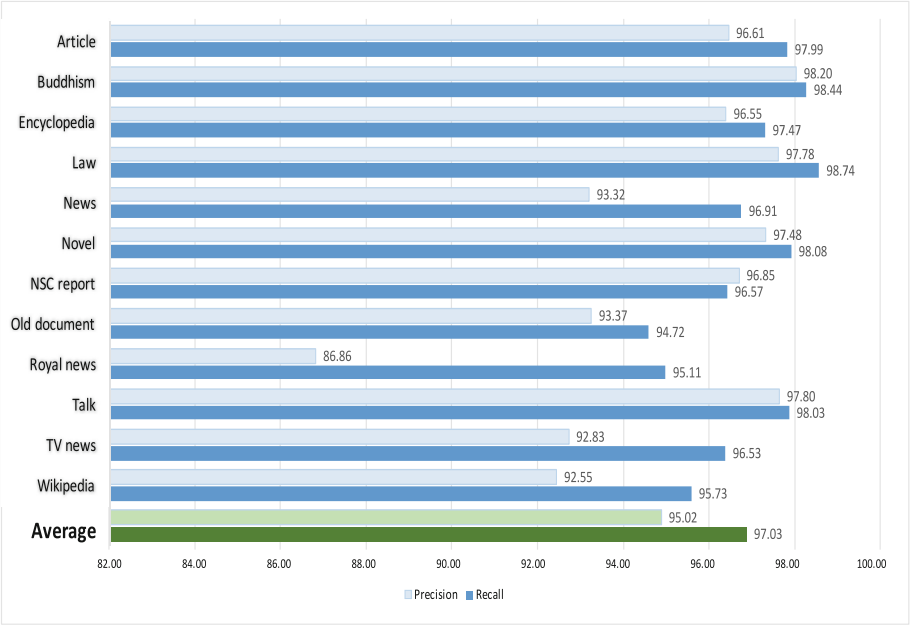


Fig. 7 Precision and recall of CRFs + forced list + cost calculation

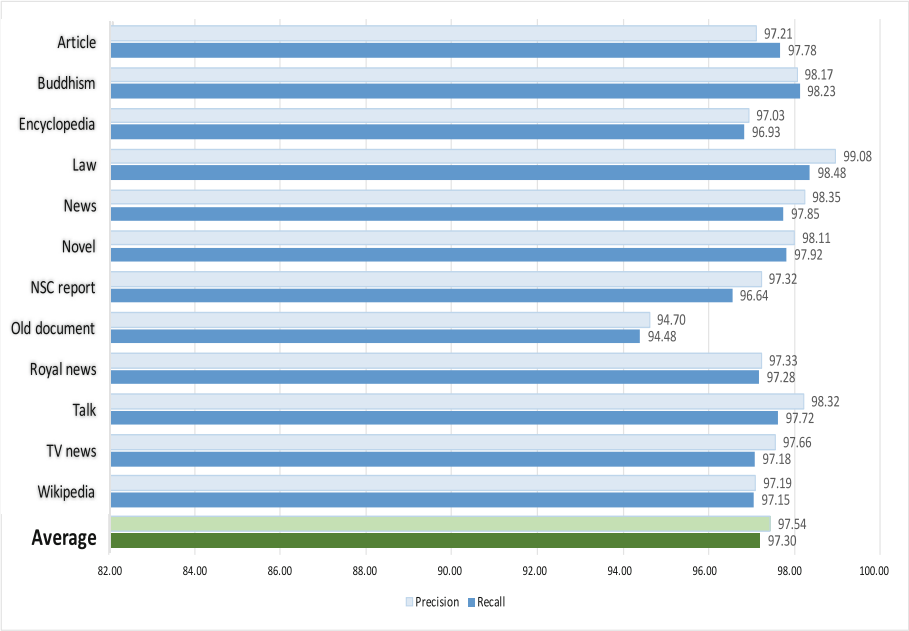


Fig. 8 Precision and recall of CRFs + forced list + new NE merging

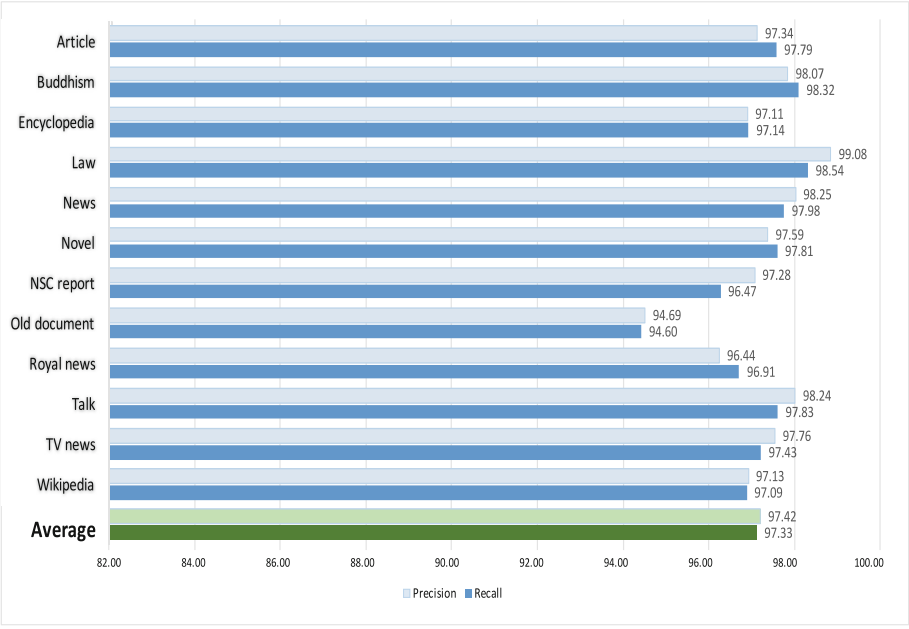


Fig. 9 Precision and recall of CRFs + cost calculation + new NE merging

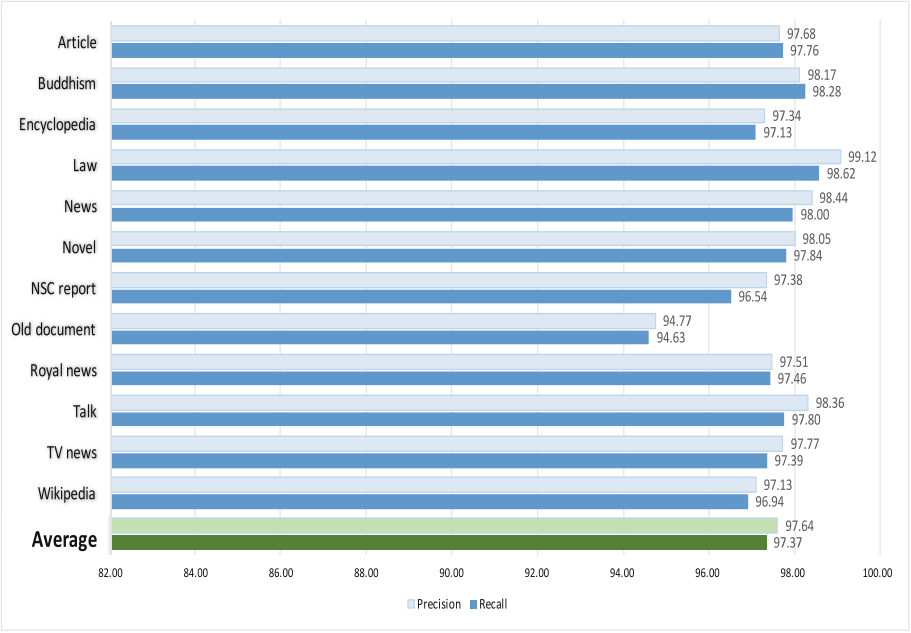


Fig. 10 Precision and recall of TLex+ 1.0

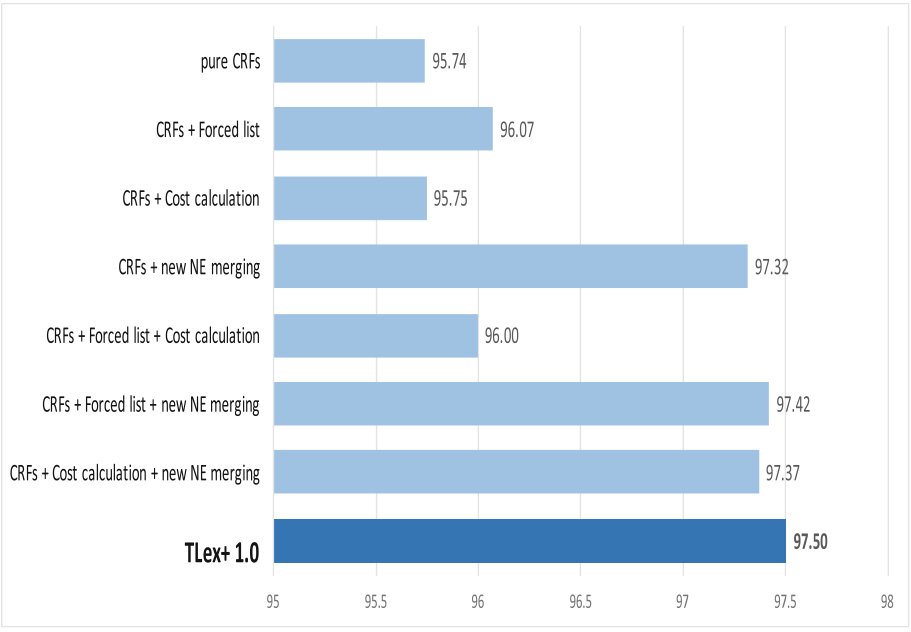


Fig. 11 F-measure comparison

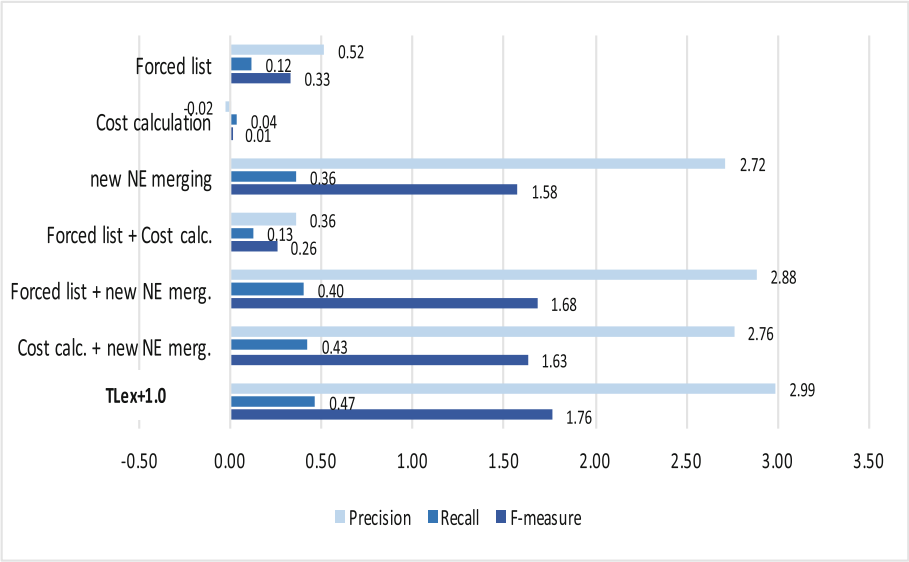


Fig. 12 Improvement comparison

5 Discussions

The results from Figs. 4, 5, 6, 7, 8, 9, 10 and 11 show the effect of adding different modules, compared to the pure CRFs in Fig. 3. As Fig. 12 shows, the forced list can improve the precision, recall and F-measure by 0.52%, 0.12%, and 0.33%, respectively, while the cost calculation can do so by only −0.02%, 0.04%, and 0.01%, respectively. The low improvement in the later may come from using LEX*i*TRON dictionary, which does not have the same word patterns as InterBEST. However, according to our experiments, using the word list extracted directly from InterBEST Corpora did not give higher scores, since InterBEST contained many incorrect words as well as too many segmentation patterns. The new NE merging using TRIE instead of string matching can improve accuracy the most, by 2.72% in precision, 0.36% in recall, and 1.58% in F-measure.

When combining together the forced list with the cost calculation, the precision, recall and F-measure improves by only 0.36%, 0.13%, and 0.26%, respectively. Further, the forced list with the new NE merging improves the precision, recall and F-measure by 2.88%, 0.40%, and 1.68%, respectively. And lastly, the cost calculation with the new NE merging improves the precision, recall and F-measure by 2.76%, 0.43%, and 1.63%, respectively.

However, the overall improvement by combining all three modules together can reach 2.99%, 0.47%, and 1.76%, respectively.

Compared to other Thai word segmentation systems that used the same scoring criteria, we have already bettered the second place system from InterBEST 2009, Suesatpanit [15], reporting 95.53% of F-measure, and also the first place system, Kruengkrai [13], reporting 96.86% of F-measure, as shown in Fig. 13.

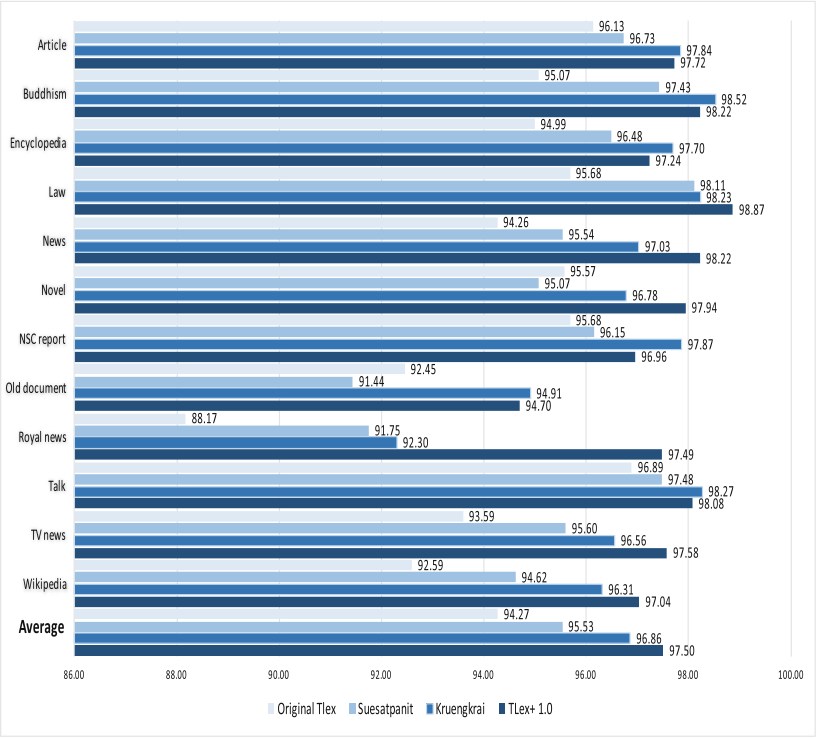


Fig. 13 F-measure comparison of the top four competitors

6 Conclusion and Future Works

We have proposed a new hybrid method, called TLex+ 1.0, using CRFs with dictionaries for Thai word segmentation. Based on the original TLex, three dictionaries have been added: the first one in pre-processing to segment Thai long expressions and unambiguous named entities, the second in post-processing to check the number of unknowns in order to select the fewest one, and the third dictionary to merge any separated NEs into one. The results, based on InterBEST 2009 evaluation, show that this hybrid method can improve the precision, recall and F-measure of the original TLex system from 93.63%, 94.91%, and 94.27% to 97.64%, 97.37%, and 97.50%, respectively and become the most accurate Thai word segmentation system presently available.

In the future, we will work more on word selection in the pre- and post-processing dictionaries to include unambiguous compound words, more expressions, and more named entities. In addition, we will also find another dictionary to replace LEX*i*TRON.

References

1. Thairatananond, Y.: Towards the design of a Thai text syllable analyzer. Master Thesis of Science, Asian Institute of Technology, Pathumthani (1981)
2. Charnyapornpong, S.: A Thai syllable separation algorithm. Master Thesis of Engineering, Asian Institute of Technology, Pathumthani (1983)
3. Theeramunkong, T., Sornlertlamvanich, V., Tanhermhong, T., Chinnan, W.: Character cluster based Thai information retrieval. In: 5th International Workshop on Information Retrieval with Asian Languages (IRAL2000), pp. 75–80, Hong Kong (2000)
4. Poovarawan, Y., Imarrom, W.: Thai syllable separator by dictionary (in Thai). In: 9th Annual Meeting on Electrical Engineering of the Thai Universities, Khonkaen (1986)
5. Varakulsiripunth, R., Ngamwiwit, J., Junwun, S., Chiwattayakul, S., Thipchaksurat, S.: Word segmentation in Thai sentence by longest word mapping (in Thai).(1989). In: Sornlertlamvanich, V. (ed.) Papers on Natural Language Processing: Multilingual Machine Translation and Related Topics (1987–1994), pp. 279–290. NECTEC, Bangkok (1995)
6. Raruenrom, S.: Dictionary-based Thai word separation (in Thai). Senior Project Report, Department of Computer Engineering, Chulalongkorn University, Bangkok (1991)
7. Sornlertlamvanich, V.: Word Segmentation for Thai in machine translation system (in Thai). In: Machine Translation, pp. 50–56. NECTEC, Bangkok (1993)
8. Meknavin, S., Charoenpornsawat, P., Kijsirikul, B.: Feature-based Thai word segmentation. In: Natural Language Processing Pacific Rim Symposium 1997 (NLPRS'97), pp. 41–46, Phuket (1997)
9. Kooptiwoot, C.: Segmentation of ambiguous Thai words by inductive logic programming (in Thai). Master Thesis of Science, Department of Computer Engineering, Chulalongkorn University, Bangkok (1999)
10. Sawamipak, D.: Construction of Thai Syntax Analysing Software Under UNIX (in Thai). Thammasat University, Bangkok (1990)
11. Runapongsa, K., Urathammakul, P.: Rule-based approach and new dictionary for Thai words segmentation (in Thai). Final Report, Faculty of Engineering, Khon Kaen University, Khon Kaen (2006)
12. Kawtrakul, A., Thumkanon, C., Seriburi, S.: A statistical approach to Thai word filtering. In: 2nd Symposium on Natural Language Processing (SNLP'95), pp. 398–406, Bangkok (1995)
13. Kruengkrai, C., Uchimoto, K., Kazama, J., Torisawa, K., Isahara, H., Jaruskulchai, C.: A word and character-cluster hybrid model for Thai word segmentation. In: InterBEST 2009 Thai Word Segmentation Workshop, pp. 24–29, Bangkok (2009)
14. Kosawat, K., Boriboon, M., Chootrakool, P., Chotimongkol, A., Klaithin, S., Kongyoung, S., Kriengket, K., Phaholphinyo, S., Purodakananda, S., Thanakulwarapas, T., Wutiwiwatchai, C.: BEST 2009: Thai word segmentation software contest. In: 8th International Symposium on Natural Language Processing, pp. 83–88, Bangkok (2009)
15. Suesatpanit, K., Punyabukkana, P., Suchato, A.: Thai word segmentation using character-level information. In: InterBEST 2009 Thai Word Segmentation Workshop, pp. 18–23, Bangkok (2009)
16. Haruechaiyasak, C., Kongyoung, S.: TLex: Thai lexeme analyser based on the conditional random fields. In: InterBEST 2009 Thai Word Segmentation Workshop, pp. 13–17, Bangkok (2009)
17. Apache Commons Collections. https://commons.apache.org/proper/commons-collections/
18. LEX*i*TRON Thai-English Dictionary. http://lexitron.nectec.or.th

An Expert System Using Ontology as Knowledge Base for Personalized Rice Cultivation Suggestion

Wichuphon Chariyamakarn[1(✉)], Poonpong Boonbrahm[1], Salin Boonbrahm[1], and Taneth Ruangrajitpakorn[2,3]

[1] School of Informatics, Walailak University, Nakorn Si Thammarat, Thailand
{wichuphon, poonpong, salil.boonbrahm}@gmail.com
[2] Department of Computer Science, Thammasat University, Pathum Thani, Thailand
taneth.rua@nectec.or.th
[3] Language and Semantic Technology Laboratory, Nation Electronics and Computer Technology Center (NECTEC), Pathum Thani, Thailand

Abstract. This work presents a framework to assist Thai rice farmers in cultivation planning. Knowledge of rice production is extracted from best practice documents and kept in ontology. Concepts relevant to rice production such as soil type of every location in Thailand and rice seeds with their strong and weak properties are semantically drawn in the ontology. The ontology associated with rules to make decision is exploited as a knowledge base for the system to infer a reasonable suggestion based on farmers' specific information, context and need. According to farmers input, different suggestions can be given including seed selection, land preparation, concern in cultivation based on preferred seed, etc. From testing, the generated recommendations were satisfied by rice production experts for precise criteria analysis and proper reason selection in inference process. An experiment result of seed selection showed that the system generated suggestions with 0.56 precision score and 0.97 recall score comparing to human expert suggestions.

Keywords: Expert system · Rice production · Reasoning based recommendation · Automatic suggestion · Ontology

1 Introduction

Rice and rice production play a crucial role in Thai society since it serves as main food for Thai citizen and a major export product. Because of its importance, many researches supported by Thai government have been conducted to study an effective cultivation and establish a best practice for rice production. Despite many researches have been made and published, rice production in Thailand is still under threat in productivity rate. The major cause is that farmers have not followed the best practice and a suggestion from research results correctly.

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_11

From observation, it reveals that Thai farmers have their trust in the best practice given by government supporting researches, but the information is excessively difficult for them to understand due to its complexity and many relevant factors. The research documents are in descriptive details and their wording is in academic literature level which is not well understood by local farmers. Moreover, there is a lot of information in such fashion in public since many research institutes and relevant government agencies are trying their best to provide their own useful information to farmers. The information overwhelming in public can confuse the readers as each of the publication shares not only a common suggested process, but also specific result of their own studies and some may conflict with others. Furthermore, the research studies often worked in a specific part of the country but gave suggestion for general area. These are difficult for farmers to select related details to their context, and they often apply incorrect suggestions which are not suitable for them because of a lacking of an apprehension.

In practice, farmers need a solution for their specific requirements and settings. This leads to several research works that design a framework focusing on a divergence of users. Therefore, the term "farmer centered development" has been introduced [1]. This concept focuses on individual farmers and gives the best practice suitable for them. The concept requires users to provide their additional information which can lead to individual problem-solving. For such case, an expert system containing farming knowledge is needed for matching an individual setting and domain knowledge. Since the research studies in the field of rice production are normally given as descriptive information, knowledge of the domain is hard to comprehend. Apparently, ontology is frequently applied to represent the domain knowledge in agriculture in many of existing works [2–6]. The reason why Ontology is often used is domain knowledge can be complicated and ontology has ability to represent semantic concepts in the knowledge and make it understandable for human and computer [7].

In this work, the main focus is to create the supporting system to recommend a suitable method for rice production according to the best practice. Specific information, context and needs are reliably considered for the recommendation to correspond to each individual setting environment. In a design of our knowledge base, deep understanding of rice cultivation processes and related concepts are focused along with knowledge of the risks involved in those processes and product. How to effectively prevent or minimize undesired factors were also considered. This can ensure a quality by design [8] for our recommendation to an individual farmer.

2 Background

2.1 Best Practice in Rice Cultivation

In this section, details on best practice in rice cultivation process are summarized. To produce rice, there are many processes that involve with several factors. The processes include seed selection, paddy preparation, planting, watering, fertilizing, pest and disease controlling, harvesting, etc. [9, 10]. For ease of understanding, a rice production life cycle [9] is illustrated in Fig. 1.

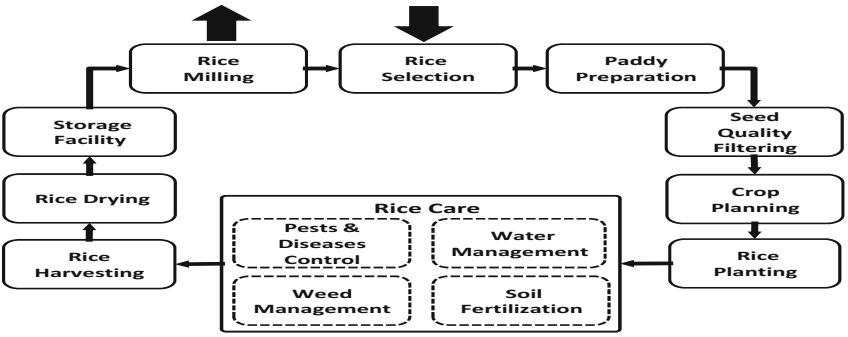


Fig. 1. A cycle of rice cultivation

An effect of factors in each process is also important. Farmers have to select the best material and process that fit their farm setting. Since there are many varieties for each factor selection, farmers should consult with experts in the field to prevent missing minor details that can yield unexpected result.

For example, a seed selection should be carefully selected depending on (1) a time of the year they plan to cultivate, (2) the environment of the land, (3) a prevalence of pest and disease in the area and time, (4) available watering source, and (5) production rate of the seed. Selecting an appropriate seed can prove a successful production rate while one missing factor can lead to a failure of the entire production. We select some seeds and a few of their attribute from the rice seed researches [9] to demonstrate them in Table 1 for reference (S, N and W notation of disease and pest factor in the table stand for Strong, Neutral and Weak, respectively).

Not only the seed selection can decide the productivity, but also other factors such as planting technique and watering system can greatly affect a production. Depending on farmer budget, the best planting technique is machine-based drill planting for the capability to control distance between each rice sprout, but this method has an expensive cost. Another famous planting technique is broadcasting method. This method is a cheap and manual method in planting, and it requires an expertise in casting since rice sprouts may group together and will not grow up as usual. The downside of it is a waste in seed spreading and it will increase the cost for buying seed, hence this method is suggested to only a low cost seed [10].

In fact, farmers nowadays recklessly ignore these data and often aim for the seed that yields the best production rate or in high demand. Thus, their product is not in good quality and insufficient to repay the cost they spend.

2.2 Related Works

In agriculture, many IT systems have been introduced to assist the farmers. We review those works and summarize them into Table 2.

From Table 2, we found that many works, such as User Centered Ontology for Sri Lankan farmers [1] and Design of Ontology Based Ubiquitous Web for Agriculture—A farmer Helping System [12], focused on using technology to comfortably provide

Table 1. A brief knowledge of rice seeds and some of related properties

Rice seed varieties	Property							
	Photoperiod sensitivity	Disease			Pest			Yield estimation (kg/1600 m^2)
		Rice blast	Bacterial leaf blight	Brown spot	Rice gall midge	Green rice leafhopper	Brown plant hopper	
RD-1	Yes	N	W	S	W	S	W	740
RD-2	Yes	W	W	S	W	S	W	765
RD-3	Yes	W	W	N	W	S	W	665
RD-4	Yes	W	W	S	S	S	S	650
RD-5	No	S	S	S	N	N	N	565
RD-6	No	W	W	N	S	N	S	670
KhaoDawk Mali-105	No	N	N	N	N	W	W	365
Niaw San-pah-tawng	No	S	N	S	S	N	S	530
MueyNawng-62 M	No	W	W	S	S	N	W	540
LeuangYai-148	No	N	S	S	W	N	W	550

Table 2. Technologies and assisting tools in agriculture domain

Title	Technology	Objective	Result
A lifecycle-based design methodology of lightweight ontology and its application to cultivating high quality mandarin orange [3]	• Semantic web, ontology	• To create ontology for lifecycle-based cultivating high quality mandarin orange	Lightweight ontology containing knowledge of mandarin orange cultivation
Intelligent livestock farm management for context data model design [11]	• Sensor data based on WSN technology	• To provide the intelligent service for users by utilizing livestock farm environment and situation information through RDF-based context model based on the raw level sensing data	The intelligent livestock farm management services using hardware to collect data and providing instruction according to situation
User centered ontology for Sri Lankan farmers [1]	• Mobile based information systems	• To provide agriculture information for the farmers in Sri Lanka	A ontology to assist farmers based on their context in Sri Lanka
Design of ontology based ubiquitous web for agriculture—a farmer helping system [12]	• Semantic web, ontology	• To use ontology is knowledge base for semantic search	A framework by integrating agriculture information
Web and mobile technologies in a prototype DSS for major field crops [13]	• Mobile services (SMS) • Web-based decision support System	• To analysis, design and implementation, evaluation, of the proposed decision support system	Web-based DSS (MAFIC-DSS) that provides intelligent consultation for possible alternative cultivations
An ontology-based knowledge representation and implement method for crop cultivation standard [14]	• Semantic web, ontology	• To create a representation method which combines domain ontology and task ontology based on crop cultivation standards	A plant standardized management based on the integration of existing resources and the developing new resources

information to farmers. Another research trend in smart farming, such as Intelligent Livestock Farm Management for Context Data Model Design [11], is to utilize hardware to collect valuable data in individual farming to manage and analyze the cause and effect to improve farming method and boost production rate. All above-mentioned researches are a decent research aiming to help farmers by gathering information and data from many sources. This trend is greatly acceptable as a supporting tool for those farmers who can apply that information in their farming. However, there are some farmers who do not excel in information acquisition and apply them into actual method, for example farmers in a developing country. A decision making service is needed to help them in indicating a solution suitable for them. In this work, we aim to develop an expert system that can help farmers in making complicated decision in rice production using knowledge extracting from government supported best practice.

3 Ontology-Based Recommendation for Farmer-Centered Rice Production

This framework is designed for assisting rice farmers by giving a list of instructions in rice production based on the knowledge extracted from a best practice. An inference engine powered by a heavyweight ontology is exploited to make a logical solution to the personal environment. Since this system concerns individual setting, each farmer must give the setting information as an input to system and a model of the recommendation is made according to it. For rules used in an inference engine, several sets of rules are defined separately for different purposes because of many factors involved in rice production process as describe in a Background Section. An overview of the proposed framework is illustrated in Fig. 2.

3.1 Individual Farmer Information

This work aims to assist a farmer in rice production by concerning personal setting and need. Therefore, information from farmer is required to collect and generated into a model of each farmer. The required data are separated into two parts. The first part is

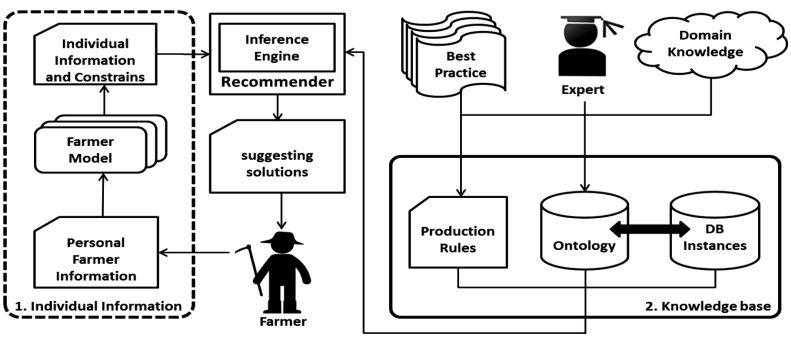


Fig. 2. An overview of the ontology based recommendation for farmer centered rice production

Table 3. Examples of farmer information for the recommendation service and their description and type

Information field	Definition	Type
Land location	Location of a paddy field planned to cultivate rice product to estimate environment, soil type, pest prevalence, etc.	Mandatory
Planting time	Time planned to start rice production to match up the seasonal environment such as rain, range of day time, etc.	Mandatory
Land size	A size of a paddy field area	Optional
Labor amount	An amount of labor in production, referable to size of the land	Optional
Owned instrument	A check box for informing available machine or tool to use in farming such as tracker, mill, etc.	Optional
Experience in rice farming	Information of previous rice production, planting technique in use, rice seed, to determine expertise of a farmer	Optional
…		
Preferable rice seed	Specify a rice seed farmer intends to use in rice production	Optional

the mandatory data that are related to field location, field size, estimated budget, etc. The second part is optional data such as specific seed that the farmer prefers, available instruments in farming, available water source, etc.

The mandatory data are a required input to give a suggestion according to best practice, but the optional data are driven to return a suggestion specifying for individual farmers. Some of the information as input to the system is exemplified in Table 3.

3.2 Rice Production Ontology

From many research studies in rice and rice production [9, 10, 15] this ontology was manually crafted to represent an axiom and semantic relations among concepts, not a taxonomy of farming concepts. In this research, Hozo Ontology Editor [16] is used as a development tool. The designed ontology contains forty-six major classes conceptualized from rice production world [7]. Some parts of Rice Production Ontology (in minimized due to limitation of space) are exemplified in a graphical image in Fig. 3. Some are chosen for explanation in detail in Table 4.

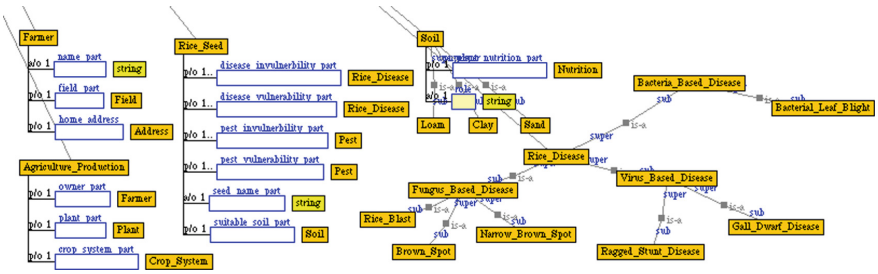


Fig. 3. Some parts of rice production ontology in hozo ontology editor UI

Table 4. Major concepts in rice production ontology

Class name	Subclass	Property	Description
Farmer	–	has **Paddy Field** has budget … has **Instrument** has preferred Rice Seed	Individual farmer information related to a rice production process including land planned to cultivate, budget in production, available instruments, etc.
Paddy field	–	has size has **Soil** has **Address** – in Area of Thailand	Land for cultivation, related information in terms of physical attribute, ownership, etc.
Soil	Clay Loam … Sand Silty Clay Sandy Loam	has **Nutrition** has pH value … has drainage rate	Knowledge of soil in agriculture and its properties, following a research studies of Soil in Thailand
Area of Thailand	Central Region – Bangkok – Pathumthani – Ayutthaya … Northern Region – Chiang Mai – Lampang …	has **Prevalence Plant Disease** has **Prevalence Pest**	A representation of specification of each location in Thailand. A taxonomy of Thai area and its information relevant to agriculture prohibition and environment
Rice seed	RD – RD-1 – RD-2 … Supanburee-1 … LebnokPattani-47 KhaoDawk Mali-105	has **Pest Invulnerability** has **Pest vulnerability** … has **Plant Disease Invulnerability** has photoperiod sensitivity has Suitable **Soil**	Varieties of Rice and their advantage or disadvantage in Rice production PS. Subclass is a list of variety name usually in Thai word

3.3 Recommendation and Rules

To obtain a suggestion, an inference engine is applied in the system. Several sets of rules are designed for different issues. In this work, OAM (Ontology Application

IF <FIELD is in AREA has PEST PREVALENCE = ***Brown Plant Hopper***> AND < FIELD is in AREA has DISEASE PREVALENCE = ***Brown Spot***>
THEN <RICE SEED = <RICE SEED has Pest Invulnerability = ***Brown Plant Hopper***> AND <RICE SEED has Disease Invulnerability = ***Brown Spot***>>
IF<RICE SEED has Pest Invulnerability = ***Brown Plant Hopper***> AND <RICE SEED has Disease Invulnerability = ***Brown Spot***>
THEN <RICE SEED = RD-4> AND <RICE SEED = RD-6>
IF <FARMER has Preferred RICE SEED = ***RD-1***>AND <FIELD is in AREA has DISEASE PREVALENCE = ***Bacterial Leaf Blight***>
THEN <FARMER *prepare*<INSTRUMENT = ***Streptomycin***> AND <INSTRUMENT = ***Oxytetracycline***> AND {*suggest preventing Fertilizer with high* ***Nitrogen***}

Fig. 4. Exemplified production rules designed for rice production recommendation

Management) framework [17], developed and provided by National Electronic and Computer Center, is used as a main tool in both instance mapping to ontology and a recommendation system. It provides a user interface to generate a production rule according to classes and their properties in a given ontology.

The rules were created following the suggestion given in the best practice. The format of the production rule is split into two components in a form as follow:

`IF <condition(s)> THEN <action>`

When the conditions are met, the action will be executed. The condition part is matched to farmer information while the action part is a list of solutions. Examples of the production rules in pseudo representation are given in Fig. 4. The system is a web-based application for easily accessibility from users. Inputs of the system are farmer information. The designed user interface is illustrated in Fig. 5.

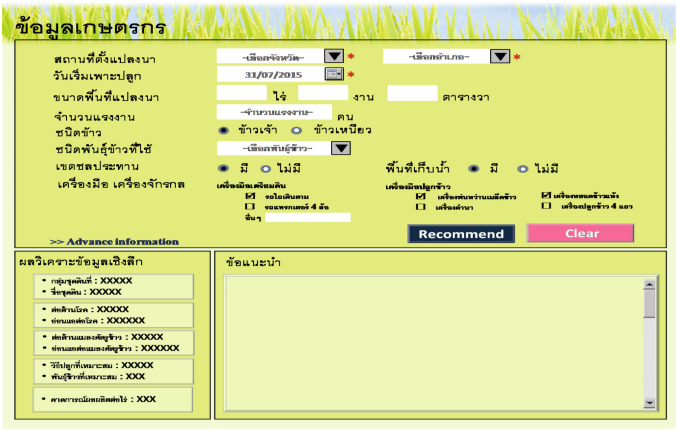


Fig. 5. The designed user interface

3.4 Scenario Usage

The first scenario is that a farmer provided only the least requirement of information to a system. The mandatory details are a land location for planting rice and a time planned for cultivation. For this case, only suggestion the service can generate is a seed selection recommendation and what to be aware in rice production. As input, the land location is specified to *Ban Pho district, Chachoengsao province* and the planned time to begin planting is June. From these two data to the system, properties of the given land location can be retrieved while an amount of rainfall and sunlight of designated time was specified. These data are the basic elements for inference engine and rules to find a matched solution. The result of the service (translated to English) is shown in Fig. 6.

The result in Fig. 6 was generated considering properties of given land location and time. The given location was reported with a prevalence of *rice green leafhopper* and *brown spot disease*; therefore, many seeds with their invulnerability are given into a possible list such as *RD-1, RD-2, RD-4, PathumThani-60*, etc. Unfortunately, some of those in the list are seeds of glutinous rice such as *RD-2* and *RD-4* that are not matched to the criterion and thus are removed from the given list. Furthermore, the given time is a month with longest daytime and this factor scopes to the seed with photoperiod insensitive property. Among those seed met the given criteria, *RD-1* seed is the best product yield per area and was selected as a solution of the system. For the given planting method, the inputted detail was insufficient hence the service returns the best option, *machine-based drill planting*, due to its superior in managing distance in rice planting and better yield rate. The automated recommendation was compared to the human expert suggestion and found that they are the same, and five human experts apparently agreed with the given reasoning.

The second scenario is different from the first. This scenario contains land location, planting time, preferred seed, and available instruments including tools and a list of fertilizers. Hence, the system does not suggest for the seed selection and method, but how to do the rice production properly. The given setting is as follow:

```
Suggesting Seed: RD-1
   Concerns:
      weak against Disease
          • Bacterial Leaf Blight
          • Ragged Stunt Disease
          • Narrow Brown Spot
          • Yellow Orange Leaf
          • Rice Blast
      weak against Pest
          • Brown Plant Hopper
          • Rice Stem Borer
          • Rice Gall Midge
      normal height : 115 cm.
      normal growth period until harvesting : 130 days
Suggesting Planting Method: Machine-based Drill Planting
```

Fig. 6. A generated solution from the first scenario

- Land location: *Phan district, Chiang Rai province*
  - Area size: 10 Rai (16,000 m^2)
- Time: *1st February*
- Preferred rice seed: *Niaw San-pah-tawng*
- Available instrument
  - Owned water storage, irrigation available
  - Rake
- Fertilizer: Chemical Fertilizer, customizable

With the given setting, the service returns solutions shown in Fig. 7.

From second recommendation, a list of instructions and concerns was generated from our knowledge base. According to land properties and farmer preferred rice seed, specific insecticide was given. Based on a highland paddy field type and given tools, planting method was selected. From expert opinion of the recommendation, 4 of 5 rice experts impressed the entire returned solution while an expert questioned on suggested planting method. The question is that *pre-germinated broadcasting* is normally not suggested to highland farming since this method has a risk on water control and may lead to failure in product yield due to uncontrollable rainfall rate. However, the system apparently has a rule concerning this issue, but with user information of available irrigation and owned water storage, the lack of water source is not in concern.

Given Seed: ***Niaw San-pah-tawng*** #Concerns:

Disease Prevalence in area against the given seed

- *Yellow Orange Leaf*

  Cause
  - carrier (green rice leafhopper)

  Preparation
  - weed elimination for every 2 weeks
  - insecticide: 1) dinotefuran 1% (2g. per sprout) or 2) etofenprox 20% (40cc./rai)

Pest Prevalence in area against the given seed: none

Planting Method: **pre-germinated broadcasting with sowing machine**

Land Preparation before planting
- remove weed root from a paddy field
- plow a paddy field roughly with rake and flood the soil for 5 cm. for 20 days
- after first plow for 20 days, plough to create regular furrows while flooded and redo after 7 days

Fertilizing after planting
- apply fertilize16-16-8 (30kg./rai) after broadcasting for 20 days
- after first fertilizing for 90 days, apply fertilize 46-0-0 (10kg./rai)

Pest and Disease Control
- apply etofenprox 20% (40cc./rai) after broadcasting seeds for 20 days, redo every 40 days

Harvesting
- release water before harvesting for 10 days

Fig. 7. A generated solution from the second scenario

4 Experiment

To test a potential of the system, we selected the cases inputted from the users as our scenarios. The seed suggestion however is the only function that users have executed. Hence, the testing cases from users are all the seed suggestion. Ten cases from anonymous users were gathered and applied to the system. The results were compared with the suggestion of three human experts for approving the correctness. In case of different answers from human experts, the only answers agreed by the three experts are listed as the result. The ten cases are given in Table 5 and a comparison of results from the proposed system and human experts are shown in Table 6, respectively. The returned answers were calculated into a precision and a recall score to measure the results of the system.

Table 5. Cases with conditions from users as testing scenarios

Case no.	Area (Tambon, Amphoe, Province)	Rice type	Month to start rice farming	Specific setting
1	Chumsaeng, Nang rong, Buriram	Normal	April	Irrigation, pond, invasion of brown plant hopper
2	Phra Non, Nakhon Luang, Ayuddhaya	Glutinous	October	Irrigation
3	Cheongdoi, Doysaket, Chiang mai	Normal	January	Spreading of rice blast disease, foothill slope
4	Chaiyo, Chaiyo, Ang Thong	Normal	May	Irrigation, spreading of rice blast disease
5	Ban Sa, Samchuk, Suphanburi	Normal	November	Irrigation, pond, spreading of bacterial blight disease and yellow orange leaf disease invasion of brown plant hopper
6	Chiang-saen, Chiang rai	Normal	December	Irrigation, pond, spreading of yellow orange leaf disease and invasion of brown plant hopper
7	Ayutthaya	Glutinous	June	Rainfall, spreading of bacterial blight disease and rice blast disease. Invasion of brown plant hopper and rice gall midge
8	Ronpiboon, Nakhon Si Thammarat	Normal	April	Rainfall, lowland (flood), spreading of bacterial blight disease
9	Jedeekham, Chiangkham, Phayao	Glutinous	November	Irrigation, pond, spreading of rice blast disease, bacterial blight disease
10	Banna, Nakhonnayok	Normal	May	Rainfall, lowland (flood), spreading of brown spot disease, acid soil

Table 6. Suggestion results comparing to experts and a precision-recall score

Case no.	System recommendation	Human expert recommendation	Precision	Recall
1	RD-7, RD-9, Chai Nat-1, Pathum Thani-1, Suphan Buri-60, Suphan Buri-1, Phitsanulok-2	Chai Nat-1, Pathum Thani-1	0.33	1.00
2	RD-2, RD-14	RD-2, RD-14, Sakon Nakhon	1.00	0.67
3	RD-33, Phitsanulok-2, Pathum Thani1, RD-39, RD-59	RD-33, Phitsanulok-2, Pathum Thani-1	0.60	1.00
4	RD-31, Chai Nat-1, Khao Tah Haeng-17, Khao Pahk Maw-148, Khao′ Jow Hawm Phitsanulok-1, Nahng Mon S-4, Pathum Thani-1	RD-31, Chai Nat-1, Pathum Thani-1	0.43	1.00
5	RD-27, RD-31, Pathum Thani1, Gow Ruang-88, Khao Tah Haeng-17, Khao Pahk Maw-148, Phitsanulok 60-1, Suphan Buri-1	RD-31, Pathum Thani-1, Suphan Buri-1	0.38	1.00
6	Chai Nat-1, Pathum Thani-1, Phitsanulok-2, Suphan Buri-1, Suphan Buri-3, Suphan Buri-4, Suphan Buri-90	Phitsanulok-2, Suphan Buri-1, Suphan Buri-3	0.43	1.00
7	RD-4, Niaw Dam Mor-37, Niaw Kiaw Ngoo-8974,Niaw Ubon-1, Ta - pow Gaew-161, Leb Meu Nahng-111, Khao Bahn Nah-432	RD-4, Niaw Dam Mor-37, Niaw Kiaw Ngoo-8974, Niaw Ubon-1	0.57	1.00
8	RD-17	RD-17	1.00	1.00
9	RD-2, RD-4, RD-10, RD-14, Sakon Nakhon, San-pah-tawng-1	RD-10, RD-14, San-pah-tawng-1	0.50	1.00
10	RD-19, Hantra-60, Prachin Buri-1, Prachin Buri-2, Khao Bahn Nah-432, Leb Meu Nahng- 111, Ayutthaya-1	RD-19, Leb Meu Nahng-111	0.33	1.00
Average			0.56	0.97

5 Discussion

By the use of ontology as knowledge base, the hierarchy and properties of ontology play an important role in suggesting the result. For example, the property of *invulnerability of disease* related to *seed* class can help to avoid undesired seed that would be weak to the disease prevalent in the input area. Moreover, knowing that the type of disease such as virus or bacterial type could help to scope the suggestion to medicine preparation and prevention method suggested in farm management.

From the experiment result, we found that the recommendation of seed selection has a high score of recall as 0.97 since most of the suggested seed types matched the suggestions of the experts. However, the precision score is 0.56 since the system

returned several answers that were not mentioned by the experts. In some cases, users clearly defined the area details such as case#1–3 while some cases, for example case#7, user only provided the province without giving an exact information of tumbon (sub-province), in which will reduce an ability of the system since the condition is vague.

From the in-sight observation, we found interesting details. Although the precision score is not much high, the alternative answers of the system are acceptable for the user given conditions and approved to be usable in those conditions by the experts. The reason which these seed types were not mentioned by the experts is that they are not the regularly used seed types nowadays. Hence, the precision score of the result does not show that the system is poorly designed, but it can provide a great variety of seeds in which are forgotten or unranked from experts' intuition.

For the recall score, the system performed well to cover most of the answers of the experts. Unfortunately, there was the case (case#2) that the system cannot return the mentioned answer. Upon consulting with the experts, they revealed that the reason they chose "Sakon Nakhon" seed because the given area condition (Phranon, Supanburi) is the "high land" which is a specific information to the local people and the seed is famous to be cultivated in the area. This information, however, did not input to the system, and as a result, the system did not return the seed even though the seed has the property for the high land cultivation in the ontology.

From the overview of the experiment, the system showed a potential to suggest seed that matches a given condition acceptably. However, the system can be improved more if the knowledge of the area such as terrain type and in depth soil information are given to the knowledge base. Moreover, the results will be more reliable and sound if the seeds can be ranked based on the previous and commonly usages in the area.

6 Conclusion and Future Work

In this paper, a framework of a recommendation system for rice farmer using agriculture knowledge represented in ontology is presented. Concepts in rice production and their relation are carefully captured in the designed ontology that is a knowledge blueprint to the system. The system is designed to give a solution based on individual farmers; hence, users need to submit their specification of contexts and needs. As an expert system, the proposed recommendation applies designed rules to generate a solution and the rules are updateable without a change in the main program. Based on the variety of rules assigned to system, several types of suggestions can be asked. To test a potential of the system, two different cases were given. The returned solutions were approved and satisfied by five experts in rice production.

In the future, we plan to add more plants to cover a majority of Thai farming. More rules will be added to system to cover other aspects in rice domain such rice milling and packaging. A cost reduction module according to Sufficiency Economy philosophy is in our plan to recycle a waste in cultivation to another process.

References

1. Walisadeera, A.I., Ginige, A., Wikramanayake, G.N.: User centered ontology for Sri Lankan farmers. Ecol. Inform. **26**, 140–150 (2015)
2. Ahsan, M., Motla, Y.H., Asim, M.: Knowledge modeling fore-agriculture using ontology. In: International Conference on Open Source Systems and Technologies (ICOSST). IEEE (2014)
3. Fujimoto, R., Aoyama, M.: A lifecycle-based design methodology of lightweight ontology and its application to cultivating high quality mandarin orange. In: IEEE 2014 21st Asia-Pacific Software Engineering Conference (2014)
4. Shyamaladevi, K., Mirnalinee, T.T., Trueman, T.E., Kaladevi, R.: Design of ontology based ubiquitous web for agriculture—a farmer helping system
5. Daiyi Li, A., et al.: An ontology-based knowledge representation and implement method for crop cultivation standard. Math. Comput. Model. **58**, 466–473 (2013)
6. Walisadeera, A.I., Ginige, A., Wikramanayake, G.N.: Conceptualizing Crop Life Cycle Events to Create a User Centered Ontology for Farmers, Part V, LNCS, vol. 8583, pp. 791–806. Springer (2014)
7. Mizoguchi, K.: Tutorial on ontological engineering. Tutorial on ontological engineering—Part 1: Introduction to ontological engineering. In: Mizoguchi, R. (eds.) New Generation Computing, vol. 21, no. 4, pp. 365–384. OhmSha&Springer (2003)
8. Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services
9. Rice Production Potential in Thailand, Rice Department of Thailand (2014) (in Thai)
10. GAP: Good Agriculture Practice, Bureau of Rice Research and Development Department, Rice Department of Thailand (2001) (in Thai)
11. Kim, H., Cho, Y., Yoe, H.: Intelligent Livestock Farm Management for Context Data Model Design. GST/SIA 2012, CCIS 338, pp. 291–296 (2012)
12. Shyamaladevi, K, Mirnalinee, T.T., Trueman, T.E., Kaladevi, R: Design of Ontology Based Ubiquitous Web for Agriculture—A Farmer Helping System
13. Antonopoulou, E., Karetsos, S.T., Maliappis, M., Sideridis, A.B.: Web and mobile technologies in a prototype DSS for major field crops. Comput. Electron. Agric. **70**, 292–301 (2010)
14. Wang, Y., Wang, Y., Wang, J., Yuan, Y., Zhang, Z.: An ontology-based approach to integration of hilly citrus production knowledge. Comput. Electron. Agric. **113**(2015), 24–43 (2015)
15. Sansri, B., et al.: Technical efficiency of Khao Dawk Mali 105 rice production in Roi-Et province. Khon Kaen Agric. J. **42** (Suppl. 1) (2014)
16. Hozo ontology editor. http://www.hozo.jp
17. Buranarach, M., Thein, Y.M., Supanithi,T.: OAM: an ontology application management framework for simplifying ontology-based semantic web application development. Int. J. Softw. Eng. Knowl. Eng.

A Knowledge-Based Approach for Generating UML Sequence Diagrams from Operation Contracts

Nasith Laosen(✉) and Ekawit Nantajeewarawat

School of Information, Computer and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University, Bangkok, Thailand
nasith@gmail.com, ekawit@siit.tu.ac.th

Abstract. Object-interaction design in a software development process is a complex task requiring comprehensive object-oriented design knowledge. An inexperienced system analyst often spends a lot of time and extensive efforts to achieve a good design. This paper presents a knowledge-based framework for facilitating object-interaction design, taking into account several commonly occurring design issues, including realization of associations and compositions, handling object creation, and determining appropriate object access paths. Along with a domain model, requirement artifacts specified in the form of operation contracts are encoded using Web Ontology Language. Generic object-interaction design solutions are represented as Jena rules, which can be instantiated into many domain-specific application contexts. Application of the framework to generation of object-interaction designs, represented as sequence diagrams, is illustrated. An evaluation schema is proposed and an evaluation example is demonstrated.

Keywords: Object interaction design · Operation contract · Sequence diagram · Rule-based system

1 Introduction

In object-oriented software development, object-interaction design determines how objects in a system interact with each other in order to complete a required task. In the Unified Process [2], input artifacts required for such design work include use cases, a domain model (a class diagram), and operation contracts. A use case determines a number of system operations, each of which represents a system event generated by an actor. An operation contract declaratively describes detailed changes to objects as the expected result of performing a system operation. Such changes are specified in terms of the postconditions of the contract, with their basic categories being object creation, association forming, and attribute modification. An object-interaction design output is usually represented as a collaboration diagram or a sequence diagram.

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_12

Designing high-quality object interactions is a difficult task especially for a novice system analyst. Extensive object-oriented design knowledge is required for making appropriate design decisions, e.g., deciding how to realize an association or a composition relationship, how to select an appropriate object that is responsible for creating a new object, how to determine a path for exchange of messages among objects, and how to manage message parameters.

In this paper, we propose a semi-automatic knowledge-based approach to facilitate sequence diagram generation from operation contracts. Web Ontology Language (OWL) is used for encoding input artifacts. The General Responsibility Assignment Software Patterns (GRASP) [2] are employed as guidelines for designing generic design solutions. Jena rules [1], which can work directly with OWL and Resource Description Framework (RDF) data, are used for representing generic design solution templates. When all information required for making a particular design decision is available, these rules make the decision automatically. When necessary, the rules delegate a design decision to a system analyst with appropriate recommended choices. Through rule instantiation, the design solutions are applied in the contexts of domain-specific design problems. The proposed system requires basic skills of object-interaction design and OWL from its user.

This paper is organized as follows: Sect. 2 reviews related works. Section 3 gives an overview of the proposed framework. Section 4 describes rules representing design solution templates and rule-based generation of sequence diagrams. Section 5 demonstrates an application of the proposed framework. Section 6 shows an evaluation schema and an evaluation result. Section 7 provides conclusions.

2 Related Works

Most studies on generating sequence diagrams take use case descriptions as their main input. Such descriptions consist of textual sentences specifying flows of interactions between actors and systems. In [3–5], methods were proposed for writing flow-of-event sentences in a use case description using certain specific structural forms, which are designed to support extraction of messages together with their senders and receivers. From such structured sentences, output sequence diagrams are generated.

Instead of restricting structural forms for writing use case descriptions, Thakur and Gupta [8] employed an intelligent natural language processing tool, the Standford parser, to automatically construct from input flow-of-event sentences Part-of-Speech (POS) tagged text and Type Dependencies (TDs) representing grammatical relationships among words. Based on POS tags and TDs, transformation rules were proposed for identifying elements of a sequence diagram. The obtained elements are then used to create commands of the Sdedit modeling tool for diagram visualization.

None of [3–5,8] takes a domain-model class diagram into consideration. Since a use case description basically specifies interactions between actors and a system, rather than interactions among objects in a domain model, a sequence

diagram generated by the above works is likely to be a high-level diagram for specifying a flow of system operations in an early phase of a development process. By contrast, a sequence diagram considered in our work is a design-level diagram, which specifies how objects in a system interact in terms of method invocations and is much closer to an implementation level.

Sawprakhon and Limpiyakorn [6] presented a model-driven sequence-diagram generation approach that takes a class diagram, a use case description, and their metamodels as input artifacts. Along with a sequence-diagram metamodel, metamodel-level mapping rules are used for transforming input artifacts into elements of a sequence diagram. The output sequence diagram is visualized using the Sdedit modeling tool. A metamodel provides a specification for constructing a well-structured diagram rather than a requirement specification describing what changes should be made to the states of relevant objects. Compared to [6], an operation contract is used in our work for specifying precisely required changes in the level of objects' states, allowing us to generate a design-level sequence diagram with method-invocation details. Such a design-level diagram can be directly used for generating program code.

3 The Proposed Framework

Figure 1 shows an overview of the proposed framework, which is described below with reference to Steps (1)–(9) indicated in the figure.

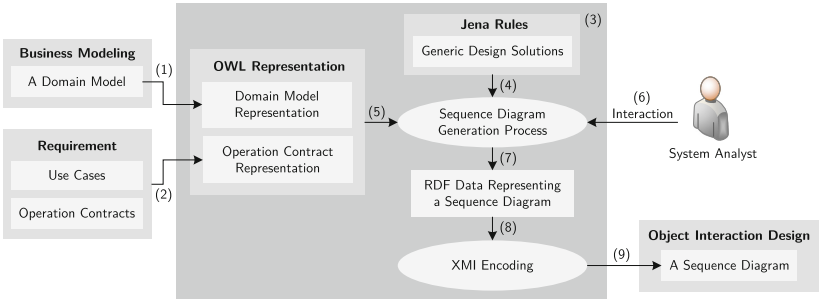


Fig. 1. The proposed framework

A domain model and requirement artifacts such as use case diagrams and operation contracts are encoded in OWL (Steps (1) and (2)). Generic design solutions are represented as Jena rules [1] (Step (3)). These rules provide a machine-processable solution template suggesting how to assign responsibilities to objects and how to design object interaction in order to fulfill the requirement specified by a given operation contract.

To apply such a solution template, variables in Jena rules are instantiated into the assertions describing input artifacts. Assertions representing the components

of an output sequence diagram are then generated by a RETE-based forward-chaining inference engine (Steps (4), (5), and (7)). During the inference process, a system analyst may interact with the rules to provide some necessary design decisions by using Java-based command-line user interface (Step (6)).

The generated assertions are translated into an XML Metadata Interchange (XMI) file (Step (8)). The obtained XMI file can be visualized in a graphical diagrammatic form by a standard UML modeling tool (Step (9)).

The generic design solutions used in this work are based on three basic patterns of GRASP, i.e., the controller pattern, the creator pattern, and the information expert pattern. Other design patterns can also be incorporated into this framework. Such incorporation is however beyond the scope of this paper.

4 Generic Design Solution Rules and Rule-Based Generation of Sequence Diagrams

Postconditions of an operation contract are the main drivers for generating a sequence diagram. We restrict our attention in this work to three types of postconditions, i.e., association forming, attribute modification, and object creation, which are the most commonly used postconditions [2].

The rules for generating a sequence diagram are divided into two main types, i.e., supporting rules and postcondition handling rules, which are introduced in Sects. 4.1 and 4.2, respectively. A rule-based process for generating a sequence diagram is described in Sect. 4.3.

4.1 Supporting Rules

Supporting rules are rules for (i) asserting essential RDF triples in preparation steps for generating a sequence diagram or (ii) performing certain special operations requested by postcondition handling rules. The supporting rules are classified into the following subtypes:

- *Association Realization* (*AR*) *Rules*: AR rules are used for realizing an association between two classes. Association realization describes how a particular association is implemented within the design model in terms of an attribute of a class. For example, an association R between classes X and Y could be implemented by adding an attribute referring to the class Y in the class X or vice versa. AR rules delegate a system analyst to make such a decision. The name of the attribute is also acquired from the system analyst. Suppose that R is implemented as an attribute of X, the multiplicity of R at the Y endpoint is used to determine the data type of the implemented attribute, i.e., if the maximum value of the multiplicity is one, then the data type is the class Y, otherwise the data type is a collection of instances of Y.
- *Composition Realization* (*CR*) *Rules*: By using CR rules, the composition relationship between a composite class X and a component class Y is automatically implemented as an attribute of X. The composition multiplicity at

the *Y* endpoint is again used for determining the data type of the implemented attribute. A system analyst is still required to provide the name of the implemented attribute.

- *Controller Creation* (*CC*) *Rules*: Based on the controller pattern, a controller object is the first object behind the user-interface layer. It is a receiver of system operation messages. CC rules are used for creating a controller class, a controller object, and a message invoking a system operation. CC rules generate the controller class name for a use case by concatenating the use case's name with the word "Handler", e.g., the controller class for the use case "SubmitDCS" is "SubmitDCSHandler".
- *Entry Path Creation* (*EPC*) *Rules*: EPC rules are used for finding objects and their classes along an access path, called *entry path*, from a controller object to a target object. The realization of associations and composition relationships obtained by using AR and CR rules are exploited for determining an entry class and an entry object for each successor element in an entry path. When there are multiple possible candidate entry classes, a system analyst will be requested to select one candidate. EPC rules are activated by other types of rules to create messages along the entry path with different message names and parameters.
- *Parameter Management* (*PM*) *Rules*: PM rules are responsible for assigning and managing parameters of messages that are used for fulfillment of a postcondition.
- *A Root Class Determination* (*RCD*) *Rule*: The RCD rule interacts with a system analyst to determine root classes, i.e., classes that can be accessed by a controller object.

The first column group in Table 1 shows the number of rules in these six subtypes.

Table 1. The number of rules in each subtype

Type	Supporting rules						Postcondition handling rules		
Subtype	AR	CR	CC	EPC	PM	RCD	AF	AM	OC
#Rules	8	3	3	44	4	1	40	1	12

4.2 Postcondition Handling Rules

Postcondition handling rules typically generate components of a sequence diagram for making changes to a state of the system in order to satisfy an input postcondition. They are divided into three subtypes based on postcondition categories as follows:

- *Association Forming* (*AF*) *Rules*: AF rules are used for creating a link connecting two target objects as an instance of an association or a composition. An object interaction sequence designed for creating such a link depends heavily on how the association or composition is realized (cf. AR and CR rules).

- *An Attribute Modification* (*AM*) *Rule*: According to the information expert pattern, a method m for modifying an attribute *attr* should be located in the class that owns *attr*. The AM rule assigns the method m to such a class and generate an invocation of m in a sequence diagram.
- *Object Creation* (*OC*) *Rules*: OC rules are responsible for fulfillment of object creation postconditions, based on the creator pattern of GRASP. To create an object of a given class, say X, OC rules first find a composite class that contains X as its component. If such a composite class exists, it is designated as the creator class for X. If it does not exist, the rules retrieve all classes that are associated with X and request a system analyst to select one to be the creator class for X. The rules then generate a message from an object of the creator class for creating an object of X. When the created object has an attribute whose type is an object collection, the empty object collection is also created for initializing that attribute.

The second column group in Table 1 shows the number of rules in each of the three subtypes.

4.3 A Rule-Based Process for Generating a Sequence Diagram

As outlined in Fig. 2, a process of sequence diagram generation is divided into two phases, i.e., an initial phase and a postcondition processing phase.

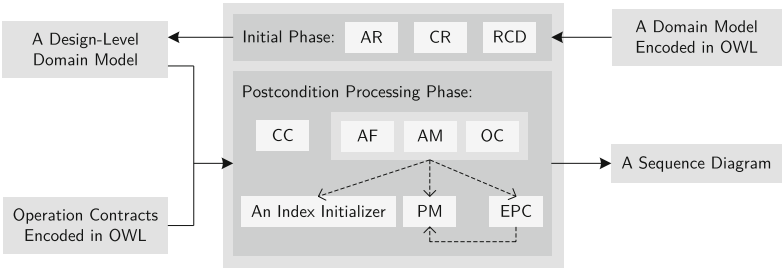


Fig. 2. The rule-based process for generating a sequence diagram

In the initial phase, essential RDF triples for generating sequence diagrams are prepared. These triples are created from a domain model encoded in OWL by using AR, CR, and RCD rules. In particular, by adding attributes for association and composition realization introduced by AR and CR rules, a design-level domain model is obtained. This model provides important input information for the next phase.

In the postcondition processing phase, one sequence diagram is generated for each operation contract. CC rules are fired for creating a message representing the system operation that is described by an input operation contract. AF, AM, and OC rules are then fired to construct parts of a sequence diagram

for satisfying postconditions of their respective types. These rules activate (i) an index initializer to obtain an initial index used for determining appropriate message ordering, (ii) PM rules for adding parameters to relevant messages, and (iii) EPC rules for creating messages along an entry path to a relevant target object. Dashed arrows in Fig. 2 depict "activate" relations between rule groups.

5 Example Application

Based on our implemented prototype tool, application of the proposed framework to a university registration system adapted from [7], is illustrated. During designing an object-oriented system, when complex problems concerning object-interaction design are broken down, basic problems are obtained. Basic problems obtained from different systems are similar, i.e., they are problems about the changes to be made to objects (cf. Sect. 1). Hence, in order to make the example application easy to understand, the simple registration system is used.

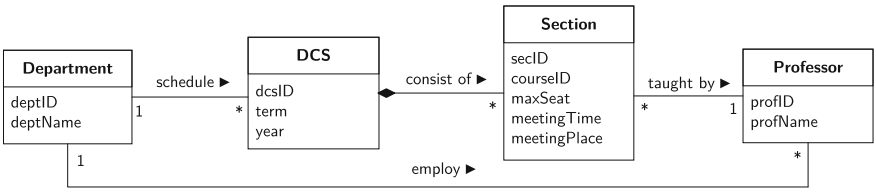


Fig. 3. A partial domain model of a university registration system

Operation:	enterDCS(deptID, dcsID, term, year)
Cross Reference:	SubmitDCS
Precondition:	none
Postcondition:	1. A DCS instance dcs was created (object creation). 2. dcs was associated with a Department, based on deptID (association formed).

(a)

Operation:	enterSection(secID, courseID, maxSeat, meetingTime, meetingPlace, profID)
Cross Reference:	SubmitDCS
Precondition:	dcs exists and an entry path to dcs has been established.
Postcondition:	1. A Section instance s was created (object creation). 2. s was associated with dcs (association formed). 3. s was associated with a Professor, based on profID (association formed).

(b)

Fig. 4. Operation contracts of the enterDCS and the enterSection system operation

A partial domain model of the registration system is shown in Fig. 3. This registration system provides a use case SubmitDCS, which is used by a department officer to create a list of sections to be offered, called a department class schedule (DCS). This use case consists of two system operations, i.e., enterDCS and enterSection. The specifications of these system operations are given by operation contracts in Fig. 4.

5.1 An Initial Phase

In this phase, AR, CR, and RCD rules are used. To illustrate an example, an AR rule is given in Fig. 5, where a variable begins with "?", and prefixes "base" and "rdf" indicate a namespace of each element in a RDF triple.

The AR rule in Fig. 5 asks a system analyst how to realize an association by using a custom built-in function, named associationRealization, and asserts necessary RDF triples describing details of the realization. For instance, a system analyst is asked to realize the association schedule between classes Department and DCS. Assume that the system analyst specifies that the association schedule is implemented as an attribute, named dcsList, in Department. Along with an individual representing the new attribute dcsList, other RDF triples describing the realization of the association schedule are added to the knowledge base by this rule.

After AR, CR, and RCD rules are activated, the design decisions made in this phase are as follows: The associations employ, schedule, and taught-by are implemented as an attribute profList in Department, an attribute dcsList in Department, and an attribute prof in Section, respectively. The composition consist-of is implemented as an attribute secList in DCS. The class Department is set to be a root class.

```
[AR1:
    (?asso rdf:type base:Association) (?asso base:ClassAtEndPoint1 ?cep1) (?asso base:ClassAtEndPoint2 ?cep2)
    (?asso base:MultiplicityAtEndPoint1 ?mep1) (?asso base:MultiplicityAtEndPoint2 ?mep2) (?asso base:hasName ?aName)
    noValue(?asso base:implementedAs ?anAttribute) generateID(?newAttribute) generateID(?ari)
    associationRealization(?cep1, ?cep2, ?mep1, ?mep2, ?aName, ?attributeName)
    ->
    (?newAttribute rdf:type base:Attribute) (?newAttribute base:hasAttributeName ?attributeName)
    (?asso base:implementedAs ?newAttribute) (?cep1 base:hasAttribute ?newAttribute) (?cep1 base:knows ?cep2)
    (?ari rdf:type base:AssociationRealizationItem) (?ari base:keeperClass ?cep1) (?ari base:keptClass ?cep2)
    (?ari base:implementedAttribute ?newAttribute)
]
```

Fig. 5. An AR rule

5.2 A Postcondition Processing Phase

After the basic RDF triples are asserted by supporting rules in the previous phase, a sequence diagram of each system operation is generated sequentially, according to the order of system operations.

5.2.1 The enterDCS System Operation To begin with, CC rules are used for applying the controller pattern (cf. CC rules). The CC rule shown in Fig. 6 asserts individuals representing a controller object and a system operation message for the system operation specified by an operation contract. When applied to the operation contract in Fig. 4a, this rule generates the components represented by the individuals i1 and i2, of the sequence diagram in Fig. 7. The controller class and parameters of i2 are generated by other CC rules.

```
[CC1:
    (?opc rdf:type base:OperationContract) (?opc base:forSystemOperation ?so) (?controller rdf:type base:Controller)
    noValue(?controllerObj base:belongsTo ?controller) generateID(?objID) generateID(?msgID)
    ->
    (?objID rdf:type base:Object) (?objID base:belongsTo ?controller) (?msgID rdf:type base:CallMessage)
    (?msgID base:hasReceiver ?objID) (?msgID base:forOperation ?so) (?msgID base:hasIndex 1)
]
```

Fig. 6. A CC rule

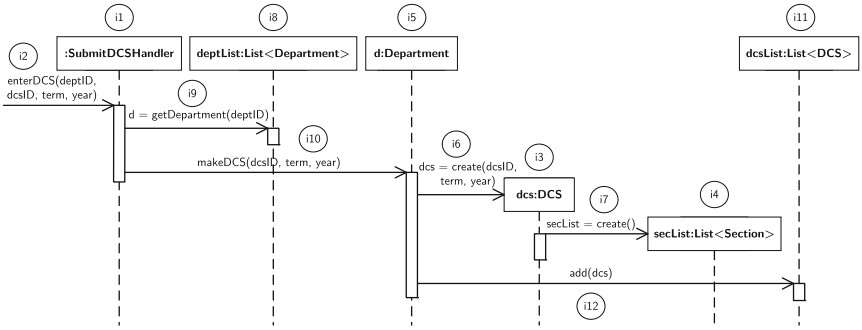


Fig. 7. The output sequence diagram of the enterDCS system operation

```
[OC1:
    (?opc rdf:type base:OperationContract) (?opc base:hasPostCondition ?oc) (?opc base:forSystemOperation ?so)
    (?oc rdf:type base:ObjectCreation) (?oc base:InstanceOf ?classOfobjToBeCreated) (?oc base:ObjectCreated ?objName)
    generateID(?objToBeCreated) getMessageIndex(?msgIndex)
    makeTemp(?initializeAttribute) makeTemp(?requestCreatorClass)
    ->
    (?objToBeCreated rdf:type base:Object) (?objToBeCreated base:belongsTo ?classOfobjToBeCreated)
    (?objToBeCreated base:hasObjectName ?objName)
    (initializeAttribute rdf:type InitializeAttribute(?objToBeCreated))
    (?requestCreatorClass rdf:type RequestCreatorClass(?classOfobjToBeCreated, ?objToBeCreated, ?so, ?msgIndex))
]
```

Fig. 8. An OC rule for creating an object

The first postcondition in Fig. 4a specifies the creation of a new object, named dcs, belonging to the class DCS. It activates the OC rule in Fig. 8. The rule in Fig. 8 generates the object dcs, which is represented by the individual i3. By using a custom built-in function, named getMessageIndex, a starting index used for fulfilling this postcondition is obtained from the index initializer. The rule in Fig. 8 also asserts an InitializeAttribute and a RequestCreatorClass assertion.

The InitializeAttribute-assertion is asserted to activate other OC rules to generate empty-collection objects. By considering the realized association of the class DCS, an empty-collection object of the class Section, which is represented by the individual i4, is created.

The RequestCreatorClass-assertion is asserted to trigger other OC rules that are responsible for finding a creator class and a creator object of dcs. Since the class Department is the composite class of the class DCS, those OC rules automatically assign Department to be the creator class and generate a creator

object. The generated creator object d is represented by the individual i5. The object name "d" is obtained from the system analyst. After the creator object is generated, a GenerateCreatingMsg-assertion is also asserted for activating the OC rule in Fig. 9.

```
[OC2:
    (?generateCreatingMsg rdf:type GenerateCreatingMsg(?creatorClass, ?creatorObj, ?classOfobjToBeCreated,
    ?objToBeCreated, ?so, ?msgIndex))

    (?classOfobjToBeCreated base:hasClassName ?cName) (?objToBeCreated base:hasObjectName ?objName)
    strConcat("make",?cName,?makeOperationName) difference(?msgIndex, 1, ?msgIndex1)
    generateID(?creatingMsg) generateID(?parameterSet) makeTemp(?requestEntryPath)
    makeTemp(?requestAddingParameters) makeTemp(?generateCreatingMsgForInitializingAttribute)
    ->
    (?creatingMsg rdf:type base:CreatingMessage) (?creatingMsg base:hasSender ?creatorObject)
    (?creatingMsg base:hasReceiver ?objToBeCreated) (?creatingMsg base:hasIndex ?msgIndex1)
    (?creatingMsg base:hasReturnedValue ?objToBeCreated) (?parameterSet rdf:type base:SetOfParameter)
    (?objToBeCreated base:hasCreatorObject ?creatorObject) (?objToBeCreated base:hasEntryObject ?creatorObject)

    (?requestAddingParameters rdf:type RequestAddingParameters(?classOfobjToBeCreated, ?creatingMsg, ?so, ?parameterSet))

    (?generateCreatingMsgForInitializingAttribute rdf:type GenerateCreatingMsgForInitializingAttribute
    (?classOfobjToBeCreated, ?objToBeCreated, ?msgIndex1))

    (?requestEntryPath rdf:type RequestEntryPath(?creatorClass, ?creatorObject, ?so, ?makeOperationName, ?msgIndex1,
    ?parameterSet))
]
```

Fig. 9. An OC rule for generating a creating message

The rule in Fig. 9 generates a creating message represented by the individual i6, with the object d being its sender and the object dcs being its receiver. This rule also asserts a RequestAddingParameters, a GenerateCreatingMsgForInitializingAttribute, and a RequestEntryPath assertion into the knowledge base.

The asserted RequestAddingParameters-assertion triggers PM rules for adding parameters to the creating message i6. The asserted GenerateCreatingMsgForInitializingAttribute-assertion activates other OC rules to generate a creating message for initializing an object-collection attribute. As a result, the individual i7 is generated. The RequestEntryPath-assertion is asserted for activating EPC rules to establish an entry path of the creator object d. Since Department is a root class, the controller object can access the list of Department. By using this information, the EPC rules generate individuals i8, i9, and i10. At this point, an entry path to dcs is also established.

The second postcondition in Fig. 4a specifies the association forming between dcs and an object of Department. The occurrence of dcs triggers AF rules to fulfill this postcondition. To start with, an AF rule asks the system analyst to specify an object of Department to be associated. Assume that the system analyst specifies the object d, another AF rule is then activated to generate the individual i11 and assert an AssociationRequest-assertion for activating the AF rule in Fig. 10. The rule in Fig. 10 generates the call message i12.

```
[AF1:
    (?associationRequest rdf:type AssociationRequest(?side1Class, ?side2Class, ?side1Object, ?side2Object))

    (?ari rdf:type base:AssociationRealizationItem) (?ari base:keeperClass ?side1Class) (?ari base:keptClass ?side2Class)
    (?ari base:implementedAttribute ?attribute) (?attribute rdf:type base:List) (?attribute base:pointToClass ?listClass)
    (?listClass base:isaListOf ?side2Class) (?side2Object base:hasCreatorObject ?side1Object)
    (?side1Object base:hasListComponentObject ?listObj) (?listObj base:belongsTo ?listClass)
    (?side2Class base:hasClassName ?cName) (?creatingMsg rdf:type base:CreatingMessage)
    (?creatingMsg base:hasSender ?side1Object) (?creatingMsg base:hasReceiver ?side2Object)
    (?creatingMsg base:hasIndex ?msgIndex) strConcat("add", ?operationName) addOne(?msgIndex, ?msgIndex1)
    generateID(?callMsg) generateID(?opID)
    ->
    (?opID rdf:type base:Operation) (?opID base:hasOperationName ?operationName)
    (?callMsg base:forOperation ?opID) (?callMsg rdf:type base:CallMessage) (?callMsg base:hasSender ?side1Object)
    (?callMsg base:hasReceiver ?listObj) (?callMsg base:hasIndex ?msgIndex1) (?callMsg base:hasParameter ?side2Object)
]
```

Fig. 10. An AF rule

5.2.2 The enterSection System Operation By applying all rules together with design decisions obtained from fulfilling the enterDCS system operation, to the operation contract in Fig. 4b, an output sequence diagram of the enterSection system operation is generated as shown in Fig. 11.

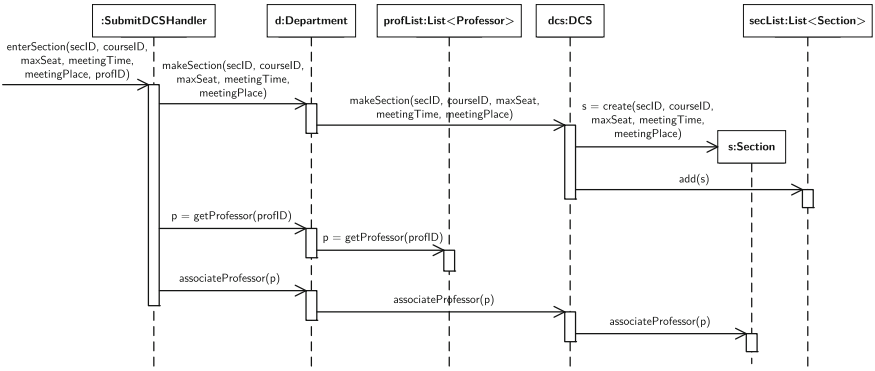


Fig. 11. The output sequence diagram of the enterSection system operation

6 Evaluation

6.1 An Evaluation Schema

An output sequence diagram is evaluated by (i) using a human expert to check whether all postconditions of a given operation contract are fulfilled, (ii) checking whether the receiver and parameters of a message are visible to their sender, and (iii) justifying reasons for sending message parameters.

Let m be a message that is sent by an object a, and b be the receiver of m or a parameter of m. Based on Larman [2], b is visible to the sender a if one of the following conditions holds:

- *Attribute Visibility* (AV): b is the value of an attribute of a. (AV may exist through association/composition realization or establishment of a connection between a controller class and a root class.)
- *Parameter Visibility* (PV): b is sent to a as an input parameter of some method m' invoked on a. (PV temporarily exists only within the scope of m'.)
- *Local Visibility* (LV): Within a method m' of a, b is created, computed, or returned to a as a result of some method invocation. (LV temporarily exists only within the scope of m'.)
- *Global Visibility* (GV): The scope of b covers the scope of a.

Reasons for sending a parameter p of a message m are also used for evaluation of m. They include:

- $R1$: p is stored as an attribute of the receiver of m.
- $R2$: p is used by the receiver of m, e.g., for some calculation, for object retrieval, etc.
- $R3$: p is passed as a parameter of a message invoked within the activation (execution specification) of m.

6.2 An Evaluation Example

To illustrate an application of the evaluation schema, the output sequence diagram of enterDCS system operation in Fig. 7 is considered. Note that (i) the system operation message (i2) is excluded from the evaluation scope, and (ii) for an object creation message (e.g., i6), only its parameters are evaluated.

Table 2 shows receiver/parameter visibility for messages in Fig. 7. For example, the message i9 with a parameter deptID (sent from the controller object to deptList) is well-formed with respect to receiver/parameter visibility, since (i) deptList is visible to the controller object through the establishment of a connection between the controller class SubmitDCSHandler and the class Department, and (ii) deptID is parametrically visible to the controller object within the scope of the method enterDCS. The scope of a method is depicted in terms of an activation in a sequence diagram. According to Table 2, the receiver and parameters of every message are visible to their sender.

Table 3 shows reasons for sending message parameters in Fig. 7. For instance, the parameter deptID of the message i9 is sent from the controller object to deptList because the receiver deptList needs deptID to retrieve the object d.

The output sequence diagram shown in Fig. 7 is considered to be correctly generated since (i) the postconditions of the enterDCS system operation are fulfilled, i.e., the object dcs is created and dcs is associated with an object of class Department, (ii) every message in Fig. 7 is well-formed with respect to receiver/parameter visibility (cf. Table 2), and (iii) reasons for sending message parameters can be justified (cf. Table 3). By using the proposed evaluation schema, the output sequence diagram in Fig. 11 is also considered to be correctly generated.

Table 2. Receiver/parameter visibility for messages in Fig. 7

Message	Sender	Receiver/Parameter	Types of visibility			
			AV	PV	LV	GV
i9	Controller	deptList	✓			
		deptID		✓		
i10	Controller	d			✓	
		dcsID		✓		
		term		✓		
		year		✓		
i6	d	dcsID		✓		
		term		✓		
		year		✓		
i12	d	dcsList	✓			
		dcs			✓	

Table 3. Reasons for sending message parameters in Fig. 7

Message	Parameter	Reasons for parameter passing		
		$R1$	$R2$	$R3$
i9	deptID		✓	
i10	dcsID			✓
	term			✓
	year			✓
i6	dcsID	✓		
	term	✓		
	year	✓		
i12	dcs		✓	

7 Conclusions

In this paper, a knowledge-based framework for generating design-level sequence diagrams is presented. A domain model and operation contracts are encoded using OWL. Generic design solutions for addressing common design problems are represented as Jena rules. They are devised based on the controller pattern, the creator pattern, and the information expert pattern. Types of rules and their responsibilities are described. Application of the proposed framework is demonstrated based on our own developed prototype system. By instantiating the rules in an application context, domain-specific sequence diagrams are generated. During the generation process, a system analyst may communicate with the rules to provide additional information required for making a particular design decision. A schema for evaluation of output sequence diagrams is proposed. Future works include (i) incorporation of other design patterns, e.g., Gang-of-Four design patterns, into the current design solution knowledge base, and (ii) implementation of a practicable CASE tool based on the proposed framework.

Acknowledgements. This work has been supported by (i) the Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Thammasat University, and (ii) the Intelligent Informatics and Service Innovation (IISI) Center, Sirindhorn International Institute of Technology (SIIT), Thammasat University.

References

1. Apache jena. https://jena.apache.org
2. Larman, C.: Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development, 3rd edn. Prentice Hall PTR, Upper Saddle River, NJ, USA (2004)

3. Laura Méndez, M., Romero, C., Pérez Herrera, K.: UML sequence diagram generator system from use case description using natural language. In: 2007 Electronics, Robotics and Automotive Mechanics Conference, CERMA 2007, pp. 360–363, Sept 2007
4. Li, L.: Translating use cases to sequence diagrams. In: Proceedings ASE 2000. The Fifteenth IEEE International Conference on Automated Software Engineering 2000, pp. 293–296 (2000)
5. Mason, P., Supsrisupachai, S.: Paraphrasing use case descriptions and sequence diagrams: an approach with tool support. In: 2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2009, vol. 02, pp. 722–725, May 2009
6. Sawprakhon, P., Limpiyakorn, Y.: Model-driven approach to constructing UML sequence diagram. In: 2014 International Conference on Information Science and Applications (ICISA), pp. 1–4, May 2014
7. Stumpf, R., Teague, L.C.: Object-oriented Systems Analysis and Design with UML. Pearson/Prentice Hall (2004)
8. Thakur, J.S., Gupta, A.: Automatic generation of sequence diagram from use case specification. In: Proceedings of the 7th India Software Engineering Conference ISEC'14, pp. 20:1–20:6. ACM, New York, NY, USA (2014)

Designing a Social Innovation Based Knowledge Support System: A Preliminary Guideline

Ahmet Suerdem(✉), Baskin Yenicioglu, and Ozge Demir

Faculty of Business, Istanbul Bilgi University, Istanbul, Turkey
{ahmet.suerdem, baskin.yenicioglu, ozge.demir}@bilgi.edu.tr

Abstract. Policy makers and funding agencies increasingly emphasize the social nature of innovation. However, focusing just on the product side of social innovation might easily reduce the concept to commercialization of social goods. Engagement of all stakeholders in the knowledge co-creation process makes innovation really social. The aim of this paper is to highlight the essential elements of a social value creation based support system that would engage all stakeholders to the innovation process. To explore this guideline, we conducted a case-study research using qualitative data collected from diverse stakeholders of an NGO involved in early age child education in Turkey. Our findings suggest that social innovation based support systems require an architecture bridging the coordination of both online collaboration and knowledge management tools to the offline communities of practice. The most important challenge to this architecture is integrating piecemeal tools and practices into a social ecological system.

Keywords: Social innovation · Innovation support systems
S-LCA PLM Knowledge governance

1 Introduction

We are witnessing a change in the traditional individual value creation focused understanding of innovation, as the economic, social, and technological systems become more globally interconnected. Accordingly, innovation is redefined to emphasize the social aspects of value creation such as: the global challenges as the driver of innovation; co-creation of value with stakeholders; gathering and managing knowledge about users; knowledge sourcing through collaborative networks and innovating public goods as well as private goods [1]. This redefinition implies a more integrative approach to innovation support systems, which should reconsider innovation not only as an economic mechanism or a technical process but as a social phenomenon by its purpose, effects, or methods. Such a system requires integrating people, data, processes and business systems to provide an information backbone for new product development.

Considering innovation's relationship with its context brought about Life Cycle Assessment (LCA) philosophy to new product development, which is increasingly

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_13

being used to evaluate the impact of product systems on environmental, economic, and social contexts from idea generation to manufacturing, distribution, use, and post-use stages [2]. Aiming to integrate people, processes, tools and technology for managing all relevant information and practices through the entire life cycle of a product [3] LCA inspired some collaborative software applications for providing a support system to new product development. Gathered under the umbrella term PLM (Product Life Cycle Management), these applications maintain an "information authoring" platform that integrate computer-based tools to harvest data from various stakeholders throughout the product life cycle and build these into the product design and development [4, 5].

However, present PLM applications, until recently, remained limited to collecting information about user requirements and building this into the product design, thus, supporting technological than social value creation. To be in line with the social nature of innovation, these applications need to go beyond supporting one way information flow from stakeholders to the business. They need to integrate Responsible Research and Innovation (RRI) principles to their systems to observe the social nature of the innovation as a more transparent, deliberative and interactive process. RRI is a "comprehensive approach of proceeding in research and innovation in ways that allow all stakeholders that are involved in the processes of research and innovation at an early stage (A) to obtain relevant knowledge on the consequences of the outcomes of their actions and on the range of options open to them and (B) to effectively evaluate both outcomes and options in terms of societal needs and moral values and (C) to use these considerations (under A and B) as functional requirements for design and development of new research, products and services" [6]. In this respect, RRI encourages knowledge partnerships among stakeholders to bring together innovative and entrepreneurial activities with social initiatives through the innovation life cycle. Accordingly, PLM applications need to integrate solutions supporting a collaborative environment where knowledge partnerships can co-create and share knowledge in a participative manner. The role of PLM as an innovation support system in the future is expected to support social value creation not only for the company and customers but also for all stakeholders [7].

This paper aims to highlight the essential elements of a social value creation based support system that would engage all stakeholders to the innovation process. For this purpose, we developed a conceptual guideline and explored its implications through a case-study research using qualitative data collected from diverse stakeholders of an NGO involved in early age child education in Turkey. First, we discuss the concept of social innovation (SI) as an engaging and empowering knowledge co-creation process. Second we suggest that knowledge co-creation in SI needs to be organized as Distributed Networks of Knowledge (DNoK) to support local tacit knowledge integration. Third, we emphasize participative knowledge governance against top-down knowledge management for governing emergent and spontaneous interactions within DNoK. Then we discuss the potential of social lifecycle assessment (S-LCA) as the basis for an empowering and engaging knowledge governance tool to support SI. Finally, we conclude that encouraging distributed tacit knowledge co-creation requires building and governing a system of social ecology tools and procedures rather than focusing on creativity tools for managing individual product features. The paper then discusses findings of the case study exploring stakeholders' existing new product development

strategies, their perceptions, expectations, and interactions during existing collaboration activities, and their visions about the ecology of a collaborative knowledge support system.

2 Social Innovation as an Engaging and Empowering Knowledge Co-creation Process

As the nature of innovation is changing to become more social and policy makers encourage RRI practices, companies are starting to be increasingly involved with corporate social initiatives (CSIs) mostly through partnerships with Non-Governmental Organizations (NGOs). The concept of social innovation (SI) has started to sit comfortably at the juxtaposition of RRI and social responsibility projects of CSIs and NGOs. SI is commonly defined as new product, service, or process ideas aiming to meet social needs, which are predominantly developed and diffused through organizations whose primary purposes are social [8]. These innovations are considered both to have positive impact for public good and also to encourage greater societal involvement in social service offerings [9]. As such, SI is considered to be a normative tool to develop social goods and services to resolve social problems. However, this limited definition of SI runs the risk of turning it into another type of commercial production process, a buzzword to cover the production of all social goods. Besides profit making, corporates also use this buzzword as a convenient public relations tool to window-dress their effects on social injustices and environmental problems [10]. The definition of SI merely as bringing innovative solutions to social and environmental problems is problematic as it focuses on the product as an end in itself. Social innovations should be different than commercial innovations in several respects. Although commercial innovations can increase a social group's wellbeing, they are primarily profit oriented. In this sense, SI draws parallels with CSIs and NGO activities which do not aim profitability but to fulfill social responsibilities. However, innovative CSIs are usually performed as a form of public relations campaign or just plain corporate philanthropy which turns them into strategic business tools to advance commercial or political agendas [11]. NGOs on the other hand, although in principle they should be non-profit voluntary organizations, increasingly have started to become façades for businesses to benefit from tax exemption or government incentives. These efforts attract criticisms for carrying a hidden political, religious or other interest related agenda under the auspices of charity [12].

Although each having different intended purposes, socially oriented commercial innovations, innovative CSIs and NGOs all tend to consider stakeholders as passive receivers of a social product. Their notion of SI is product focused where the intended end is a prescribed solution to a social problem. They identify co-creation as the collection of product information and user requirements from the stakeholders and building this into the product features. Hence, their use of support systems is limited to idea and information generation tools, where co-created knowledge becomes the intellectual property of the organization and materializes as a commercial product or a charity to be passively consumed by the users. However, as SI aims to bring about social change through mostly immaterial new social practices than new product

development, its support systems need to concentrate on the active engagement of different stakeholders to the innovation process. Social innovations only materialize in transformed attitudes, behaviors, and perceptions of stakeholders that will ultimately become institutionalized as new social practices [13]. Hence, stakeholder interactions need to go beyond information exchange and entail an emerging collaborative knowledge co-creation process.

Accordingly, SI concept needs to be re-identified with a clear and specific definition emphasizing the engagement side of the process to demarcate itself from its common product focused definition. This redefinition needs to entail the "social" both in its ends and in its means. More than about innovating products to solve social problems, this definition should emphasize engaging and empowering stakeholders in the innovation process [14]. In this vein, Cajaiba-Santana [15] redefine SI as "the capacity to address social needs that traditional policy seems increasingly unable to tackle; the empowerment of groups and individuals; and the willingness to change social relations". According to this redefinition, engaging local stakeholders to co-create knowledge and solve their problems in a deliberative manner is both more democratic and also more efficient as it mobilizes local knowledge and human resources.

3 Distributed Networks of Knowledge Supporting Social Innovation

In this vein, we can say that classic PLM tools cannot provide support systems fulfilling the engagement requirements of this definition. Still focused to the end product than the empowerment aspects of the product life cycle, they are designed as enterprise collaboration tools collecting and managing data from detached stakeholders to be built into product or service features. As PLM tools clearly demarcate between external and internal boundaries of the organization, they do not have supporting mechanisms for interactive knowledge co-creation and sharing between stakeholders [16]. Since collaborative SI is knowledge intensive, immaterial, and involves rich flows of knowledge from diverse, complementary, and often conflicting sources [17] knowledge management is of critical importance during this process. The loss of tacit knowledge remains a major obstacle to the successful implementation of PLM to social innovations [18].

Classic PLM tools can successfully manage flow of technological knowledge among institutional stakeholders such as universities, research institutes and corporate R&D departments [19]. Technological knowledge is explicit and can be articulated in formal language and conveyed through codifiable data, scientific formulae, technical specifications etc. Tacit knowledge on the other hand, comes from implicit procedures, practices, symbols, values, and emotions [20]. Non-institutional stakeholders often possess more tacit than explicit knowledge about fulfilling local needs and existing production processes [21]. Tacit knowledge is often embedded in the context and is difficult to codify. While existing PLM support systems can manage codified knowledge flows, they are yet far from integrating tacit knowledge into the innovation process [22]. Tacit knowledge can be effectively generated through regular, extensive, and trustworthy personal interactions, can only manifest itself in-use, and can diffuse within social networks. It is distributed among diverse actors and is "captured" when

the knowledge bearer engages in a network organized as a community of practice. Hence, immaterial and knowledge intensive SI requires more than information management and should take into account the distributed characteristics of tacit knowledge co-creation.

Hence, SI support systems need to be organized as Distributed Networks of Knowledge (DNoK) to serve the engagement aspects of the innovation process [23]. DNoK consist of multiple co-located communities of practice where geographically dispersed participants are connected through a knowledge network. DNoK are distributed decentralized systems lacking an overseeing 'mind' where all the participants bring about their knowledge to bear to solve a problem [24]. In this respect, managing knowledge flows in DNoK can be somewhat paradoxical as the most important characteristics of such networks are emerging, self-organizing groups, developing over time, built upon trust and personal relationships [25, 26]. The existence of and interactions within DNoK are spontaneous and beyond the control mechanisms of formal organizations [27]. Knowledge is co-created in-use as a result of series of interactions between networks of stakeholders. Relevant actors contribute their varying skills and knowledge to the process in every stage of the innovation lifecycle. Created assets and returns are not the property of particular actors but open to network participants. Clearly, in such a system managing knowledge requires a different philosophy than existing PLM tools can offer.

4 Knowledge Management or Knowledge Governance?

Since distributed knowledge is constructed and shared within networks of communities of practice, trying to top-down engineer and 'manage' the emergent, spontaneous interactions therein should be counterproductive to the whole process. SI support systems require integrating elements of participative governance to establish a framework of accountability to both internal and external users, stakeholders and the wider community. Such knowledge governance tools should be concerned with the ways in which stakeholder communities cooperate with each other and the instruments and procedures that enable these interactions. Hence, knowledge governance needs to encourage stakeholder participation to SI support systems, by enabling different distributed voices to be heard and also re-distributing the returns of the co-created knowledge.

In this respect, social lifecycle assessment (S-LCA) approach which focuses mainly on interrelationships and behaviors of stakeholders has the potential to provide a new philosophy for a governance based SI support system. S-LCA aims to assess the socio-economic performance of products and services with respect to their key stakeholders and various social issues along their life cycle. It includes the assessment of actual and potential, positive as well as negative social and socio-economic impacts of products or services on stakeholders [28]. However, S-LCA support systems are still in their infancy and there is still no consensus on a complete and commonly accepted methodology for their implementation [29]. S-LCA concept is thus usually confused with classic LCA with merely a social analysis twist that considers macro societal aspects and consult institutional stakeholders, leaving out the tacit experiences and

distributed knowledge of grassroots actors. SI support systems need to engage, empower, and integrate diverse stakeholders to bring about social change in their lives. To be able to support SI, S-LCA systems need to consider the distributed nature of tacit knowledge and enable the democratic participation of all stakeholders as co-creators. Such systems should allow for hybrids of existing elements, crossing disciplinary boundaries, bridging qualitative and quantitative research, enabling creative bricolage of knowledge and building new relationships between previously separate individuals and groups.

Hence, the design, development and implementation of open systematic infrastructures are very important to integrate S-LCA to SI as a support system. Rather than focusing on individual creativity tools to manage the impact of product features, such support systems need to build and govern a social ecology to encourage distributed tacit knowledge co-creation. They require a meta-system for social creativity where different stakeholder communities are networked together to exchange practice and learning.

5 Designing a Social Ecology for Governing Collaborative Knowledge Construction and Social Creativity

Designing a social ecology for governing collaborative knowledge construction and social creativity can be complicated as SI involves several intermingled stages that do not involve a linear succession but proceed like iterative multiple spirals [30]. Organization of collaborative activities during these intermingled stages is an important issue since this may involve solving conflicting requirements from different actors [31]. While balancing potential conflicting issues requires specific support systems [32], research in this area is usually scarce [33].

SI support systems can both be considered as online platforms for virtual collaboration and also as networks of physical social innovation hubs committed to capacity-building. Such hubs can act as communities of practice where people can "see and touch" social innovations and where new solutions can be conceived, prototyped and scaled [30]. Hub networks can act as settings for distributed knowledge creation through exchange of new practices and tacit knowledge. Within these networks, the practice of social innovation can be collaboratively developed and made visible to all stakeholders. A multi-layered support system is required to organize the concerns and collaborative activities of diverse communities of practice within such a complex social ecology.

Architecture design is an essential step in devising such a support system. Existing literature in this field only focuses on technical system architecture design without considering human and social aspects. Only a few studies such as Sherman et al. [34] concentrate on the collaboration, communication, knowledge sharing, and review process during software architecture design. Designing architecture without consulting the concerned stakeholders to resolve conflicting demands may create serious usability problems turning the created system into impractical gadgets. In conventional architecture design, such consultation usually occurs at the prototyping and piloting stages. As we are considering a non-linear and iterative SI process, we review its support

system together with stakeholders during the idea generation stage rather than testing and reviewing the prototype. Designing such architecture should be an ongoing process and to meet SI principles it should be engaging and empowering stakeholders to the design process. In order to explore the guidelines for developing an ecological SI support system design, we conducted a case-study research using qualitative data collected from stakeholders of an NGO involved in primary and secondary education in Turkey.

6 Empirical Study

6.1 Method and Settings

The main objective of our study is to explore stakeholders' existing strategies for new education program development, their perceptions, expectations and interactions during existing collaboration activities, their existing online collaboration practices, and their visions about the ecology of a collaborative knowledge support system. Hence, we aim to concentrate on the collaboration, communication, and knowledge sharing aspects of a possible architecture design. Before designing the prototype, we asked participants to review their existing collaboration tools and their requirements for an efficient knowledge support system to maintain a social ecology. Participants were direct internal and external stakeholders of Educational Volunteers Foundation of Turkey (TEGV) including the general manager, director of education programs, managers of individual education programs, a member of the board of directors, CSR manager of a sponsor company, manager of an education unit, volunteers and students. TEGV is one of the largest and most reputable NGOs in Turkey operating in the field of education. It aims to support formal state education by creating and implementing education programs and extracurricular activities for children aged 6–14. In doing so they are working closely with a wide variety of stakeholders including the private sector, volunteers, educators, students and their parents as well as the local communities they operate in.

We conducted in-depth interviews with a small number of respondents from each group to explore their perspectives on existing new concept development practices, participation activities, and expectations from a collaborative platform in order to generate ideas for concept building for further research. For designing the research, we first identified relevant stakeholders, then what information is needed and from whom. We followed a snowball sampling strategy starting from TEGV managers to identify and reach interviewees from other stakeholder groups. Before each interview, we designed interview protocols to guide the rules during the administration and implementation of the interviews. We followed an iterative process during the structuring of the protocols, starting from literature research then grounding the protocols for further interviews to the previous ones. For each group, we fine-tuned the protocols and customized them according to the particular characteristics of the stakeholder group the respondent belongs to without deviating from the concept guidelines. At the analysis stage, we coded the gathered data to develop and refine interpretations of the interviews. We organized the coded data according to the regularities and patterns into thematic categories.

6.2 Findings and Discussion

6.2.1 Existing Product Development Strategies

Idea generation and conceptualization of projects is ad-hoc and sporadic, usually following popular commercial trends. The process is strictly linear, where usually idea comes from the sponsor, concept is developed by TEGV, and external experts develop the product. Education programs are considered as technical issues and are designed on a piecemeal basis by independent experts. These experts and TEGV R&D finalize the product through constant feedbacks. Corporate sponsors usually prefer ideas that would reinforce their corporate images, whereas TEGV agrees to develop products in return for the sponsorship. Mostly, institutional stakeholders formulate the ideas and promote the concepts. These stakeholders usually have dedicated departments for developing projects. Sometimes non-institutional stakeholders within the organization such as volunteers also participate at the idea generation stage. However, internal and external boundaries are clearly demarcated, hierarchical and professional standards for project development are strictly reinforced. The finalized product goes to the field for a pilot application. During this stage, voluntary workers are trained according to the program procedures. The responsible unit takes feedback from the volunteers and set the objectives for the project. Units do not participate in the main concept development but are consulted for polishing the features of the program.

6.2.2 Information Flow and Interaction Between Stakeholders

Information flow is not interactive but unidirectional; top-down and bottom-up. Interactive product development tools such as PLM are not used to collect and manage information from stakeholders. Information is collected usually through online structured surveys. Sometimes, unstructured data is collected through open ended surveys and group discussions. The programs designed by the experts usually do not fit to the realities of the field and hence despite this hierarchically structured process more flexible dialogical information sharing between the units and the management is allowed.

Interaction between TEGV and its stakeholders occurs in several ways. First, TEGV develops the program and leaves the PR and the promotion of the project to the marketing departments of the sponsors. This makes the project a part of the brand communication strategy of the sponsors. During this process there is a strict division of labor between the stakeholders. Second, the corporate contributes to the project with its volunteer professionals for capacity building for TEGV. These volunteers act as experts and mentors during the application of the program. This aspect encourages participation and active interaction especially during the unofficial idea generation and conceptualization stages. It increases the feelings of belonging and satisfaction of the employees and corporate management observes this as a long run reputation building exercise. Third way includes sponsorship relations where firms sponsor projects for reputation management without actively participating in them. These activities are funded by social responsibility budgets of firms as simple marketing activities for reinforcing their brand reputation. The last way includes a fully commercial relationship where firms sell solutions for the project and provide technical support and

consultancy. TEGV criticizes such relations for being commercialized, mechanical, and reified. They are critical about existing product focused project development strategies and aim for more sustainable, engaging and voluntary experiences during project development.

6.2.3 Towards a DNoK for Flexible Innovation Process

TEGV is ready to have more flexible and open relations for developing innovative programs in line with their new strategic frame and mission. According to the new frame, TEGV drops its broadly defined mission in favor of a narrowly defined domain comprehending three basic PISA education areas (reading, math and sciences) plus arts. They want these principles to act as an open frame where internal and external stakeholders can form project based autonomous teams at the grassroots level. They aim to empower project teams at field level by creating communities of practice with some decentralization experiments. They expect that ensuing trust and interaction based relations between stakeholders will lower partner circulation and hence transaction costs. This will also increase efficiency as decentralized project teams will have incentives to own projects and produce more customized products. Finally, a more flexible decentralized approach will stimulate a more fluent information flow among stakeholders providing a more dynamic feedback mechanism. The most important challenge is seen as coordination and knowledge management since this process tends to produce massive amounts of unstructured information. Neither TEGV nor its corporate sponsors use any integrative support systems for this process.

6.2.4 Knowledge Support System for DNoK

TEGV managers think that open innovation cannot be a solution to the challenges of DNoK as they find the information produced and number of people involved unmanageable and chaotic. Their vision of a knowledge support system is not limited to an open online platform. They think that knowledge production and sharing can be more deliberative at the local unit level. Designing knowledge support system as a "democratic assembly" where the features of the program is decided and voted by the public is not compatible with their mission. They fear that such an open platform can lead to politicization of issues. They want to remain at the technical level and at an equal distance to all political movements. Their policy strictly requires that education programs should be designed by academic experts according to scientific principles. However, they believe that programs can be designed according to open innovation and deliberative principles at local hubs where all the stakeholders, including children can have their say. These hubs can act as communities of practice during knowledge co-creation. Online collaboration and knowledge support systems can be effective in helping these communities share their practices and knowledge, and form teams with other hubs around specific projects. Besides this horizontal knowledge flow, support systems can help a dialogical process between top management and local communities of practice.

6.2.5 Knowledge Governance Architecture of the Platform

The governance boundaries of such co-creation efforts should be determined according to the new mission aiming to develop programs that address PISA subjects. As long as they stay within these boundaries, communities of practice will be free to negotiate and develop projects, form teams with other hubs around these projects and submit their proposals to a scientific committee. Hence, the co-creation platform requires a multi-layered and complex user roles and rights system which would be open to all stakeholders in a controlled manner. The platform will be open to the extent that proposed projects should be practical and in line with the strategic mission. Each participant will be able to see others' contributions, give feedback, and propose projects to put ideas into practice. Requirements for coordination should be different at the unit-hub, inter-hub team, and top management levels. The governance of the platform needs to be transparent, accountable, and dialogical (i.e. not just passive access to information and reports but also having a possibility for an interactive-critical dialogue). Present information sharing tools at hand are the organization website, Chatter, and common social media such as Facebook, Whatsapp and Twitter. The priority for a knowledge support system requires integrating these piecemeal efforts to a larger co-creation platform.

7 Conclusion

Policy makers and funding agencies increasingly emphasize the social nature of innovation as a result of public concerns for a more responsible research and innovation process [6]. However, focusing just on the product side of SI might easily turn the concept into a buzzword reducing it merely to the production and commercialization of social goods [10]. Engagement of all stakeholders in the knowledge co-creation process makes innovation really social. Supporting knowledge co-creation during this process requires tools for participative knowledge governance than top-down knowledge management. Designing the architecture for integrating such tools requires building and governing a social ecology system instead of focusing on creativity tools for managing the features of individual products [30].

Our findings suggest that stakeholders around TEGV are aware of the engagement aspects of innovation and ready for integrative support systems. A social innovation based knowledge support system requires an architecture bridging the coordination of online collaboration and knowledge management tools to the offline communities of practice. The most important challenge to this architecture is integrating piecemeal tools and practices into a collaborative knowledge co-creation platform designed as a social ecological system. Knowledge governance will be the backbone of such a system. Classical knowledge management systems are expert systems for formalizing technological knowledge by categorizing according to predefined ontologies and as such cannot cope with the complexities of the multi-layered social ecology architecture of SI. Therefore, creating a semantic web that would help to classify tools, topics, teams, and team members is essential to support the governance of distributed tacit knowledge emerging in this architecture. We suggest a supervised, learning ontology

based semantic web which involves constant engagement of communities of practice to the classification process as its core element. Such a semantic web will bridge pragmatics to semantics for an emergent knowledge ontology across the whole lifecycle of SI.

This research aims to explore and produce heuristic guidelines concerning the requirements for a social innovation based knowledge support system. The generalization of its findings is limited and needs to be triangulated with other case studies and confirmatory analyses done with other methods. Further research results can provide material for user requirements for designing social ecological support system architecture. The prototype of this system can be tested in a pilot incubation living-lab environment for further fine tuning to provide good practice example for NGO based social innovation.

References

1. FORA: The New Nature of Innovation. OECD. FORA, Copenhagen. http://www.newnatureofinnovation.org/full_report.pdf (2009)
2. O'Shea, T., Golden, J.S., Olander, L.: Sustainability and earth resources: life cycle assessment modelling. Bus. Strategy Environ. **22**, 429–441 (2013)
3. Lee, J.Y., Choib, S.S., Kima, G.Y., Noh, S.D.: Ubiquitous product life cycle management. Int. J. Comput. Integr. Manuf. **24**(7), 627–649 (2011)
4. Dekkers, R., Chang, C.M., Kreutzfeldt, J.: The interface between 'product design and engineering' and manufacturing: a review of the literature and empirical evidence. Int. J. Prod. Econ. **44**(1), 316–333 (2013)
5. Ameri, F., Dutta, D.: Product lifecycle management: closing the knowledge loops. Comput. Aided Des. Appl. **2**(5), 577–590 (2005)
6. European Commission: Options for Strengthening Responsible Research and Innovation—Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation. EC, Luxembourg. https://ec.europa.eu/research/swafs/pdf/pub_public_engagement/options-for-strengthening_en.pdf (2013)
7. Terzi, S., Bouras, A., Dutta, D., Garetti, M., Kiritsis, D.: Product lifecycle management—from its history to its new role. Int. J. Prod. Lifecycle Manag. **4**(4), 360–389 (2010)
8. Mulgan, G.: The process of social innovation. Innovations (Spring), 145–162 (2006)
9. Murray, R., Caulier-Grice, J., Mulgan, G.: The Open Book of Social Innovation. The Young Foundation and Nesta, London (2010)
10. Grisolia, F., Ferragina, E.: Social innovation on the Rise: yet another buzzword in a time of austerity? Salute e Società **1**, 169–179 (2015)
11. Shamir, R.: Socially responsible private regulation: world culture or world-capitalism? Law Soc. Rev. **45**(2), 313–336 (2011)
12. Millar, C.C.J.M., Choi, C.J., Chen, S.: Global strategic partnerships between MNEs and NGOs: drivers of change and ethical issues. Bus. Soc. Rev. **109**, 395–414 (2004)
13. Cajaiba-Santana, G.: Social innovation: moving the field forward. A conceptual framework. Technol. Forecast. Soc. Change **82**, 42–45 (2014)
14. Huddart, S.: Renewing the future: social innovation systems, sector shift, and innoweave. Technol. Innov. Manag. Rev. 5–9 (2012)
15. Cajaiba-Santana, G.: Social innovation: moving the field forward. A conceptual framework. Technol. Forecast. Soc. Change **82**, 42–45 (2014)

16. Borsato, M., Estorilio, C.C.A., Cziulik, C., Ugaya, C.M.L.: An ontology building approach for knowledge sharing in product lifecycle management. Int. J. Bus. Syst. Res. **4**(3), 278–292 (2010)
17. Lee, S., Maa, Y., Thimm, G., Verstraeten, J.: Product life-cycle management in aviation maintenance, repair and overhaul. Comput. Ind. **59**(2/3), 296–303 (2008)
18. Ibrahim, R., Paulson Jr., B.C.: Discontinuity in organizations: identifying business environments affecting efficiency of knowledge flows in PLM. Int. J. Prod. Lifecycle Manag. **3**(1), 21–36 (2008)
19. Un, C.A., Cuervo-Cazurra, A., Asakawa, K.: R&D collaborations and product innovation. J. Prod. Innov. Manag. **27**, 673–689 (2010)
20. Inkpen, A.C., Wang, P.: The China-Singapore Suzhou Industrial Park: a knowledge transfer network. J. Manag. Stud. **43**, 779–811 (2006)
21. Faems, D., Van Looy, B., Debackere, K.: Interorganizational collaboration and innovation: toward a portfolio approach. J. Prod. Innov. Manag. **22**, 238–250 (2005)
22. Ming, X.G., Yang, J.Q., Lu, W.F., Ma, D.Z.: Technology solutions for collaborative product lifecycle management—status review and future trend. Concurr. Eng. **13**(4), 311–319 (2005)
23. Hustad, E., Bechina, A.A.: Facilitating distributed networks of knowledge: a boundary management perspective. In: Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning (2011)
24. Tsoukas, H.: The firm as a distributed knowledge system: a constructionist approach. Strateg. Manag. J. **17**, 11–25 (1996)
25. Brown, J.S., Duguid, P.: Organizational learning and communities-of-practice: toward a unified view of working, learning and innovation. Org. Sci. **2**(1), 40–57 (1991)
26. Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation. Cambridge University Press, Cambridge (1991)
27. Wenger, E., Snyder, W.M.: Communities of practice: the organizational frontier. Harv. Bus. Rev. **78**(1), 139–145 (2000)
28. UNEP/SETAC Guidelines for social life cycle assessment of products, UNEP/SETAC Life Cycle Initiative, UNEP. http://www.oeko.de/oekodoc/908/2009-023-en.pdf (2009)
29. Benoit, C., Norris, G.A., Valdivia, S., et al.: The guidelines for social life cycle assessment of products: just in time! Int. J. Life Cycle Assess. **15**, 156–163 (2010)
30. European Commission: Financing Social Impact: Funding social innovation in Europe—mapping the way forward. EC, Luxembourg. http://ec.europa.eu/enterprise/policies/innovation/files/funding-social-innovation_en.pdf (2012)
31. Driessen, P.H., Hillebrand, B.: Integrating multiple stakeholder issues in new product development: an exploration. J. Prod. Innov. Manag. **30**(2), 364–379 (2013)
32. Winn, M.I., Keller, L.R.: A modeling methodology for multi-objective multistakeholder decisions: implications for research. J. Manag. Inq. **10**(2), 166–181 (2001)
33. Kaler, J.: Evaluating stakeholder theory. J. Bus. Ethics **69**(3), 249–268 (2006)
34. Sherman, S., Hadar, I., Levy, M., Unkelos-Shpigel, N.: Enhancing software architecture via a knowledge management and collaboration tool. In: 9th International Conference on Knowledge, Information and Creativity Support Systems KICSS'14, 6–8 Nov 2014, Limassol, Cyprus

A Study of Cost, Revenue, and Profit Efficiency of Commercial Banks in Cambodia Using DEA Approach

Sokim Phang(✉) and Morrakot Raweewan

School of Management Technology, Sirindhorn International Institute of Technology, Thammasat University, Bangkok, Thailand
sokimphang@gmail.com, morrakot@siit.tu.ac.th

Abstract. This paper examines the cost, revenue, and profit efficiency of the Cambodian banking system over the period 2010–2013, using Data Envelopment Analysis. Results show that, on average, the cost efficiency decreased from 86% in 2010 to 71% in 2013. The different development can be seen in the case of revenue efficiency, when average revenue efficiency increased from 77% in 2010 to 90% in 2012, and dropped to 65% in 2013. The profit efficiency was 78% in 2010 and 2011, and then increased to 81% in 2012; however, it reduced to 73% in 2013. Big banks had, on average, higher cost, revenue, and profit efficiency than smaller counterparts, while foreign banks, on average, were more cost, revenue, and profit efficient than Cambodian banks. The results suggest that commercial banks in Cambodia still have potential for cost savings as well as revenue and profit increases.

Keywords: Commercial bank · Cost efficiency · Revenue efficiency · Profit efficiency · Data envelopment analysis models

1 Introduction

Data Envelopment Analysis (DEA) is a linear programing based nonparametric method for evaluating efficiency of institutions which uses multiple inputs to generate multiple outputs. The basic DEA model was firstly introduced by Charnes et al. [1] in 1978 based on assumption of constant return to scale, and then modified by Banker et al. [2] in 1984 based on assumption of variable return to scale. DEA has been widely adopted nowadays and extended by many scholars in modern banking sector research society to study efficiency of banks (e.g. Sherman and Gold [3], Rangan et al. [4], Paradi and Zhu [5], Bader et al. [6], Kočišová [7], Manlagñit [8], Jayaraman and Srinivasan [9], Ariff and Can [10], Camanho and Dyson [11], Giokas [12], Noulas et al. [13], Manandhar and Tang [14], Schaffnit et al. [15], Jaffry et al. [16], Drake et al. [17], and Sahin et al.[18]).

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_14

The cost, revenue, and profit efficiency models are among the most important extended DEA models used to measure efficiency of banking systems. In cost efficiency model, a firm minimizes costs reflected by how close its costs to what a best-practice firm's would be for generating equal level of outputs at certain prices of inputs and outputs [10]. In revenue efficiency model, the objective of a firm is to maximize total revenue by producing combination of outputs that bounce maximum revenue at applicable output prices [6]. The profit efficiency model measures how close a firm comes to generating the maximum profit given an amount of inputs and outputs and a level of their prices. Profit efficiency considers the effects of selection of vector of production on both costs and revenues. The object of a firm is to choose a good combination of inputs and outputs that yield the maximum profitability in the market with known price of inputs and outputs [10].

This study adopts the DEA approach with the main objective to evaluate cost, revenue, and profit efficiency of the commercial banks in Cambodia during the period 2010–2013. This study tries to answer three research questions: (i) Are commercial banks in Cambodia operating efficiently in term of cost minimization, revenue and profit maximization? (ii) How has cost, revenue, and profit efficiency of commercial banks in Cambodia changed over time? (iii) What can be done for inefficient banks to improve the efficiency?

The structure of this paper is organized as follows. Section 2 provides a review of relevant literature. Section 3 contains used methodology in detail, followed by Sect. 4 discusses the empirical analysis and results. Finally, the paper ends with conclusions remarks in Sect. 5.

2 Literature Review

Sherman and Gold [3] were among the earliest researchers to apply the DEA application to measure efficiency in the banking sector. The scholars employed the DEA model to evaluate the level of bank branch efficiency and adjusted inputs and outputs to suggest ways to eliminate inefficiency of the branches. They found that DEA provides a meaningful insight for improving bank branch efficiency that cannot be obtained by other techniques, in the area of productivity improvement. A later study, by Rangan et al. [4], introduced the first two-stage contextual DEA for banking applications. A DEA application was employed to measure the technical efficiency of US banks. The study results suggested that the US banks can produce the same levels of their outputs by consuming only 70% of the actual current inputs. These findings indicate that these banks were wasting resources.

Motivated by its powerful optimization ability and meaningful insightful results, DEA has been popularly used by modern research scholars to evaluate the performance in the banking sector around the globe. There exists a considerable growth of research literature in the field. In this sense, according to a survey of 195 published papers, Paradi and Zhu [5] have found that studies using DEA applications at the bank level focused mainly on four areas, including (i)

evaluating efficiency overtime (due to market structure, economic environment change, banking liberalization and deregulation), (ii) international comparison of banks, (iii) bank performance benchmark and improvement, and (iv) effects of ownership and bank types.

In addition, many studies have focused on evaluating and comparing branch network cost and profit efficiency of a parent bank. For example, the works of Manlagñit [8], Camanho and Dyson [11], Giokas [12], and Noulas et al. [13] have evaluated branch cost efficiency, and the works of Jayaraman and Srinivasan [9], and Manandhar and Tang [14] have analyzed branch profitability. These studies have evaluated the ability of a branch to produce current outputs (loans, deposits, number of service transactions, revenues and profits) at minimal costs, using the same level of current inputs (labors, assets and capitals) to generate maximal profits. The quantitative data of inputs and outputs, as well as information of price of each input and output at each branches, have been used for analysis.

Moreover, in the case of using DEA applications for cost and benefit optimization of commercial banks, Bader et al. [6] have used the DEA approach to measure cost, revenue, and profit efficiency of Islamic banks and conventional banks from 1990 to 2005, with an intermediation approach. They found no major differences of overall efficiency between bank systems. Those banks did not fully utilize the current resources at the most efficient level, and there was indication of inefficiency in usage of the resources. Therefore, to become cost, revenue, and profit efficient, both bank systems need to eliminate the inefficient resources. In another study, Kočišová [7] has applied the DEA approach to evaluate cost, revenue, and profit efficiency of Slovak and Czech commercial banks during 2009–2013, using an intermediation approach. In their study, they compared the efficiency of both countries and they found that the Czech banks were more cost, revenue, and profit efficient than Slovak banks for the whole period analyzed.

3 Methodology

Data Envelopment Analysis (DEA) is a theoretically sound framework for performance analysis. Technically, DEA represents the set of nonparametric, linear programing techniques used to construct empirical production frontiers and evaluate relative efficiency [15,19]. Schaffnit et al. [15] explains that by using relevant inputs and desirable outputs of process with similar units (e.g. stores, mutual funds, hospitals, bank branches, etc.), DEA identifies: (i) the efficient frontier, or envelopment surface, consisting of best practice units; (ii) efficiency measures for each decision making unit (DMU) that reflects its distance to the frontier; (iii) an efficient reference set, or peer group, for each inefficient DMU; and (iv) efficient targets for each inefficient DMU (projections onto frontier).

Two alternative approaches are available in DEA to determine the technical efficiency frontier, input-oriented and output-oriented models. The following DEA model is an input-oriented model (known as BCC model) under the assumption of variable return to scale (VRS) where inputs are minimized and outputs are kept at their current levels.

$$\theta^* = min\ \theta \tag{1}$$

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io} \quad i = 1, 2, \ldots, m;$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \leq y_{ro} \quad r = 1, 2, \ldots, s;$$

$$\sum_{j=1}^{n} \lambda_j = 1 \quad j = 1, 2, \ldots, n; \quad \lambda_j \geq 0$$

where: θ is input-oriented technical efficiency of DMU_o in input-oriented DEA model; x_{io} is consumed input amount of ith for DMU_o; y_{ro} is produced amount of rth output for DMU_o; x_{ij} is consumed amounts of ith input of DMU_j; y_{rj} is produced amount of rth output of DMU_j; and λ_j are nonnegative scalars [19].

Since $\theta = 1$ is a feasible solution and the optimal value to (1), $\theta^* \leq 1$. If $\theta^* = 1$, then the current input levels cannot be reduced (proportionally), indicating that DMU_o is on the frontier. Otherwise, if $\theta^* < 1$, then DMU_o is dominated by the frontier. θ^* represents the (input-oriented) efficiency score of DMU_o [19].

When price data available, then it is possible to measure the technical, allocative efficiency as well as cost, revenue, and profit efficiency [7]. In this study, we adopt popular extensions of basic DEA models to measure the cost, revenue, and profit efficiency of commercial banks in Cambodia. The performance of each bank is evaluated by comparing its performance with the best performing counterparts analogous to concept of efficiency. To calculate cost, revenue, and profit efficiency, we use the DEA excel solver provided by Zhu [19]. Using input and output prices, Zhu summarizes the associated cost, revenue, and profit DEA models under assumption of variable return to scale (VRS) as shown in Table 1.

Following Bader et al. [6], Kočišová [7], and Ariff and Can [10], among others, this study uses the intermediation approach for choices of inputs and outputs, and calculation of input price and output prices of banks. In the intermediary approach, banks are assumed to function as intermediaries in the financial markets. Banks obtain funds from drawing deposits, and turn deposits and other funds into productive investments, such as loans and other earning assets. This conversion requires labor and capital [16–18]. Accordingly, we considers total fixed assets (x_1), total employees (x_2), and total funds (x_3) as inputs, and total loans (y_1) and total other earning assets (y_2) as outputs. Each of these inputs and outputs generate cost and income, respectively. Therefore, the price of inputs and outputs can be computed as follows. On the input side, the price of labor (p_1) is the ratio of total labor expenses to total number of employees; price of fixed assets (p_2) is the ratio of total depreciation expenses to total value of fixed assets. The price of total fund (p_3) is the ratio of total interest expenses to total of funds. On the output side, the price of total loans (q_1) is the ratio of total interest incomes to total loans, and the price of other earning assets (q_2) is the ratio of total operating incomes to total other earning assets.

Table 1. Cost, revenue, and profit efficiency models

Cost	Revenue	Profit
$\min \sum_{i=1}^{m} p_{io}\tilde{x}_{io}$ (2)	$\max \sum_{r=1}^{s} q_{ro}\tilde{y}_{ro}$ (3)	$\max \sum_{r=1}^{s} q_{ro}\tilde{y}_{ro} - \sum_{i=1}^{m} p_{io}\tilde{x}_{io}$ (4)
subject to	subject to	subject to
$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \tilde{x}_{io}$	$\sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io}$	$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \tilde{x}_{io}$
$i = 1, 2, \ldots, m;$	$i = 1, 2, \ldots, m;$	$i = 1, 2, \ldots, m;$
$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}$	$\sum_{j=1}^{n} \lambda_j y_{rj} \geq \tilde{y}_{ro}$	$\sum_{j=1}^{n} \lambda_j y_{rj} \geq \tilde{y}_{ro}$
$r = 1, 2, \ldots, s;$	$r = 1, 2, \ldots, s;$	$r = 1, 2, \ldots, s;$
$\sum_{j=1}^{n} \lambda_j = 1$	$\sum_{j=1}^{n} \lambda_j = 1$	$\sum_{j=1}^{n} \lambda_j = 1$
$j = 1, 2, \ldots, n;$	$j = 1, 2, \ldots, n;$	$j = 1, 2, \ldots, n;$
$\tilde{x}_{io}, \lambda_j \geq 0$	$\tilde{y}_{io}, \lambda_j \geq 0$	$\tilde{x}_{io} \leq x_{io}, \tilde{y}_{ro} \geq y_{ro}, \lambda_j \geq 0$

where: m is number of inputs; s is number of outputs; i is mth input; r is sth output; p_{io} is unit price input ith DMU$_o$ (DMU$_o$ represents decision making unit under evaluation); q_{io} is unit price outputs rth for DMU$_o$; $\tilde{x}_{io}$ is ith input that minimizes the cost; $\tilde{y}_{ro}$ is rth output that maximizes revenue; x_{io} is ith input for DMU$_o$; y_{ro} is rth output for DMU$_o$; x_{ij} is consumed amounts of ith input of DMU$_j$; y_{rj} is produced amount of rth output of DMU$_j$; n is number of DMU; and λ_j are nonnegative scalars

The samples of this study consist of 24 commercial banks operating in Cambodia from 2010 to 2013. All data were obtained mainly from the annual supervision reports of National Bank of Cambodia which has comprehensive information of financial and ownership of all the financial institutions in Cambodia. Additional data and double check are done from official annual reports of the banks.

4 Empirical Analysis and Results

4.1 Cost, Revenue, and Profit Efficiency

This section describes practical applications of methods for measuring cost, revenue, and profit efficiency of commercial banks in Cambodia during 2010–2013 using DEA. Figure 1 illustrates the DEA results for measuring efficiency of the major commercial banks in Cambodia. According to Fig. 1, on average, the cost efficiency (CE) decreased from 86% in 2010 to 71% in 2013. When we look at the average revenue efficiency (RE), it increased from 77% in 2010 to 90% in 2012; however, it dropped to a minimum of 65% in 2013. The average profit efficiency (PE) was 78% in 2010 and 2011, and grew to 81% in 2012; however, it reduced

to 73% in 2013. The cost, revenue, and profit efficiency did not show an obvious trend throughout the observation period. It was observed that the ratios tend to decrease slightly over time, but the evidence was inconclusive. One of the main reasons explains drops in cost, revenue, and profit efficiency of Cambodian banking sector in 2013 was the adverse effects of political uncertainty and labor unrest since the second half of 2013. It caused private sector deposit growth to slow significantly, reducing from 25.2% by end-2012 to 14.2% year-on-year by end-2013 [20].

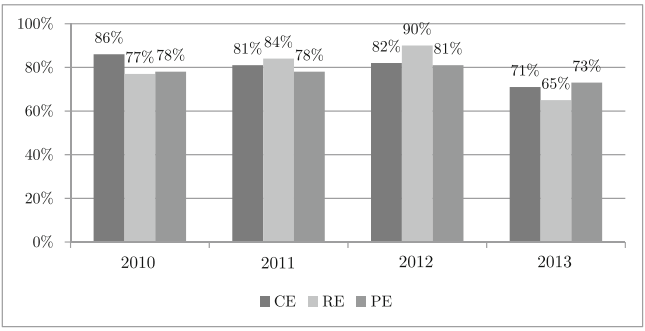


Fig. 1. Average CE, RE, and PE of commercial banks in Cambodia

For the whole observed period of 2010–2013, on average, the cost, revenue, and profit efficiency of commercial banks in Cambodia were 80%, 79%, and 78%, respectively. The results indicate that the Cambodian banking sector still has potential of cost savings, as well as revenue and profit increases of approximately 20%.

The results of cost, revenue, and profit efficiency can be illustrated with an individual bank. Consider Bank 5 in 2013 in Table 2. This bank reached a cost efficiency of 100%. This indicates that Bank 5 was cost efficient in the peer group, which means this bank used the minimum possible cost (fixed assets, total employees, and total funds) for producing its observed outputs (total loans and total other earning assets). When we look at revenue efficiency we observe that the revenue efficiency is not 100% (RE = 61.99%). This means that Bank 5 did not produce the maximum possible outputs by using the given inputs. So there exists a space for potential revenue increases in the case of Bank 5. The observed value of revenue was 3,490 billion KHR, and the optimal value was 4,394 billion KHR. To become revenue efficient Bank 5 should increase its revenue by 20.58%. This potential revenue increases in the case of Bank 5 can be obtained by reducing total loans 1,041 billion KHR and by increasing the total other earning assets to 3,353 billion KHR, while maintaining the given outputs prices. The optimal combination of outputs allows Bank 5 to reach maximum revenue and shifts onto efficiency frontier. The profit efficiency of Bank 5 reached 100% indicating that this bank generated the maximum profit with the observed combination of inputs and outputs with their prices.

Table 2. CE, RE, and PE of commercial banks in Cambodia

Year	2010			2011			2012			2013		
Commercial Bank	CE (%)	RE (%)	PE (%)	CE (%)	RE (%)	PE (%)	CE (%)	RE (%)	PE (%)	CE (%)	RE (%)	PE (%)
Bank 1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	75.20	100.00
Bank 2	60.83	73.97	60.12	62.77	73.29	55.00	64.24	79.64	65.50	100.00	100.00	100.00
Bank 3	94.53	72.78	94.43	81.65	49.55	40.43	95.61	99.46	98.84	63.92	32.39	31.12
Bank 4	100.00	100.00	100.00	81.65	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Bank 5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	61.99	100.00
Bank 6	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Bank 7	100.00	22.19	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Bank 8	100.00	100.00	100.00	100.00	100.00	100.00	24.29	94.33	41.01	7.33	19.08	13.55
Bank 9	34.65	72.57	39.12	47.28	68.02	47.14	60.55	96.50	74.25	14.53	14.14	11.07
Bank 10	69.65	49.99	34.81	62.22	45.54	27.11	67.54	38.33	23.16	61.12	59.70	58.94
Bank 11	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	46.74	100.00
Bank 12	100.00	56.76	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	58.60	100.00
Bank 13	52.84	55.20	35.92	36.35	69.84	40.20	46.23	75.80	58.08	25.08	46.60	50.90
Bank 14	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Bank 15	100.00	88.68	100.00	100.00	78.33	100.00	100.00	100.00	100.00	100.00	36.80	100.00
Bank 16	60.18	79.18	64.62	63.36	74.36	85.17	100.00	100.00	100.00	37.13	32.40	30.46
Bank 17	100.00	100.00	100.00	88.88	83.95	81.29	91.30	98.21	96.62	100.00	100.00	100.00
Bank 18	87.17	28.34	14.94	79.77	46.05	33.66	19.83	57.82	4.47	2.59	25.01	2.43
Bank 19	100.00	86.25	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	85.84	100.00
Bank 20	75.73	61.78	52.50	63.31	65.19	51.33	66.27	84.68	78.00	24.30	50.93	56.59
Bank 21	100.00	90.81	100.00	55.06	93.41	79.55	100.00	100.00	100.00	42.26	41.61	38.37
Bank 22	100.00	100.00	100.00	100.00	100.00	100.00	80.02	95.98	93.57	100.00	100.00	100.00
Bank 23	70.17	70.60	59.09	82.92	89.70	84.20	90.22	80.12	73.52	100.00	100.00	100.00
Bank 24	64.45	47.78	28.08	38.89	70.48	35.04	54.45	67.38	43.25	22.93	75.70	61.88

4.2 Attributes to Bank Efficiency

A variety of attributes are used to measure the cost, revenue, and profit efficiency of the banks in Table 2. Therefore, the banks are classified into groups according to bank attributes, derived from data, and the mean of each group is compared to examine how bank attributes affect banks efficiency. First the banks are classified into two groups based on size, either big or small banks. A bank is define as a big bank (BB) when it is ranked as one of the four largest commercial banks in Cambodia, while a bank is defined as a small bank (SB) when its total assets amount to less than one-tenth of the average amount of total assets of entire banking sector. Second, the banks are classified into two groups based on ownership structure, as Cambodian or foreign banks. A bank is defined as a Cambodian bank (CB) when the ratio of Cambodian share exceeds 50%; otherwise, it is a foreign bank (FB).

Figures 2, 3 and 4 show average cost, revenue, and profit efficiency, respectively, of small-bank group, big-bank group, foreign-bank group, and Cambodian-bank group. After that, the average cost, revenue, and profit efficiency of small-bank group are compared with average cost, revenue, and profit efficiency of big-bank group. The average cost, revenue, and profit efficiencies,

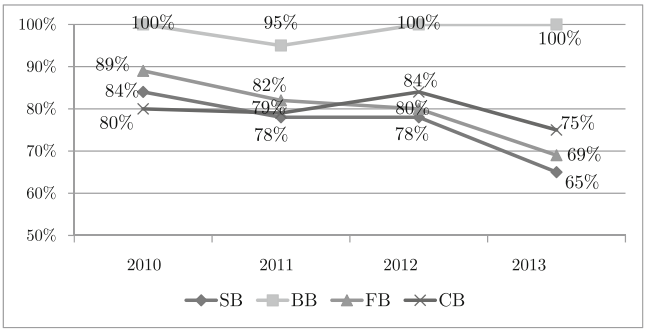


Fig. 2. Average cost efficiency of SB, BB, FB, and CB

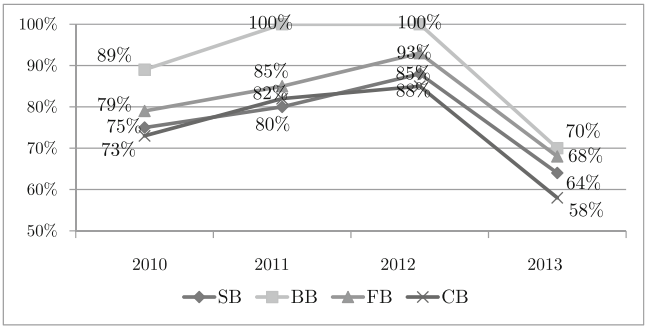


Fig. 3. Average revenue efficiency of SB, BB, FB, and CB

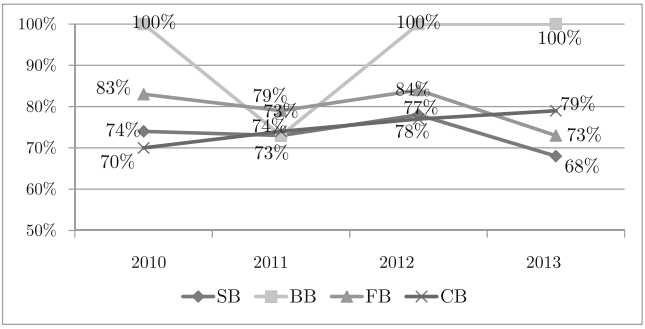


Fig. 4. Average profit efficiency of SB, BB, FB, and CB

except the profit efficiency in 2011, of big banks are higher than the small-bank group for the whole observation period. This result shows that big banks utilize their resources better than small banks. As popular belief hold that big banks, generally, can achieve substantial economies of scale and outperform small competitors. The result, interestingly, is consistent with previous studies.

When the average cost, revenue, and profit efficiency of foreign-bank group are compared with that of Cambodian-bank group as illustrated in Figs. 2, 3 and 4, the average cost, revenue, and profit efficiency of foreign banks tend to be higher than Cambodian banks. However, it is unclear which of these two groups is better than the other because the ranking of efficiency changes each year. The results suggest that foreign banks use their resources better than the Cambodian banks.

5 Conclusions

This study has the main purpose to evaluate the cost, revenue, and profit efficiency of commercial banks in Cambodia over the period of 2010 to 2013. The level of efficiency was measured using Data Envelopment Analysis which applied an intermediation approach. Results show that, on average, cost efficiency decreased from 86% in 2010 to 71% in 2013. The different development can be seen in the case of revenue efficiency, when average revenue efficiency increased from 77% in 2010 to 90% in 2012, and dropped to 65% in 2013. The profit efficiency was stable in 2010 and 2011, as equal to 78%, and then grew to 81% in 2012; however, it reduced to 73% in 2013. When we look at efficiency according to attributes we can see that big banks are more cost, revenue, and profit efficient than smaller counterparts. Moreover, foreign banks, on average, are more cost, revenue, and profit efficient than Cambodian banks. However, it is unclear which of these two groups is better than the other because the ranking of efficiency changes each year. Results suggest that the Cambodian banking system still has potential for cost savings as well revenue and profit increases.

The results suggest important policy implications. First is to increase competitive capacity of commercial banks in Cambodia by measuring cost, revenue,

and profit efficiency, using a DEA approach. Second is to enlarge the size of business operations of individual banks. However, this study focuses on investigating the cost, revenue, and profit efficiency of commercial banks but does not analyze their determinants and dynamics. The authors acknowledge this limitation of this paper and leave it for further research.

References

1. Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978)
2. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag. Sci. **30**(9), 1078–1092 (1984)
3. Sherman, H., Gold, F.: Bank branch operating efficiency: evaluation with data envelopment analysis. J. Bank. Finance **9**(2), 297–315 (1985)
4. Rangan, N., Grabowski, R., Aly, H.Y., Pasurka, C.: The technical efficiency of us banks. Econ. Lett. **28**(2), 169–175 (1988)
5. Paradi, J.C., Zhu, H.: A survey on bank branch efficiency and performance research with data envelopment analysis. Omega **41**(1), 61–79 (2013)
6. Bader, M.K.I., Mohamad, S., Ariff, M., Hassan, T.: Cost, revenue, and profit efficiency of islamic versus conventional banks: international evidence using data envelopment analysis. J. Islamic Econ. Stud. **15**(2), 24–76 (2008)
7. Kočišová, K.: Application of data envelopment analysis to measure cost, revenue and profit efficiency. Statistika **94**(3), 47–57 (2014)
8. Manlagñit, M.C.V.: Cost efficiency, determinants, and risk preferences in banking: a case of stochastic frontier analysis in the philippines. J. Asian Econ. **22**(1), 23–35 (2011)
9. Jayaraman, A., Srinivasan, M.: Analyzing profit efficiency of banks in india with undesirable output—nerlovian profit indicator approach. IIMB Manag. Rev. **26**(4), 222–233 (2014)
10. Ariff, M., Can, L.: Cost and profit efficiency of chinese banks: a non-parametric analysis. China Econ. Rev. **19**(2), 260–273 (2008)
11. Camanho, A., Dyson, R.: Cost efficiency measurement with price uncertainty: a dea application to bank branch assessments. Eur. J. Oper. Res. **161**(2), 432–446 (2005)
12. Giokas, D.I.: Cost efficiency impact of bank branch characteristics and location: an illustrative application to greek bank branches. Manag. Financ. **34**(3), 172–185 (2008)
13. Noulas, A.G., Glaveli, N., Kiriakopoulos, I.: Investigating cost efficiency in the branch network of a greek bank: an empirical study. Manag. Financ. **34**(3), 160–171 (2008)
14. Manandhar, R., Tang, J.C.: The evaluation of bank branch performance using data envelopment analysis: a framework. J. High Technol. Manag. Res. **13**(1), 1–17 (2002)
15. Schaffnit, C., Rosen, D., Paradi, J.C.: Best practice analysis of bank branches: an application of dea in a large canadian bank. Eur. J. Oper. Res. **98**(2), 269–289 (1997)
16. Jaffry, S., Ghulam, Y., Pascoe, S., Cox, J.: Regulatory changes and productivity of the banking sector in the indian sub-continent. J. Asian Econ. **18**(3), 415–438 (2007)

17. Drake, L., Hall, M.J., Simper, R.: Bank modelling methodologies: a comparative non-parametric analysis of efficiency in the Japanese banking sector. J. Int. Fin. Markets Inst. Money **19**(1), 1–15 (2009)
18. Sahin, E., Yılmaz, B., Akgün, A.: Measurement of efficiency in the turkish banking sector in the 2009–2011 period: a dea approach. Mediterr. J. Soc. Sci. **4**(10), 744–752 (2013)
19. Zhu, J.: Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets. Springer Science and Business Media, LLC, Boston, MA (2009)
20. World bank. http://www.worldbank.org/content/dam/Worldbank/document/EAP/cambodia/Cambodia_EAP_Update-Apr2014_ENG.pdf (2014)

Mapping Knowledge Management for Technology Incubation

Jarot S. Suroso(✉), Abba Suganda Girsang, and Ford Lumban Gaol

Bina Nusantara University, Jakarta, Indonesia
{jsembodo, agirsang, fgaol}@binus.edu

Abstract. Knowledge Management is an approach which is based on the understanding that the task of the organization, which is understood by both of the reuse of knowledge and how the success of re-use of knowledge that has been created. Knowledge management is a series of activities that are used to identify, create, describe, and distribute knowledge. Business and Technology Incubator is one of alternative models of business development through incubation system that can help entrepreneurs to create and grow their business. This review will discuss mapping knowledge management for business and technology incubator that provides guidance to new business, support facilities and transfer of technology and business, especially for small and medium enterprises which will be a case study in Tegal regency, Central Java, Indonesia. The result showed that mapping knowledge management for business and technology incubator can help entrepreneur winning the competition in their business.

Keywords: Knowledge management · Technology incubation
Business incubation · Small-medium enterprise

1 Introduction

Currently, it was realized that the organization is no longer superior relies solely on financial resources, buildings, land, technology, market position, and tangible assets other, but instead relies on knowledge assets (intangible). This is because knowledge is a resource that has the characteristics of strategic advantage. Knowledge management in recent years has become an interesting topic for discussion. Many interested parties to develop this concept both on the level of business and other non-profit organizations. Knowledge management concept originated and evolved in the business world. As a concept, knowledge management is applied in order to increase and improve the operation of the company in order to gain competitive advantage and increase profits. The basic objective is to harness knowledge for organizational excellence. Other objectives are: 1. Facilitating the transition period the old with the new. 2. Minimize the loss of "knowledge" because the release of employees. 3. Knowing the resources and critical knowledge areas owned and 4. Develop a method to prevent the loss of intellectual property companies. In relation to the incubation of technology, knowledge management is expected to encourage innovative entrepreneurs to grow into a strong and high-competitive businessman. This review will discuss the mapping knowledge management used in business and technology incubator that is an institution that

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_15

provides guidance to new businesses by providing support facilities and the transfer of technology and business, especially on small and medium enterprises, which will be a case study in Tegal regency, Central Java, Indonesia.

2 Knowledge Management

The term knowledge management first appeared in the business world. Knowledge Management is an attempt to increase the useful knowledge in the organization, including cultural familiarize communicate between personnel, provide opportunities for learning, and promote the sharing of knowledge. Where this effort will create and sustain an increase in the value of the core business competencies by leveraging existing information technology. It is excerpted from the opinions McInerney as follows: "Knowledge Management (KM) is an effort to increase of useful knowledge within the organization. Ways to do this include encouraging communication, offering opportunities to learn, and promoting the sharing of knowledge Appropriate artifacts" [1].

Nonaka and Takeuchi argue that the fundamental reason why Japanese company success, because they maintain the creation of organizational knowledge. The creation of knowledge is achieved through the introduction of synergic relationship between tacit and explicit knowledge [2]. Ikujiro Nonaka and Hirotaka Takeuchi (1995), distinguishing between tacit and explicit knowledge, and knowledge conversion model divide into 4 ways as follows:

a. Tacit knowledge to Explicit knowledge; called externalization process
b. Tacit knowledge to Tacit knowledge; called Socialization process.
c. Explicit knowledge to explicit knowledge; called Combination process.
d. Explicit knowledge to tacit knowledge; called Interlization process.

The essence of Knowledge Management is intended to represent a planned and systematic approach to ensure full use of basic knowledge of the organization, plus the expertise, competence, ideas, innovation, and potential individual ideas to create a more efficient organization and effective. Knowledge Management is an information management paradigm that comes from the idea that pure knowledge is actually embedded in the minds and mind of every human being. Thus it is necessary to build a mechanism for disseminating information and experience from the existing human resources in order to increase the knowledge of each actor activities in an organization [3].

With the development of technology, the issue of how to get a high level of competitiveness is also experiencing a transition from the use of (natural) resources to technology. The world has witnessed how countries are given the abundant natural resources would be a backward country. Indonesia one of them. Conversely a country that has often been at the forefront limitations because they were able to transform limitations into an opportunity. In the technological development of the most essential thing is the ability and knowledge. According to Nonaka and Takeuchi (1995) Japanese company's success is determined by the skills and expertise in the creation of knowledge organization. To perform this mapping knowledge management used

10-step method of knowledge management roadmap drawn up by Amrit Tiwana are broadly divided into four phases as follows [4]: i.e. 1. Evaluation of infrastructure, 2. Analysis, design and development of Knowledge Management System (KMS), 3. System deployment and 4. Evaluation (Fig. 1).

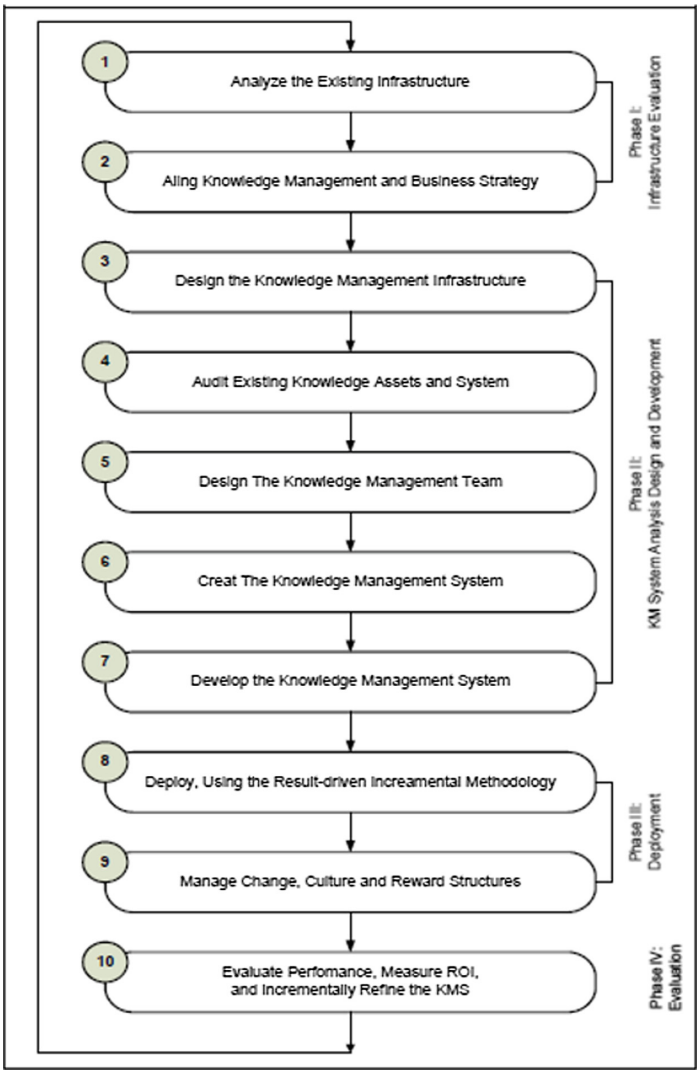


Fig. 1. 10-step knowledge management roadmap (Source: [4])

3 Technology Incubation

Business and technology incubator—hereinafter referred to as the incubator—has been developed in several developed countries. In the United States, for example, the incubator has grown since the early 1980s. In Indonesia, the incubator was developed in 1992, with the government's initiatives to develop cooperation with the university incubator.

Business and Technology Incubator (BTI) is one of the alternative models of business development through incubation system that can help the problems faced by entrepreneurs in order to create and grow a business world that is robust, efficient, especially for Small and Medium Enterprises. Or in other words, a business and technology incubator is an institution that provides guidance to new businesses by providing support facilities and the transfer of technology and business. Incubator role in the development of SMEs have also been getting a response and support from experts. Such support has spurred growth in many countries incubator. It is estimated that there are already more than 1,500 incubators in the world. Most of them are in the United States and European countries as well as more than 250 incubators are in developed countries.

Incubator role is very important for entrepreneurs/SMEs beginners (tenant innovative technology-based), as Incubator is designed to assist new and emerging businesses become independent entrepreneurs through a series of integrated assistance include the provision of office facilities, production testing, market testing, management consulting, technology, marketing and finance, training, and the creation of business networks both locally and internationally.

In general, the incubator is managed by a staff with a highly efficient management by providing services "7S", namely: space, shared, services, support, skill development, seed capital, and synergy [5].

a. Space means incubator provides a place to develop business in the early stages.
b. Shared means that the incubator provides office facilities that can be used together, for example, receptionist, conference rooms, telephone systems, fax, computers and networks, as well as security systems work.
c. Services includes the management consulting and market issues, financial and legal aspects, trade and information technology.
d. Support in this case the incubator helps access to research, professional networking, technology, international, and investments.
e. Skill development can be done through training, and other human resource capacity building.
f. Seed capital can be done through an internal revolving fund or by helping small businesses access to sources of financing or existing financial institutions.

g. Synergy is a tenant and increased cooperation network with the universities, research institutions, private businesses, professionals as well as with the international community.

Different Types of existing incubator can be known by sponsors who support it. Currently, there are approximately 4 types of incubators [6] that can be used as a reference in the development of incubators in several countries, namely:

a. Research, University, Technology-based Business Incubator

Development on the basis of research in universities and research institutions. The focus of the program is to provide services for personnel who are trained to become an entrepreneur is to transfer technology to enter the market (technology push) and meet a wide range of opportunities available (market pull).

b. Public-private Partnership, the Industrial Development Incubator

These incubators generally live in an urban environment or industrial estate, where large companies could be involved in the development of small businesses as vendors for components and services.

c. Foreign Sponsors, International Trade and Technology

The focus is usually incubator program for the development of international collaboration, technological and financial, to facilitate the entry of foreign small and medium enterprises in the local market (domestic).

d. Other types of incubators

Incubator is a focus on a specific group development program. Incubator located in universities and research and development institutions, both public and private in general is kind of research incubator, university, technology-based business incubator. However, there are some differences between the incubators developed at the university with an incubator developed in research and development institutions.

Incubator contained in colleges more starting from a business idea and more require seed capital to realize its business. Meanwhile incubators (technology-based business incubator) developed by research and development institutions in the birth of new businesses based on technology, start-up incubation of existing prototypes are then performed assistance to be able to enter the market (Technology Push). Secondly, from business ideas, prototypes and technological services incubated as the market needs (market pull). In a simple conception of incubators that are common to research and development institutions shown in Fig. 2, which describes the start of the establishment of the incubator institutional infrastructure must be owned and incubation process to be passed by the tenant.

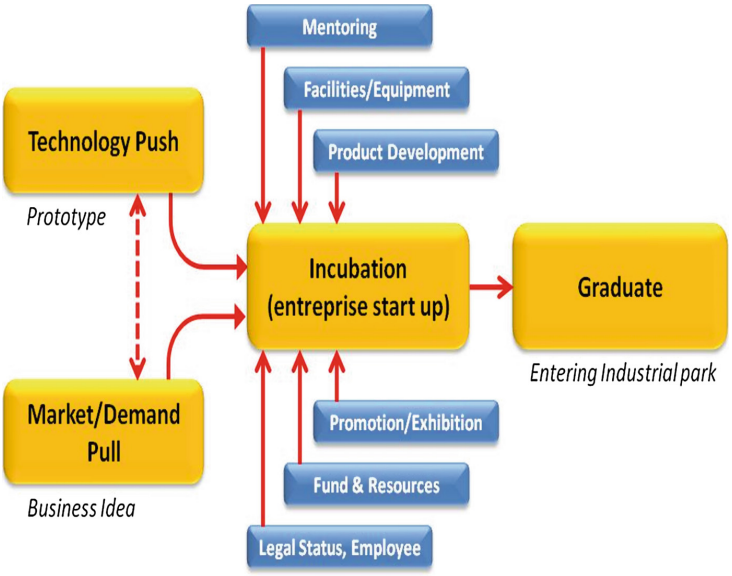


Fig. 2. Concept of incubator in research and development institutions

4 Mapping the Knowledge Management for Technology Incubation

In carrying out this survey, we used descriptive research methodology. Descriptive research is one of the types of studies that aim to present a complete picture of the social setting or the relationship between the phenomena tested. In this study, researchers have had a clear definition of the subject of the research and who will use the questions in digging the required information. The purpose of descriptive research is to produce an accurate picture of a group, describe the mechanism of a process or relationship, give a complete picture either in the form of verbal or numerical, will present the basic information of a relationship, create a set of categories and classify the study subjects, describes a set of steps or processes, as well as to store contradictory information regarding the study subjects.

Mapping target is a cluster actors metal processing industry in Tegal, Central Java. The instrument used was the questionnaire contained in the annex. From the mapping results are then analyzed and resulted in a recommendation in terms of knowledge management, to improve the ability of industry cluster Tegal. Explanation of the results of the survey are as follows:

1. Composition Role In Industry Cluster

Most businesses in Tegal regency has a role as a supplier industry with a percentage of 54 and 31% other actors act as supporting industries in the industrial cluster.

2. Ability Level Information Technology

In communicating, the business/industry/institution in Tegal most widely used mobile phone (hp). In communications, they rarely use the internet, as most businesses in Tegal not know or do not understand about how to use the internet or they call himself "clueless". The greatest percentage of mobile phone use compared to other means of communication, namely by 48% and the least is internet/satellite is 4%.

Although only a small part in Tegal regency industry players who use the internet, but they used it well. Most use the Internet to communicate with its partners. But it is unfortunate few who use the Internet to exchange knowledge. Whereas it is necessary to supplement existing knowledge.

3. Knowledge Held

Of the total respondents only 33% said that the knowledge possessed sufficient, while largely claimed to have sufficient knowledge to support their performance.

4. Acquire Knowledge Mechanism

In improving the performance of companies/institutions in Tegal, how they obtain information/knowledge about largely done with the training, while some are doing internships at other companies. Very few who have to learn on their own initiative.

In improving the knowledge of being owned, they stated that local governments play a very great in helping improve their knowledge. This happens because the local government, especially in Tegal regency relevant agencies often conduct training/workshop. In addition approximately 20% of colleges also play a part in helping to improve the knowledge of the businesses/organizations in Tegal regency.

While all sources of knowledge acquired businesses/organizations in Tegal 25% comes from the district/city, only 17% were from outside the province. Of the respondents search the source of knowledge gained by involving other parties, mostly manifested in the form of a contract with the specialist/expert.

5. Knowledge Management Mechanism

After all the knowledge acquired, 29% of respondents stated that the knowledge that has been gained socialized to other employees in the environment where they work [7]. While only about 24% of respondents were directly test/apply the knowledge they have acquired.

The businesses/organizations in Tegal regency in documenting the knowledge they gain, the most widely documented in book form. Approximately 26% of respondents documented through journals, only 5% said that they were documenting the knowledge gained in the form of a website. In fact they do not have a database system for managing information/knowledge acquired.

Some 29% of respondents said that the constraints they face in acquiring and developing the knowledge needed due to its limited human resources and a lack of technology and existing infrastructure. But the most widely encountered obstacles are insufficient charge.

6. Mechanisms Dissemination of Knowledge

(a) Internal Sharing

Of the total respondents, the majority of states that there is no mechanism to share knowledge internally for employees. Most respondents were in the company/institution is no mechanism to share knowledge internally, they use the periodic training as a way to share this knowledge, followed by conducting a limited training. Nothing at all who use mailing lists as a way to share knowledge.

5 Conclusion

Several conclusions can be drawn from the discussion in the previous chapters are as follows:

a. Knowledge management is an approach which is based on the understanding that the task of the organization, which is understood by both the reuse of knowledge, and how the success of the re-use of knowledge that has been created.
b. Knowledge management is a series of activities that are used by the organization or company to identify, create, describe, and distribute knowledge for reuse, unknown, and studied in the organization. This activity is typically associated with an objective organization and aimed to achieve a specific outcome such as shared knowledge, improved performance, competitive advantage, or higher levels of innovation.
c. Business and Technology Incubator (BTI) is one of the alternative models of business development through incubation system that can help the problems faced by entrepreneurs in order to create and grow a business world that is robust, efficient, especially for Small and Medium Enterprises. Or in other words, a business incubator and technology is an institution that provides guidance to new businesses by providing support facilities and the transfer of technology and business.
d. Incubator role is very important for entrepreneurs/SMEs beginners (tenant innovative technology-based), as Incubator is designed to assist new and emerging businesses become independent entrepreneurs through a series of integrated assistance include the provision of office facilities, production testing, market testing, management consulting, technology, marketing and finance, training, and the creation of business networks both locally and internationally.
e. The incubation process stages tenant to tenant graduation (graduates) can be simplified into three (3) successive phases of activity, namely; Pre-incubation phase, the incubation stage, and Stage Post Incubation.

References

1. McInerney, C.: Knowledge management and the dynamic nature of knowledge. J. Am. Soc. Inf. Sci. Technol. **53**(12), 1009–1018 (2002)
2. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, Oxford (1995)

3. Dalkir, K.: Knowledge Management in Theory and Practice. Elsevier Butterworth–Heinemann, Oxford (2005)
4. Tiwana, A.: The Knowledge Management Toolkid: Practical Technique for Building a Knowledge Management System. Prentice-Hall, Inc., London (1999)
5. Foster, R.: Innovation: The Attackers Advantage. Summit Books, New York (1996)
6. Drucker, P.: Innovation and Entrepreneurship Management. Prentice-Hall, New York (2005)
7. Firestone, J.M., McElroy, M.W.: Doing Knowledge Management, The Learning Organization Journal, Vol. 12, No. 2. Emerald Group Publishing, Ltd. (2005)

Author Biographies

Dr. Ir. Jarot S. Suroso, Associate Professor at Information Systems Management Department, BINUS Graduate Program-Master of Information Systems Management, Bina Nusantara University Jakarta, Indonesia. His major research interests include management information system, competitive intelligence, knowledge management, e-learning, multimedia and research methodology. He can be reached at jsembodo@binus.edu.

Abba Suganda Girsang, Lecturer at Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University Jakarta, Indonesia. His major research interests include intelligence system, and data mining. He can be reached at agirsang@binus.edu.

Dr. Ford Lumban Gaol, Lecturer at Computer Science Department, BINUS Graduate Program-Doctor of Computer Science, Bina Nusantara University Jakarta, Indonesia. His major research interests include intelligence system and soft computing. He can be reached at fgaol@binus.edu.

Competition Boosts Innovativeness: The Case of CMS Evolution

Andrzej M. J. Skulimowski[1,2(✉)] and Inez Badecka[1,2]

[1] AGH University of Science and Technology, Chair of Automatic Control and Biomedical Engineering, Decision Science Laboratory, Al. Mickiewicza 30, 30-059 Krakow, Poland
ams@agh.edu.pl, inez.badecka@pbf.pl
[2] International Centre for Decision Sciences and Forecasting Progress & Business Foundation, J. Lea Street 12B, 30-048 Kraków, Poland

Abstract. This paper presents two software evolution models where the intensity of innovation plays a primary role. The first model describes the generation of new releases, irrespective of whether they bring an essential software improvement to its users or not. The other model takes into account the emergence of essential new functionalities in the releases. Both models assume additionally that the innovation leaders follow a different innovation strategy than their competitors. The leaders' assumption is that the competing teams follow a catch-up strategy without making radical improvements. There emerge quasi-linear econometric models with vector autoregression. These models are exemplified, calibrated and validated based on historical release logs published by three major open-source CMS providers. Then they are applied to predict further system functionality trends until 2025 and beyond. Model comparison makes it possible to find a relation between innovativeness and creativity measurements, where the latter is derived from the proportion of essential innovations in all releases.

Keywords: ICT foresight · Seed technology · Content management systems
Vector autoregression · Innovativeness quantification

1 Introduction

Enterprise innovation policy usually reflects a compromise between the market success potential of new products and features, as well as innovation-related expenditure and mobilization of human resources. While finding an innovation policy equilibrium depends mostly on financial factors, the deployment of innovation policies may differ in various companies, depending on attitude towards innovativeness and the overall innovation culture. The fastest generation of innovations can be observed in information and communication technologies (ICT), specifically in the software sector [15]. Studying this phenomenon will also allow us to better understand the evolution of the information society.

While IT deployment trajectories in large enterprises have been extensively studied by many researchers cf. e.g. [1, 3, 4], and the related demand for ICT and ICT innovation is well-known to market experts, the role of innovation for the small and

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_16

medium-sized enterprises (SME) and freelance programmer markets is yet to be investigated. The results of recent enterprise trends research [20], cf. also [21], indicate that the most important ICT functionality is the information and communication capabilities offered by Web 2.0 CMS (Content Management System) technologies, which are likely to be expanded further to cater for additional business needs. This observation led us to introduce the idea of a '*seed technology*' [22]. The latter term refers to a particular technology first implemented in an enterprise, and then followed by further compatible technologies. The concept primarily applies to ICT, with CMSs often playing the seed technology role for SMEs in different sectors of the economy. CMSs are well-known for their innovation generation intensity, which is a key factor in software evolution. In addition, the domination of open-source CMSs allowed us to eliminate factors related to commercial interests and marketing, which might obscure the role of creativity. This justified the selection of CMSs as a playground for investigating innovation generation models and as one of main drivers in the development of Enterprise Resource Planning (ERP) software for SMEs.

The main purpose of a typical CMS is to allow a non-expert user to edit the content of a website and customise its appearance via a user-friendly interface, called an administration panel. The information content of a website is thus separated from the technical aspects of its appearance. CMSs generate web pages automatically or semi-automatically, and the information entered by an authorized editor is stored in a database. The CMS then generates dynamic webpages based on the content of this database and on a particular page design template. This allows for a more flexible and convenient content management than with static HTML files. Thanks to the widespread availability of CMSs, in particular the open source CMS software, the development of semantic web was fostered during the recent decade. Depending on the target applications, CMSs can be classified into Content Management Framework, Page-Based Systems, Module-Based Systems, Content Object Systems, and Enterprise Content Management (ECM), cf. [19, 23]. The latter class of systems is focussed around an integrated information resource management systems in enterprises or institutions. The ECM based on CMS most closely implement the ERP development principles [22] that were the motivation for the models presented in this paper.

A rapid growth trend is observed in the development of dynamic control methods of website content and other CMS functionalities. In Sects. 2 and 3 it will be shown how this may be driven by the competition among the main open-source CMSs developer teams. These results differ slightly from those presented in [22], where the selection of 'essential' software versions to the competition model was more tolerant.

2 The Evolution of Content Management Systems

In this and the following sections, we present two innovation-driven software evolution models of the three most popular CMSs, namely WordPress [25], Joomla [13], and Drupal [6], cf. also [19]. The first model describes release generation irrespective of whether essential improvements of software functionality have been made. The second model takes into account the emergence of user-oriented essential new functionalities in the new releases. The terms '*essential*' release, *improvement of functionality*, mean that the change of software system concerned contained an innovation *relevant to*

the users—a subjective assessment, which has been verified independently by several CMS practitioners, as reported in [23]. For example, a release that contains merely a bug elimination is not essential, even if this is relevant to the users.

Based on the release logs, both models assume that Drupal, the innovation leader, does not take into account the activity of the other CMS providers when making their innovation development decisions. The underlying leaders' assumption is that their competitors follow a catch-up strategy and do not make any radical technological improvements. Both models use quasi-linear vector autoregression to predict further functionality evolution prospects until 2025 and beyond. Model comparison makes it possible to find a relation between innovativeness and creativity measurements, where the latter is derived from the proportion of essential innovations in all releases.

Unlike the classical Lehman's software evolution models [17] and their extensions (cf. e.g. [11]), the modelling approach presented in this paper focuses on software users rather than the developers. Since the users do not necessarily know the technical details of codes, even the simplest characteristics such as code size or the number of defects that are key factors in Lehman-type models cannot be admitted as model variables. Instead, the observable variables are the number and time of releases, the functionalities offered in each release, the number of software modules, the number and opinions of users etc. First two of them were included in the models presented in this paper. Under some more specific assumptions, the key question, namely which of the competing open-source systems is most prospective from the user's point of view can be answered. It turns out that the developer's creativity, manifested in the ability to invent new functionalities and architectures, plays a key role in ensuring the competitiveness of software offered.

2.1 The Notion of a Seed Technology

Growing demand creates competition among suppliers as well as concepts and ICT architectures. It results in the implementation of novel and more convenient solutions. The proper choice of the ICT development strategy in an enterprise, which best suits its needs and goals, becomes essential. In the case of an SME, this technological trajectory may strongly depend on the *seed technology*, defined as the *first technology from a specific technological area, which is implemented in the given enterprise*. In many SMEs, this is the technology that was chosen to build the company's website, as its first step into the ICT area. Therefore, choosing the right CMS can be extremely important to the company's future ICT strategy. The notion of seed technology also explains the difference between micro, small and larger companies as regards ICT deployment and alignment. In micro and small enterprises there is often a single initial ICT, while large companies implement usually several technologies from the same area in different departments independently. In addition, the needs analysis for a recent foresight project [21] revealed that many micro- and small enterprises plan to build their ERP system as an ECM with an open-source CMS [23]. Such systems can be referred to as ERP-CMS [23] to highlight the role of CMS in their development. In the case of medium-sized enterprises, this process may be more diversified and largely depends on the sector of the economy. Large organizations are most likely to implement customized ERP systems supplied by one of the few main providers [8].

2.2 A Brief Presentation of Three Most Popular Open-Source CMS

For a detailed analysis of the innovation process, we have selected the three CMSs that comprise most of the open-source content management market [2]. Their characteristics are briefly outlined below.

Drupal [6] is a well-established CMS developed as an open-source platform since 2001. It offers advanced features that allow its users to build diversified web applications, from social media such as blogs, forums and newsletters to complex collaborative environments and enterprise management systems. Drupal has a modular architecture that facilitates innovation development. Drupal-specific taxonomy makes it easier to organize content according to predefined categories. The CMS is equipped with a system of PHP templates ("skins") and many special tools, which are useful in the business sector. Moreover, Drupal Open Enterprise distribution has been developed (https://www.drupal.org/project/openenterprise) with the aim to facilitate the use of Drupal for simple business purposes. It is equipped with a specialized set of modules supporting the ERP/ERM functionalities of applications built with Drupal.

Joomla (or "**Joomla!**") was initially released in 2005 as a successor to Mambo, the CMS that gained popularity owing to its flexibility and reliability, but its further development was abandoned. Like Drupal, Joomla has a modern modular structure that facilitates the addition of new functionalities implemented as additional modules. This architecture, which makes it easier to build new releases, contributes to the increasing popularity of Joomla and allows enterprises to build CRM (Client Relationship Management), ERP/ERM and general e-commerce applications (cf. [13]).

WordPress [25], currently the most popular CMS, is probably the easiest to install, operate and configure. Moreover, to its popularity contributes the fact that it is particularly well-suited for blogging. WordPress is written in PHP and uses a MySQL database. It has been created by a community gathered around the Wordpress.org project and is distributed under a GNU—General Public License. Due to its growing range of functionalities and ease of use, WordPress is also increasingly popular in the business sector. With the launch of its enterprise-oriented commercial spin-out, WordPress VIP, predictions that forecasted CMS development trending towards increasingly enterprise-management oriented systems [23] were fulfilled, cf. also https://vip.wordpress.com/2014/05/22/enterprise-wordpress-content-trends.

Having analyzed all hitherto Drupal, Joomla, and WordPress releases, it emerged that over its entire development history, Drupal has been the leader in implementing innovations, cf. [22, 23, 24] and the next section. This fact has an important impact on building CMS innovation process models, which will be studied in more detail in the following sections of this paper.

2.3 Characteristics of the CMS Innovation Race

Table 1 presents information about the history of new releases that contain essential functionalities of the aforementioned CMSs, in form of a timeline. The data was gathered and verified in [21] and [23]. This is the main input to the evolution models presented in the Sect. 3.

Table 1. Intervals in months between the introduction of improvements to major CMSs in relation to the innovation leader

Functionality/milestone description	WordPress		Joomla		Drupal	
	$f_{1,k}$	$\tau_{1,k}$	$f_{2,k}$	$\tau_{2,k}$	$f_{3,k}$	$\tau_{3,k}$
Release of the first stable version of the system	28	28	56	56	0	0
Improvement of the positioning quality, swapping page elements, installer from the browser level, recursive tree, menu/submenu	1	36	21	56	0	35
Modular system architecture, plug-ins, improvement of graphic file management system	38	40	54	56	0	2
Advanced language system, translation of the administration panel, world-wide expansion	1	40	6	57	0	2
Comprehensive page-view management system, components, widgets, templates and skins	10	49	22	61	0	39
User accounts management, filtration, authentication systems, anti-spam, user accounts profiling	33	58	41	58	0	17
Advanced tools: WYSIWYG, RSS, metadata, Ajax, positioning	8	67	10	61	0	51
Integration with external tools, blogger-type systems, and advanced file management	13	68	14	61	0	47
Modern script libraries like JQuery, automatic plug-in installer	4	84	43	114	0	72
Emphasis on security, authentication, compliance with standards, data encryption	37	103	3	61	0	58
Drafts, notes, advanced caching (buffering) and backup techniques	19	115	26	115	0	88
Multisites and multiposting—multiple sites management from a single installation	15	120	19	116	0	97
Admin panel development, dashboard, desktop management	6	134	4	124	0	120
A breakthrough in the source code structure, compliance with PHP 5 and MySQL 5	3	135	3	127	0	124
[*last stable CMS version release*]	0	140	0	132	0	129

Source Own collation based on https://api.drupal.org/api/drupal/CHANGELOG.txt/7/source, https://wordpress.org/about/roadmap/, https://docs.joomla.org/Category:Version_History

The first column of Table 1 contains the specifications of relevant CMS functionalities. The second column shows the time parameters $f_{1,k}$, $k = 1,...,K$, which describe the delay of the k-th WordPress functionality introduction in comparison to the time when the first system had introduced it, for all K essential functionalities specified in Table 1. As we already mentioned in the previous section, in the case of the CMSs analyzed in this paper, the most advanced system has always been Drupal.

The *4th* column contains a similar delay in functionality introduction by Joomla, for the above reason also in comparison to Drupal. The latter is characterized by the data in the *6th* and *7th* columns. The *6th* column contains zeros, because each functionality was introduced first by Drupal. The *3rd*, *5th* and *7th* columns contain the accrued time in months $\tau_{I,k}$ that elapsed from the first release of Drupal to the appearance of the functionalities specified in the first column in Wordpress, Joomla, and Drupal, for $i = 1, 2, 3$, respectively.

As an illustrative example, 28 months after the first version of Drupal was launched in January 2001, the first version of WordPress appeared (i.e. in May, 2003). The next, improved version of WordPress was released in January 2004, i.e. eight months after the first version, thus the next value in the second column of the table is 28 + 8 = 36.

The identification of relevant functionalities and milestones illustrated in Table 1, together with the version data, form the basis on which to build two technological evolution models mentioned in the introduction and presented in the next Sect. 3.

3 Innovativeness Scenarios

There are no explicit and widely-accepted models which describe the feedback between the appearance of open source software innovations releases and users' acceptance of them. General discrete-time innovation diffusion models usually refer to the well-known Norton-Bass model [12, 14]. When applied to open-source software, in the absence of price equilibria, the major role that the essential technological progress plays must be taken into account. Therefore, we will investigate a new class of models suitable for a simplified customer-oriented CMS evolution modelling, where the dependent variable is time of market release of an innovation by the development team. Explanatory variables are time lags between consecutive essential releases (cf. Sect. 2), of the same as well as of other systems.

The first model (Sect. 3.1), which is based on the ideas presented in [22, 23] will serve as a benchmark. We will take into account all consecutive releases of each system that bring innovations relevant to the user community. 'Patchwork' releases, which only eliminated a bug or solved a security problem related to a new web threat, were excluded from consideration. The second model relies on the assumption that the innovation dynamics is driven by the most relevant functionalities. Therefore, the time lapses since the introduction of the same (or equivalent) functionality by the competitors are used as explanatory variables. This model is presented in Sect. 3.2.

Moreover, based on the empirical evidence provided in Table 1 in Sect. 2, in both models we have assumed that the technological leader (Drupal) team developed their system autonomously, without taking into account the innovations introduced by other systems. Due to the lack of situations during the observation period, where Drupal

could react to the leadership of another system, observation data to estimate the coefficients $b_{3,j}$ in Eq. (1a)–(1c) below were not available. Consequently, these coefficients that would describe the hypothetical Drupal's team decisions dependence on other systems and the overall model that would not admit the above Drupal's 'autonomy' assumption could not be estimated in a statistically relevant way. However, the missing coefficients could be defined by analogy, without assessing their goodness of fit. Such functionality evolution model and its comparison to the Scenario II (cf. Sect. 3.2 of this paper) has been presented in [24].

Since the construction of the above models is based on ex-ante assumptions that can be finally tested and verified after the models are built and running, it is justified to call these models '*Scenarios*' referring to the technological foresight vocabulary [21]. The first one can be termed "*Incremental Progress*", while the second will be termed "*User Benefit*" Scenario, as the relevance of the new systems' functionalities is assessed based on an anticipated user community response. In the next subsections, we will present further assumptions that led to the formulation of the econometric models, and forecasts of CMS technology development generated in both scenarios.

3.1 Innovation Generation in Scenario I

This subsection presents the baseline model that corresponds to "Incremental Progress" Scenario I. It takes into account the data provided in Table 1 supplemented by *3* (Drupal, WordPress) and *7* (Joomla) releases with minor, yet essential improvements. As mentioned above, it simply fits the available data to a quasilinear econometric model of software release generation derived in [22].

Let $x(i)$, $y(i)$, and $z(i)$ denote time intervals between next (i-th) release of WordPress, Joomla, and Drupal, respectively. The main idea of the model used in Scenario I is to investigate the relationship between the time of next improvement introduction by a particular development team and the frequency of similar innovations in the past, created by the same as well as by competitor teams.

We assume that the time interval of the subsequent system improvement depends linearly on the $n - 1$ previous time intervals of system improvements as well as on the frequency of improvements emerging in other systems. Based on a timeline analysis (Table 1), which presented Drupal as the leader in the implementation of all previous innovations, we have additionally assumed that the Drupal development team operates autonomously, without taking into account subsequent versions of WordPress and Joomla, because the innovations of the other two platforms are already present in Drupal. These innovations are not influenced by competitive pressure on the Drupal's development team (such pressure may, however, result from introducing marketing or organisational innovations, which are not necessarily linked with the technological innovation introduced by the successive versions of software). These assumptions lead to the formulation of the following model of innovation creation in all three systems analyzed in this paper:

$$\begin{aligned} x(k+1) = {} & a_{1,1}x(k) + a_{1,2}x(k-1) + \cdots + a_{1,n1}x(k-n_1+1) \\ & + b_{1,2}v_{1,2}(k) + b_{1,3}v_{1,3}(k) + c_1 + \varepsilon_1 \end{aligned} \tag{1a}$$

$$y(k+1) = a_{2,1}y(k) + a_{2,2}y(k-1) + \cdots + a_{2,n2}y(k-n_2+1) + b_{2,1}v_{2,1}(k) + b_{2,3}v_{2,3}(k) + c_2 + \varepsilon_2 \quad (1b)$$

$$z(k+1) = a_{3,1}z(k) + a_{3,2}z(k-1) + \cdots + a_{3,n3}z(k-n_3+1) + c_3 + \varepsilon_3, \quad (1c)$$

where:

$v_{1,2}(k)$—average frequency of introducing of a new version of Joomla calculated on the basis of $P_{1,2}(k)$ time intervals between versions of this system directly preceding the k-th version of WordPress;

$v_{2,1}(k)$—average frequency of introducing of a new version of WordPress calculated on the basis of $P_{2,\ 1}(k)$ time intervals between versions of this system directly preceding the k-th version of Joomla;

$v_{j,3}(k)$, for $j = 1, 2$—average frequency of introducing of a new version of Drupal calculated on the basis of $P_{j,3}(k)$ time intervals between versions of this system directly preceding the k-th version of WordPress—for $j = 1$, and Joomla—for $j = 2$;

ε_i are normally distributed residual errors with variance δ_i^2, i.e. $\varepsilon_i \sim N(0,\delta_i)$, for $i = 1, 2, 3$;

$a_{i,m}$, $b_{i,j}$, and c_i, $i = 1, 2, 3$; $m = 1,\ldots n_i$, $j = 2, 3$ are the regression coefficients, n_i—the autoregression ranks. Specifically, the autoregression coefficients $a_{i,1},a_{i,2},\ldots,a_{i,ni}$ describe the dependence of the length of the time interval needed for a new version of the i-th system to appear on the time lapses between the previous versions introduced in the same system, i.e. WordPress, Joomla, and Drupal, respectively, for $i = 1, 2, 3$.

The coefficients $b_{i,j}$ for $i = 1, 2$, and $j = 1, 2, 3$, $i \neq j$, describe the following relations:

- for $i = 1$, $j = 2, 3$, $b_{i,j}$ is the coefficient of multivariate linear regression which explains $x(k)$ with $v_{1,j}(k)$,
- for $i = 2$, $j = 1$, $b_{i,j}$ is the coefficient of multivariate linear regression which explains $y(k)$ with $v_{2,\ 1}(k)$, and
- for $i = 2$, $j = 3$ $b_{i,j}$ explains $y(k)$ with $v_{1,3}(k)$.

Observe that the nonlinearity in Eq. (1a)–(1c) resides in terms $v_{i,j}(k)$ because k in any of the Eq. (1a) or (1b) must be synchronized with the actual version indicator in the remaining systems. Therefore the indicators used to calculate the frequency $v_{i,j}(k)$ do not depend only on k, but also on the indicator of the current version of the j-th system at the time $t_{k,I}$ when the k-th version of the i-th system was released. For example, if at time $t_{k,1}$ corresponding to the k-th version of WordPress in the Eq. (1a), the other systems released their $k_2(t_{k,1})$-th and $k_3(t_{k,1})$-th versions, respectively, then $v_{1,j}(k)$ will be calculated based on the $k_j(t_{k,1}) - P_{1,j}(k)$, $k_j(t_{k,1}) - P_{1,j}(k) + 1,\ldots, k_j(t_{k,1})$ versions of the *j-th* system, for $j = 2, 3$. This dependence implies a specific order of simulating technological evolution with Eq. (1a)–(1c) so that essential recent versions of all systems are always taken into account (during the observation period) or calculated (during the forecast period) up to the same moment of time. Based on numerical experiments, in the above model we admitted a simplifying assumption, namely as $P_{i,j}(k)$, we could take the constant value $P_{i,j}(k) = 3$, for $i = 1, 2$, $j = 1, 2, 3$, $i \neq j$. This value allowed us to obtain a sufficient significance of the model.

After computing the coefficients of (1a)–(1c) with the least squares method, we will get the regression function relating the expected time of a new innovation release by each system—as dependent variables—to the average time lags between the introduction of technological innovations in all systems. The lags are calculated to the releases directly preceding the latest innovation in the *i*-th system, prior to the *k*-th improvement of the system described by the Eq. (1a) or (1b). The trend drift coefficients c_i, *i = 1, 2, 3* should vanish after the model (1a)–(1c) was integrated a sufficient number of times for autoregression purposes to yield a stationary time series for each system. However, they re-appear when calculating the forecasts for the original time series.

The Eq. (1a), (1b) take into account the pressure of competition. According to the initial assumption, Eq. (1c) describes time lags between subsequent Drupal improvements only. It turned out impossible to find a statistically significant model where the variables were simple time lags between subsequent innovations. From the definition of system variables as time lags, and Eq. (2):

$$x(t) - x(t-n) = [(x(t) - x(t-1)] + [x(t-1) - x(t-2)] + \cdots + [x(t-n+1) - x(t-n)], \quad (2)$$

it follows that the above difference operation is equivalent to the averaging the integrated original time series. With this approach, a stationary time series was found after one integration step. Then, the integrated variables were averaged *s*-times, s = 2, 3, 4…, until both a significant regression function and model coefficients were found. The statistical significance of averaged models was investigated with the *F* (Fisher-Snedecor) statistics [7] and goodness of fit tests with the determination coefficient R^2. Tests confirmed the significance of calculation results for once integrated and triple averaged data, which correspondents to the case where all variables characterizing the frequency of innovations are calculated as the average of three time lags between innovations. Finally, the forecasting model used in Scenario I has the form:

$$x(k+1) = 0.5405\,x(k) - 0.1262\,v_{1,2}(k) + 0.1916\,v_{1,3}(k) + 2.6040 \quad (3a)$$

$$y(k+1) = -0.0291\,y(k) - 0.5882\,v_{2,1}(k) + 2.9171\,v_{2,3}(k) - 13.0582 \quad (3b)$$

$$z(k+1) = 0.7368\,z(k) - 0.2281\,z(k-1) + 4.1404 \quad (3c)$$

As the significance of the regression function does not guarantee that the coefficients are significant, we calculated the confidence intervals for each of them, cf. Table 2. The standard deviations $\delta_{:=}(\delta_1.\delta_2.\delta_3)$ of the residual errors ε_i in Eq. (3a)–(3c), *i = 1, 2, 3*, amount to *(1,56;2,06;2.18)*. To complete the usual statistical analysis of forecasts, we calculated the confidence intervals for the forecasted number of innovations.

Table 2. The confidence intervals for the Scenario I innovation forecasts (α= *0,05*)

	Forecast horizon	Number of innovations	Confidence interval
Drupal	2025-07	35	(33,37)
	2030-05	41	(38,43)
WordPress	2025-05	35	(32,38)
	2030-01	43	(40,46)
Joomla	2025-03	32	(30,34)
	2030-07	41	(39,43)

The forecasting procedure was first run until the last release of the most slowly evolving system reached the year 2025, i.e. the foresight horizon of the project SCETIST [21]. Since the confidence intervals for 2025 were acceptable, it was pursued for the horizon 2030. The resulting forecasts are visualised in Fig. 1.

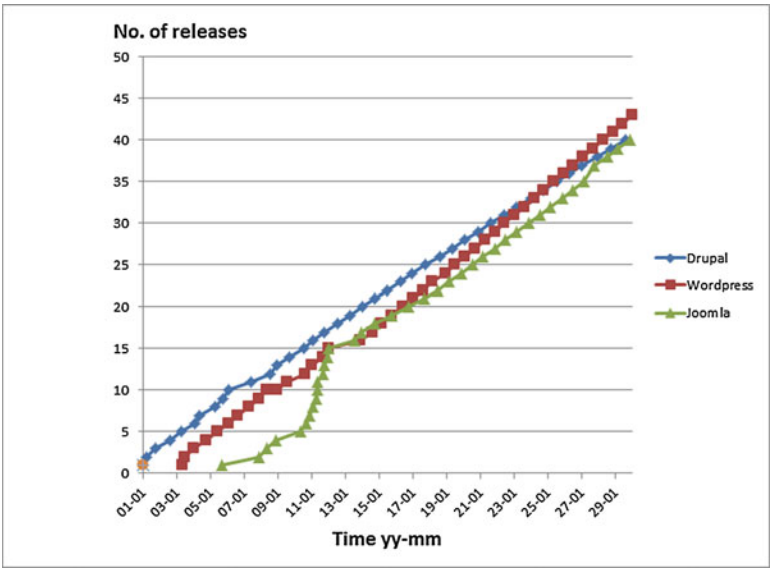


Fig. 1. Forecasted number of innovations for each of the systems investigated in Scenario I

The predicted numbers of releases with essential innovations of each system for 2025 and 2030 with their confidence intervals at level $1 - \alpha = 0{,}95$ are given in Table 2.

The chart shows that Drupal, which was assumed to proceed with its development autonomously, may lose its innovation leadership around 2025. WordPress, the present market leader, will then become the innovation leader and remain so afterwards. Joomla will catch up Drupal in about 10 years from now. By 2025 the WordPress team will implement in its releases the total number of innovations on *6* higher than the numbers of those implemented in Joomla and *2* more than Drupal.

3.2 Innovation Trajectories in Scenario II

The final observation made in the previous subsection, namely that Drupal may lose its leadership, relates to the question of how it would react in such a situation, assuming that all developer teams take into account their competitors' activity when implementing their current innovation policies. As mentioned in Sect. 3.1, we have not yet observed a situation where Drupal was overtaken by another system, so a release generation model cannot be derived from observational data. Instead, it can only be assumed that the Drupal team response would be similar to that of the other teams, which would give us an idea of the missing model coefficients. As we have already mentioned, the model building details in this case are given in [24].

This section presents the Scenario II model, which differs from Scenario I in the preceding section on the assumption that only releases that bring considerable functional progress are to be taken into account, while the independent variables are time lapses since the introduction of the same functionality in other systems. To build the forecasting model, we identified such improvements (cf. Table 1) and calculated the time lags (in months) between subsequent (i-th) essential system improvements for WordPress, Joomla, and Drupal. Moreover, the notation is similar to that used in Scenario I and model (1a)–(1c)–(3a)–(3c), and the case analysed is the same.

Consequently, these assumptions led us to the formulation of an analogous model of essential innovation generation in the above three CMSs as described by Eqs. (1a)–(1c) and (3a)–(3c) above. Model coefficients are provided in Eq. (4a)–(4c) below.

$$x(k+1) = 1.2277\,x(k) - 0.8622\,v_{1,2}(k) - 0.0418\,v_{1,3}(k) + 1.1153 \tag{4a}$$

$$y(k+1) = 0.1067\,y(k) - 0.3561\,v_{2,1}(k) - 0.3718\,v_{2,3}(k) + 19.0255 \tag{4b}$$

$$z(k+1) = 1.1662\,z(k) - 0.6892\,z(k-1) + 4.6565 \tag{4c}$$

The quantity of functionalities implemented in each of the investigated systems forecasted in Scenario II for the years 2025 and 2030 are presented in Table 3 and Fig. 2. Let us observe that a higher accrued number of functionalities of a system may increase its competitiveness when they are aligned with the users' needs.

Table 3. Innovation forecasts with the functionality-based model and their confidence intervals with the p-value α=*0,05*

	Forecast horizon	Number of innovations	Confidence interval
Drupal	2025-09	30	(28,32)
	2030-03	35	(32,38)
WordPress	2025-06	29	(27,32)
	2030-05	34	(31,37)
Joomla	2025-12	26	(24,28)
	2030-10	35	(31,39)

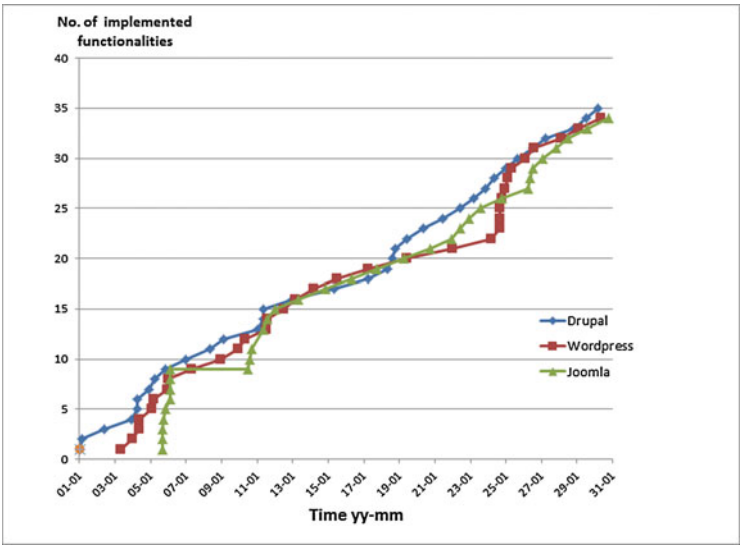


Fig. 2. Forecasted number of innovations for each of the CMSs in Scenario II

The statistical significance of the model used in Scenario II has been determined with the same approach as in Scenario I. Integration performed only once yielded a stationary time series for all three systems, while the triple averaging of input data was necessary to obtain significant regression function and model coefficients. Moreover, the residual error characteristics are similar to (3a)–(3c). This model differs from that presented in [22] as the functionality propagation has been additionally assumed rule-dependent. The empirical developer team rules concerning the order of releasing functionalities already introduced by their competitors were derived from the observations of their behaviour until 2013. The probabilities of applying a rule was assumed to be proportional to the frequency of its occurrence during the observation period. When simulating the future system development the rules were selected with a pseudo random number generator using the probabilities thus estimated.

3.3 A Comparison of Scenarios I and II

To ensure the comparability of essential releases and essential functionalities we have calculated the ratio of average number of essential functionalities and the average number of essential releases for all three systems analyzed, as a function of time. The results are presented in Fig. 3.

The ratios functionality/release were derived from data and the essential releases were concerted to essential new functionalities. The ratios amount to 0.6 for Wordpress and 0.58 for Drupal. For Joomla, there was no single ratio as these decrease in time with a constant rate of 0.3/year in a significant way. For the latter case, a linear functionality/release trend was assumed. A comparison of both scenarios after the above conversion is shown in Table 4. As we can see, there was a considerable difference in WordPress innovative behavior between the scenarios, with a more intensive innovative

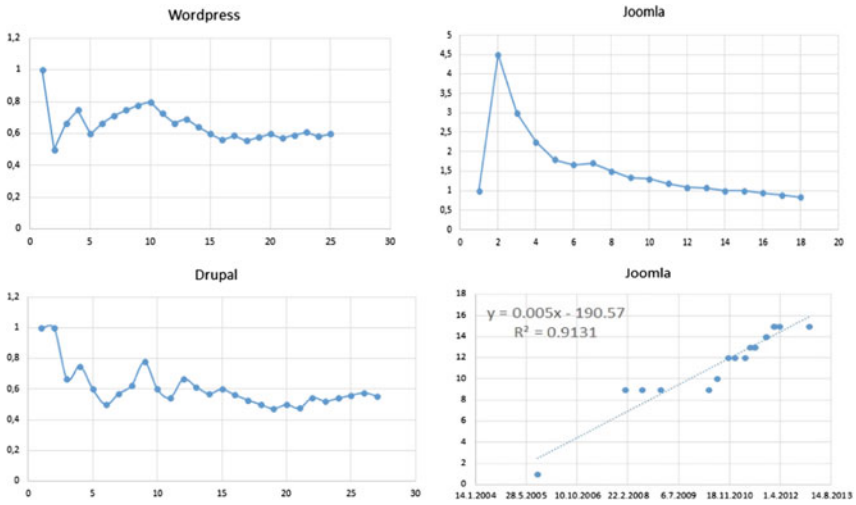


Fig. 3. The ratios of the number of '*essential new functionalities*' and '*essential innovations in new releases*' (vertical axis) for Wordpress (top left), Drupal (bottom left), and Joomla (top right). New release counts are represented on the horizontal axes, except the time trend of the Joomla's new functionalities appearance, which is shown in the bottom right chart

activity in Scenario I. The other two systems generated new releases and functionalities according to the same pattern. This difference may indicate that WordPress considers releasing new versions as more important than new functionalities.

Table 4. Number of innovations in the CMS releases until 2030—a scenario comparison

Forecast horizon	Drupal		WordPress		Joomla	
	Scenario I	Scenario II	Scenario I	Scenario II	Scenario I	Scenario II
2020	28	23	26	20	25	21
2025	35	29	35	28	32	26
2030	41	35	43	34	41	35

Another comparison is provided in Table 5, which shows the value of a new indicator "*CMS innovativeness index*", calculated as ratio of the forecasted yearly average number of essential innovations (new functionalities) during the period of 2012–2025 (4th row in Table 5) and the observed average during the period of 2005–2012 (3rd row).

Table 5. The innovativeness index based on forecasted average number of essential innovations in the CMS releases until 2030—Scenarios I and II compared

	Drupal		WordPress		Joomla	
	Scenario I	Scenario II	Scenario I	Scenario II	Scenario I	Scenario II
No. of innovations in 2005–2012	10	7	10	10	14	14
No. of innovations in 2012–2025	17	15	20	14	18	11
Mean innovations no. per year in 2005–2012	1.43	1	1.43	1.43	2	2
Mean innovations no. per year in 2012–2025	1.31	1.15	1.53	1.08	1.38	0.85
Average innovation growth (innovativeness index)	0.92	1.15	1.07	0.76	0.69	0.43

The above results show that the assumptions used to define both scenarios may have a remarkable influence on the innovativeness image of the CMSs concerned. The order of the CMSs according to their innovativeness index is different in both scenarios. Drupal, the innovation leader, may lose its leadership to Wordpress if the market strength of the competing systems is more likely to depend on the number of versions released (Scenario I). It will remain a strong player in either case. The Joomla's innovativeness scores are lowest among three CMSs in both cases. However, an insight into the business sector targeted market plans of all systems allows us to conclude that innovations related to the ERP/ERM released first by Drupal will play the driver role in the enterprise CMS market until at least year 2025.

4 Discussion and Conclusions

This paper presents the forecasts of software innovation generation exemplified by three most popular open source CMSs in two scenarios following two different assumptions concerning the innovation policies of their development teams. According to the first assumption, every new release of a competing system with a new functionality can be taken into account as a potential market threat. The model based on the second assumption filters out minor software modifications and allows the developer teams to react to the introduction by competitors of essential new functionalities only. Both models exclude changes made due to software maintenance only.

The research reported in [23] indicates that further progress in CMS development will be mainly due to an expansion in the enterprise market by including ERP/ERM functionalities and modules. Trend extrapolation with vector autoregression performed in Sect. 3 allowed us to predict CMS functionality development prospects until 2025. According to these forecasts, it emerged that even the CMSs that currently appear less advanced but are still in the innovation race will be able to implement

enterprise-management functionalities until 2020. In one of the business scenarios under consideration, the now less technologically advanced WordPress may catch up or even surpass the current enterprise CMS leader Drupal by 2025. There exists a synergy with other CMS and ERP development trends, such as the further development of dedicated open source ERP [26]. Based on the hitherto market reactions, which were rather moderate, we suppose that the latter class will not be a competitive solution for non-IT enterprises due to the relatively high programming literacy requirements compared to the database solutions created with Drupal or Plone (cf. www.plone.org). It is more likely that other CMSs will follow the leaders and reach a sufficiently advanced level of enterprise resource management modules during the next decade.

From the above reasons one can consider the open-source CMSs as both a seed technology (cf. Sect. 2.1) and a key technology for the development of integrated enterprise applications in SMEs [9]. The advanced functionalities of leading CMSs will allow the SMEs to integrate their ICT systems around their websites and intranets [5]. We argue that this will lead to a different way of enterprise application integration in SMEs [16] than the gradual implementation of full-scope professional ERP systems provided by specialized software companies, starting from installing 'lighter' versions of software built around their accounting, inventory and sales modules. Contradicting this expectation, many SMEs increasingly use open-source CMSs with their emerging ERM extensions. This trend is driven by open-source SQL database development and their wide employment in CMSs, as well as by the common use of Apache http servers and their API. This makes it possible to combine different database applications that are typically used to manage enterprise resources with web content management, CRM [18] and internal content exchanges via an enterprise intranet. These have been referred to in Sect. 2 as ERP-CMSs.

The emergence of ERP-CMSs is a relevant though sometimes neglected trend of ICT deployment in SMEs. Yet, it may have an important impact on market shares within the software industry [9, 10], with vendors and authorized resellers of professional ERP solutions losing out to CMS developers. A considerable fraction of SMEs will be able to develop and/or upgrade ERP-CMS software with their own human resources. In Sect. 2, we presented the features of the most popular open-source CMSs that made possible the emergence of ERP-CMSs. However, the main drivers of ERP-CMSs are the overall economic development, the average ICT infrastructure expenditure in a given sector of the economy as well as the average degree of ICT literacy in a country or region.

Acknowledgements. This research was supported by the foresight project "Scenarios and Development Trends of Selected Information Society Technologies until 2025" (SCETIST) co-financed by the ERDF within the Innovative Economy Operational Programme 2006–2013, Contract No. WND-POIG.01.01.01-00-021/09.

References

1. Asprey, L., Middleton, M.R.: Integrated document management for decision support. In: Burstein, F., Holsapple, C. (eds.) Handbook on DSS, vol. 1, pp. 191–206 (2008)

2. Builtwith (web technology lookup). trends.builtwith.com/cms. Accessed Dec 2016
3. Cardoso, J., Bostrom, R.P., Sheth, A.: Workflow management systems and ERP systems: differences, commonalities, and applications. Inform. Tech. Manag. **5**, 319–338 (2004)
4. Castellina, N.: SaaS and Cloud ERP Trends, Observations, and Performances. Aberdeen Group (2011)
5. Devos, J., Landeghem, H., Deschoolmeester, D.: Using bricolage to facilitate emergent collectives in SMEs. In: Proceedings of the 6th European Conference on Information Management and Evaluation, pp. 82–90 (2012)
6. Drupal web site. Drupal Foundation. www.drupal.org. Accessed Feb 2017
7. Fox, J.: Applied Regression Analysis, Linear Models and Related Methods. Sage (1997)
8. Grabski, S., Leech, S., Schmidt, P.: A review of ERP research: a future agenda for accounting information systems. J. Inf. Syst. **25**(1), 37–78 (2011)
9. Hailu, A., Rahman, S.: Evaluation of key success factors influencing ERP implementation success. In: Proceedings of the 2012 IEEE 8th World Congress on Services, pp. 89–91 (2012)
10. Hallikainen, P., Kivijarvi, H., Nurmimaki, K.: Evaluating strategic IT investments: an assessment of investment alternatives for a web content management system. In: Proceedings of the 35th HICSS, pp. 2977–2986 (2002)
11. Herraiz, I., Rodriguez, D., Robles, G., Gonzalez-Barahona, J.M.: The evolution of the laws of software evolution. A discussion based on a systematic literature review. ACM Comput. Surv. **46**(2), Article #28, 28 (2013)
12. Jiang, Z., Jain, D.C.: A generalized Norton-Bass model for multigeneration diffusion. Manag. Sci. **58**(10), 1887–1897 (2012)
13. Joomla web site. www.joomla.org. Open Source Matters, Inc. Accessed Feb 2017
14. Kapur, P.K., Sachdeva, N., Singh, O.: Generalized discrete time model for multi generational technological products. In: Proceeding of the International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 717–723 (2015)
15. Keller, A., Hüsig, S.: Ex-ante identification of disruptive innovations in the software industry applied to web applications: the case of Microsoft's vs. Google's office applications. Technol. Forecast. Soc. Change **76**, 1044–1054 (2009)
16. Lee, J., Siau, K., Hong, S.: Enterprise integration with ERP and EAI. Commun. ACM **46**(2), 54–60 (2003)
17. Lehman, M.M.: Programs, life cycles, and laws of software evolution. Proc. IEEE **68**(9), 1060–1076 (1980)
18. Leina, Z., Tiejun, P., Guoyan, R., Chengbin, F.: Development and implementation of ERP/CRM system based on open source software to small and medium-sized enterprise in China. In: ICICTA, 2008, vol. 2, pp. 725–730 (2008)
19. Patel, S.K., Rathod, V.R., Parikh, S.: A statistical comparison of open source CMS. In: 3rd International Conference on Trendz in Information Science and Computing (TISC), pp. 182–187 (2011)
20. Rosso, S.: Trends in Enterprise WordPress Content (2014). https://vip.wordpress.com/2014/05/22/enterprise-wordpress-content-trends. Accessed Dec 2015
21. Skulimowski, A.M.J. (ed.): Scenarios and Development Trends of Selected Information Society Technologies until 2025 (SCETIST). Final Report [in Polish]. Progress & Business Publishers, Kraków. www.ict.foresight.pl (2013)
22. Skulimowski, A.M.J., Badecka, I.: Competition Boosts innovativeness: a quantitative software evolution model. In: Theeramunkong, T., Yuizono, T., Skulimowski, A.M.J. (eds.) Proceedings of the 10th International Conference on Knowledge, Information and Creativity

Support Systems, Phuket, Thailand, 12–14 Nov 2015, pp. 433–446. Thammasat University (2015)
23. Skulimowski, A.M.J., Badecka, I.: Development prospects of content management systems and their applications in SMEs [in Polish]. In: Skulimowski, A.M.J. (ed.) Trends and Development Scenarios of Selected Information Society Technologies, Chapter 5, Advances in Decision Sciences and Future Studies, vol. 1. Progress & Business Publishers, Kraków (2016)
24. Skulimowski, A.M.J., Badecka, I.: Software innovation dynamics in CMSs and its impact on enterprise information systems development. In: Tjoa, A.M. et al. (eds.) CONFENIS 2016, LNBIP 268, pp. 309–324. Springer International Publishing AG (2016). https://doi.org/10.1007/978-3-319-49944-4_23
25. WordPress web site. www.wordpress.org. WordPress. Accessed Feb 2017
26. Zhang, Y.: An analysis on the ERP application status of the small and medium, manufacturing enterprises. In: Luo, Q. (ed.) 2010 Second International Conference on Communication Systems, Networks and Applications Proceedings, 29 June–01 July 2010, Hong Kong, vol. 2, pp. 165–168 (2010)

Large Scale Mobility Analysis: Extracting Significant Places Using Hadoop/Hive and Spatial Processing

Apichon Witayangkurn[1(✉)], Teerayut Horanont[2], Masahiko Nagai[1], and Ryosuke Shibasaki[1]

[1] Center for Spatial Information Science, The University of Tokyo, Komaba, Tokyo, Japan
{apichon, nagaim}@iis.u-tokyo.ac.jp, shiba@csis.u-tokyo.ac.jp

[2] Sirindhorn International Institute of Technology, Thammasat University, Bangkok, Thailand
teerayut@siit.tu.ac.th

Abstract. We describe extracting people significant places for mobility analysis from a real-world large scale dataset collected by mobile operator. The total data consisted of 9.2 billion GPS points including approximately 1.5 million individual user trajectories accumulated for a year. We conducted the experiments on the dataset by using stay point extraction and density based cluster to extracting significant places from a sparse dataset. We also proposed an approach to derive types of locations especially home and work place by using classification features and inference model. The relevant features including ranking in clusters, number of days that data appeared, night time, and day time were identified and evaluated. Several inference models are evaluated in the experiment. With limited number of ground truth data, Random Forest model could achieve 99.2% accuracy for inferring home and work location. Additionally, Spatial Population Census were employed to indirectly compare the classification results with ground truth. Furthermore, to enable real-world application, we presented a technique to utilize Hadoop/Hive, a cloud computing platform, allowing full-scale data processing. As a result, the proposed method is able to discover home and work locations of users with positive results after checking the census. In addition, by using Hadoop platform, an extraction process is able to perform on the whole dataset with only about 1.53 days compared with a single application which took 32.73 days.

Keywords: GPS trajectories · Mobile phone · Mobility analysis
Spatial data mining · Hadoop

1 Introduction

With the increasing popularity of mobile devices embedded with positioning function, people are able to acquire their location as well as location history easily. Besides, it may lead to have massive spatio-temporal dataset of people's trajectories. The analysis

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_17

of people's trajectories can discover knowledge about human mobility such as favorite places, daily activity and life pattern. By mining long period of user's trajectories, the most often visited place can be determined such as Home and work location. In order to obtain position data, it usually use an application installed in a mobile device to collect data and most of the existing researches used this method with limited number of participating volunteers. Despite that, data can be also obtained by mobile operators which are more effective than mobile users because it can increase easily the number of users and continues collecting data for long periods. In the real-world operator-based data collection, point data are not so reliable and often have errors and missing points. Although, significant place extraction has been investigated for years, all existing research studies used a dataset collected by volunteers with high rate of data such as 1 point/sec and 1 point/min which is far different from our large-scale dataset. A suitable technique is needed to handle such issues and by processing large scale dataset, many applications can be developed such as person trip surveys, population estimation and important event detection with effect estimation. In summary, this paper aimed to preliminarily explore the whole dataset for the possibility of extracting significant places from sparse temporal data as well as to introduce a method of processing large scale dataset that increase processing speed for spatial data. The contribution of the paper lies in four aspects:

- We evaluated classification features using feature selection model to identify suitable features for extracting type of places.
- We evaluated several inference models to infer type of places and also proposed an easy implementation technique.
- We proposed using Hadoop/Hive with spatial processing functions to process location extractions from a large scale dataset.
- We developed location extractions function on Hive to allow large scale processing.
- We evaluated the significant place extraction methods on sparse temporal data by using operator-based large scale mobile GPS datasets collected from about 1.5 million users in Japan over one year period.

2 Related Work

Mining the trajectories of people has become an attractive research area during the past years. Most of the studies have been focused on extracting meaningful place of people [1, 2], understanding people moving pattern [3], and predicting movement of people [1]. Estimating stay point's techniques are basically used distance and time differences with adjustable thresholds as criteria for detection [4, 5]. To extract significant places from the stay points, cluster algorithms are applied such as k-means cluster, density-based cluster and grid-based cluster [4, 5]. For our research, we intend to find the most accurate location at point level rather than region meaning that the accuracy of the centroid is necessary. Hence, high density of point data and distance from centroid are used as criteria for evaluating cluster function. Identifying or labeling types of places is an important task especially for Home and Work places because it is a fundamental knowledge enabling further research steps. In [1, 2, 5], the authors

extracted significant locations using several cluster techniques but did not exactly label types of places in outputs. The work in [3] is able to infer location types using Relative Markov Network (RMN) which is different from our approach. In addition, all previous studies [1–3, 5] conducted the experiments on a dataset collected from GPS devices or mobile phones with labeling activities by volunteers rather than using real-world or inattention human datasets. Those datasets mostly had cleaned location data and high temporal density. The real-world data are not so reliable and usually have errors and missing points [6]. Regarding our experiment dataset, GPS points of users are provided by the mobile operators with user permission which much reflect real human mobility than data conducted by volunteers and also data are still being collected everyday which lead to possible advanced research in larger scale such as people mobility in each city or between cities. In summary, the main differences of our dataset compared with those obtained from other studies are shown in following Table 1.

Table 1. Related work with GPS data

	Source	Data rate	Duration	Users	Ground truth
[1]	User (GPS device)	1/s	4 months	35	All
[3]	User (GPS device)	1/s	1 weeks	5	All
[2]	User (Mobile)	1/min	3 weeks	28	All
[5]	User (Mobile)	>1/min	5 months	8	All
Our	Operator (Mobile)	1/5 min	1 year	1.5 m	0.01%

Cloud computing platform is an option for processing such a large data in a range of terabytes to petabytes with dynamically scalable and virtualized resources [7, 8]. Hadoop is an open source large-scale distributed data processing that is mainly designed to work on commodity hardware [9] meaning it does not require high performance server-type hardware. Hive is a data warehouse running on top of Hadoop to serve data analysis and data query by providing SQL-like language called HiveQL [10, 11]. Hive allows users familiar with SQL language to easily understand and able to query data. In our previous research [12], Hadoop/Hive with spatial enable produced very good results in comparison by reducing time from 24 h of database processing to 1 min. In this study, we developed a location extraction module combining with spatial knowledge on Hadoop/Hive to allow parallel processing and enabling full scale data processing.

3 System Overview

3.1 Mobile Dataset

The dataset was collected from about 1.5 million real mobile phone users in Japan over one-year period, and the data collection is being continued. The total number of records was 9,201 million and about 600 GB in size. Auto-GPS function was activated on a mobile phone to send current location data to server in every 5 min; however, it was

depended on several factors such as no signal and battery preserved function. For example, location sending function would be automatically turned off if there was no movement detected. That resulted in only 37 points collected a day in average. In addition, geo-locations were acquired and calculated from GPS, Wi-Fi and Mobile cell tower. The accuracy of position was defined in three levels range from 1 to 3 (the highest accuracy). With the accuracy defined by operator, 39% of points had accuracy in the 500 m range. Another 23% had 100 m range accuracy and the remaining 38% of points had accuracy in 10 m. Figure 1 depicts the distribution of GPS data and number of points in difference cities (top 10). Density of point cloud was directly reflected with the size of the city. For example, Tokyo, the capital city of Japan, had the highest density of point data. Considering the privacy issues, we used these datasets anonymously.

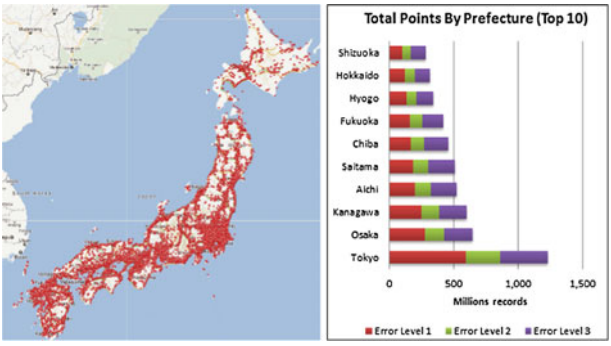


Fig. 1. Data distribution in Japan

3.2 Significant Places Extraction

In order to effectively use inference model as classifier, suitable model, classification features and training data are necessary. We separated tasks into two main parts as shown in following figure. The left-box part is to evaluate and build classifier and another is for labeling location such as Home and Work location. In addition, we used percentage of residential building in grid to indirectly compare results of ground truth and dataset (Fig. 2).

Stay Point Detection and Clustering. Most human activities are routine and people tend to spend time in the same places in their daily life. GPS points represent spatio-temporal location of people defined by $P = (p_1, p_2, \ldots p_n)$ where p = (id, time, lat, lon) and n = a total number of points. Connecting consecutive points of a user in a day according to their time series, user trajectory can be obtained. To extract the signification places particularly home and work places, we first applied stay point extraction algorithm [4, 5] based on the spatial and temporal values of points. In the algorithm, a

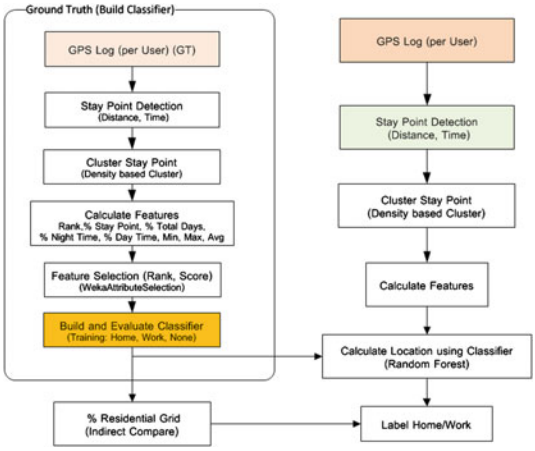


Fig. 2. Significant place extraction

stay point represents a geographic region which a user stays for a while. Space distance and time difference between observed points as shown in following constraints were applied as criteria for detecting stay points.

$$\text{Distance}(p_{start}, p_{end}) < D_{threh} \text{ and } \text{TimeDiff}(p_{start}, p_{end}) > T_{threh}$$

where D_{threh} and T_{threh} are adjustable parameters. D_{threh} is the maximum distance covering a place considered as a stay point. T_{threh} is the minimum time that users spend in the same places. In the experiment, a stay point was detected if $T_{threh} > 20$ min and $D_{threh} <= 150$ m. The Haversine formula was used to calculate the great-circle distance between two points instead of Euclidean distance to increase distance accuracy. The stay point extraction was applied to extract stay points of user in each day and kept as a list of stay points defined by $SP = (sp_1, sp_2, \ldots sp_m)$ where sp = (lat, lon, start-time, end-time) and m = a total number of stay points. The latitude and longitude of a stay point is the centroid of all points in the stay points. Daily stay points of users were combined together and used as an input for evaluating cluster algorithm. We tested several cluster functions such as KMeans, IterativeKMeans, DBSCAN (Density-based spatial clustering of applications with noise), and OPTICS (Ordering Points To Identify the Clustering Structure) to find a suitable algorithm by using high density of points and distance from the centroid as criteria. As shown in Fig. 3, centroids of two KMeans were far from highest density locations and IterativeKMeans was a bit better than KMeans because it automatically found suitable number of clusters using iterative and evaluating clusters, it was still away from dense point groups. Density-based clusters produced good results. Centroid of cluster was pointed to almost center of highest density groups. However, we chose to use DBSCAN clusters for easier implementation in Hadoop/Hive.

Fig. 3. Evaluate cluster algorithms

Classification Features. This section deliberates the classification features used in proposed method including number of stay points, cluster rank, stay time related feature, number of day and night and day time.

Number of stay points (%). From the outputs of clustering, number of stay points of each cluster was computed, and then percentages were calculated comparing between all clusters. This feature was used to classify importance of each cluster since people tend to visit their important places more frequently than others.

Rank in clusters. Apart from number of stay points, we ranked each cluster by number of stay points in clusters. Rank 1 was the highest number of stay points. This feature was used to capture top most portions of cluster results. For example, first 2 top ranks mostly dominated more than 50% of stay points and usually are home or work places.

Stay time (Min, Max, and Average). For each output cluster, we calculated stay time in minute for each stay point and then calculated the minimum stay, the maximum stay and the average stay. These features were used to discriminate short stay and long stay clusters for people who tended to stay home for longer period then other places.

Number of days that points appear (%). Not only number of visit places but number of visit days was also important. For example, people have to stay home almost every day or have to go to work on weekdays. To calculate this feature, we counted unique day from stay points in each cluster and then computed percentage comparing all days from all clusters.

Night and Day time stay (%). By analyzing and classifying time duration of each stay point in the day of the week and 24-h period, home and work places can be possibly derived. We used periods from 22:00 to 06:00 for night time and 08:00–18:00 for day time. Some period was omitted due to the high possibility of being commuting time. Then, percentage was calculated comparing the sum of all values in its own cluster. These were two main features for specific home and work locations.

Hour/ Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sun	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Mon	0	0	1	1	0	0	0	3	4	10	25	28	19	13	17	19	28	32	22	3	0	0	0	0
Tue	1	1	1	0	0	0	0	2	5	12	25	22	18	11	10	18	31	33	28	4	1	0	0	0
Wed	1	1	0	0	0	0	0	0	3	15	28	26	19	11	18	17	22	32	24	4	0	0	0	0
Thu	0	2	1	0	0	0	0	1	4	13	24	25	21	15	18	19	27	36	27	1	0	0	0	0
Fri	0	0	0	0	0	0	0	1	5	11	24	33	25	16	15	14	27	37	33	5	0	0	0	0
Sat	1	1	1	0	0	0	0	1	3	15	25	28	22	14	19	24	34	37	15	1	0	0	0	0

Fig. 4. Weight-matrix of stay points in a cluster

Feature Selection Technique. Information Gain were used and we used three output types consisted of HOME, WORK and NONE. The details about training data used in feature selection including classifiers are discussed in Sect. 4. The ranking and score of all classification features are shown in Table 2. We can see that the four top ranked classification features are the top most important features for the classification.

Table 2. Feature selection (rank, score)

Feature	Rank	Score
Rank in cluster	1	0.60
No. of stay point (%)	2	0.52
Night time (%)	3	0.45
No. of day (%)	4	0.42
Day time (%)	5	0.32
Maximum stay	6	0.27
Average stay	7	0.26
Minimum stay	8	0.13

3.3 Data Processing with Hadoop

Hadoop is an open source cloud computing software framework for data intensive and distributed application. Many projects are developed under Hadoop; however, in this research, we only used HDFS and Hive. The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware meaning it can be deployed on low-cost hardware with fault-tolerant. HDFS delivers high throughput for accessing to data and is suitable for systems that have large scale data sets [9]. Hive is a data warehousing package providing SQL-like language called HiveQL for querying and processing data in Hadoop [10]. It targets users who are familiar and comfortable with SQL to do ad hoc query, summarization and data analysis. Basically, when issuing the query, Hive allows developers to create custom functions that could be any function based on user requirement. There are three types of functions including User-Defined Function (UDF), User-Defined Aggregate Function (UDAF) and User-Defined Table Function (UDTF). With a technique proposed in [12], Hive can be used for spatial processing. The processing consists of three main stages including data loading,

data preprocessing and location extraction. The processing was started from loading raw mobile data files to the Hadoop cluster as a table using Hive with two queries shown in a following figure. One was for creating data table in Hive and another one was for loading data to Hive. Table partitioning was also used for separating data by date. In preprocessing step, the original data format ordered by time and for faster accessing, we converted data to another structure which grouped data with the same user ID together and put them in an array column in each user record. Hence a size of one record was bigger but the total records are reduced equally to a number of users which was about 1.5 million (Figs. 4 and 5).

```
CREATE TABLE zdcdata_raw (uid string, dt string, lat string, lon string, err int, alt double)
PARTITIONED BY (dateonly string)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS TEXTFILE;
```

```
Load data local inpath '/zdcdata/20100801.csv'
overwrite into table zdcdata_raw PARTITION (dateonly='2010-08-01');
```

```
CREATE TABLE zdcdatauid (uid string, dateonly string, total int, arr ARRAY<ARRAY<string>>)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\t'
COLLECTION ITEMS TERMINATED BY ','
MAP KEYS TERMINATED BY '|'
LINES TERMINATED BY '\n'
STORED AS SEQUENCEFILE;
```

```
INSERT OVERWRITE TABLE zdcdatauid
select uid, pdt as dateonly, count(dt) as total,CreateTrajectories(dt,lon,lat,err,alt) as arr
from zdcdata group by uid;
```

```
select uid, StayCluster(ExtractStayPoint(arr,20,150),200,5,'hwtrainingdata.csv')
from zdcdatauid
```

Fig. 5. Data loading queries and scripts for new structure

The figure shows queries used for creating and loading data into a new table. CreateTrajectories is a function made for creating time ordered array data. However, the original data structure was still retained and recoverable by using "Lateral View" function to extract array data to row-based data (Figs. 6, 7 and 8).

uid	interval_list
00001	[1,2,3]
00002	[4,5]

uid	interval_list
00001	1
00001	2
00001	3
00002	4
00002	5

Fig. 6. An example of lateral view

In the extraction stage, we developed two main functions. They were used for extracting stay points from raw log data and extracting cluster from stay points respectively. An example of the query is shown in Fig. 9. Another function is for finding point in polygon named FindGrid. The grid_file is a file with 2 columns in CSV-like but using ":" as separator. First is code and another is polygon data in Well-known text format. The file is loaded to Hadoop distributed caches.

- ExtractStayPoint (array, min_time, max_distance)
- StayCluster(staypoints, epsilon, min_points)
- FindGrid(latitude, longitude, grid_file)

For implementation, we used Java language for development since Hadoop and Hive are native Java application. Java Topology Suite (JTS), which was a java-based spatial library, is used for supporting spatial calculation such as find points in geometry and spatial index for fast searching geometry. For data mining techniques, we used Java Machine Learning Library (Java-ML) for clustering, feature selection and classification. Our Hadoop cluster is depicted in Fig. 7. It consisted of five computers with the same specification that is 8-Cores Xeon 2.6 GHz, 8 GB memory, and two of 2 TB disk to increase I/O performance. The operating system is CentOS 6.0 64-bit. Gigabit switch was used for communication among cluster nodes. One computer ran as NameNode and others as DataNodes and TaskNodes. Hive service was run on the same machine with NameNode. In total, the cluster had 32 cores, 32 GB memory and 16 TB storages. It could have up to 28 tasks running at the same time (4 * 7 Cores). The version of Hadoop was 0.20.2, and the version of Hive was 0.8.0 (Fig. 10).

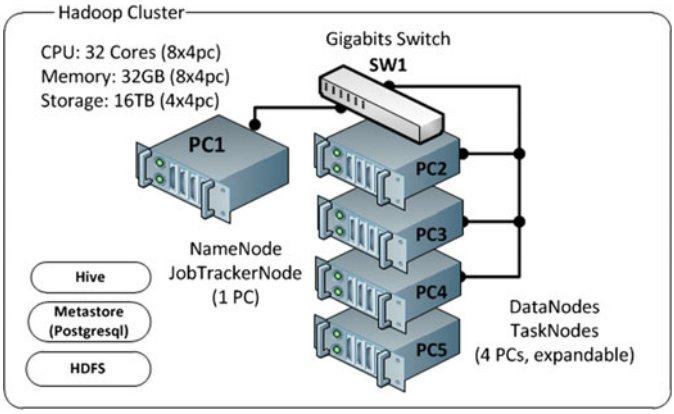


Fig. 7. Testing systems for Hadoop cluster

4 Evaluation

4.1 Stay Point and Cluster

We processed GPS trajectories for extracting stay points and then cluster stay points. We selected trajectories of one user to show the results. The selected user had 32,237 points in a year. In total, 1,140 stay points were extracted from the trajectories with time threshold of 20 min and distance threshold of 150 meters. For DBSCAN cluster, we used epsilon = 200 m with 0.05% of all stay points as minimum points. The results are shown in Fig. 11 and Table 3. Two clusters including Cluster 1 and Cluster 2 contained a number of stay points significantly more than other clusters. It means that the user stayed in these two areas more frequently than other areas. Hence, there was a high possibility that these two areas were home and work place. By classifying stay points using temporal features in each cluster in Table 1, the result of Cluster 1 and 2 are shown in Fig. 8.

Table 3. sample result of stay points clustering

Cluster	Lat, lon	Point count	% night time	Other features
Cluster 1	35.65157, 139.794052	584	2.25	…
Cluster 2	35.731584, 139.824243	214	74.71	…
Cluster 3	35.719668, 139.830936	12	19.23	…

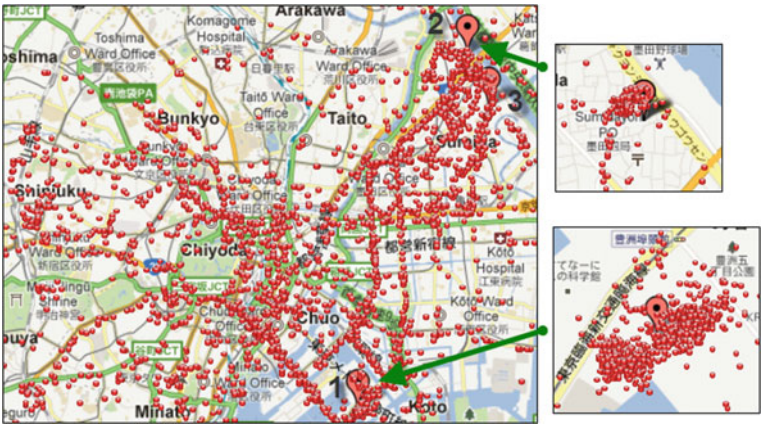


Fig. 8. The location of cluster result

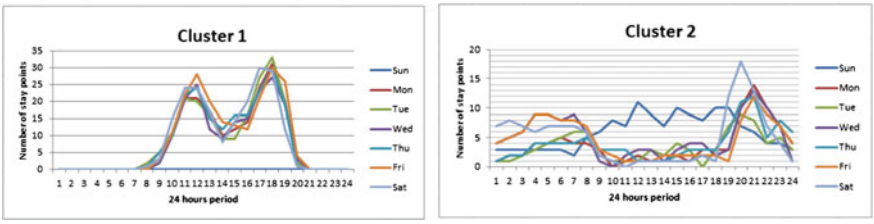


Fig. 9. Stay points in time periods of cluster 1 and 2

The graph of Cluster 2 clearly shows that the user tended to stay in this location during the day time from 9:00 to 19:00 and from Monday to Saturday. This activity was likely to be working time of usual people. Hence, Cluster 2 could be derived as a work place of the user. On the other hand, Cluster 1 graph shows that the user tended to stay at night time from 19:00 to 8:00 on every day except Sunday that the user stayed in this area all time. Therefore, this cluster could be home of the user because people are usually stay home at night time. Moreover, we could also derive the working period of the user. For this one, the user might work from Monday through Saturday and start working from 9:00 to 18:00.

4.2 Training Data

Out of all datasets with 1.5 million users, we had ground truth information of only 139 users including locations, transportation modes and activities. For ground truth, we extracted stay points and clustered all stay points as described previously, and then calculated all classification features stated previously. After that, we manually labeled locations of all cluster results including three output values: HOME, WORK and NONE. The example of training data is shown in Fig. 10.

RANK	Min	Max	Avg	Total Day	Total Point	Night Time	Day Time	Value
1	31.00	486.00	305.43	100.00	50.66	84.31	1.57	HOME
2	29.00	970.00	320.46	80.00	32.60	7.56	60.47	WORK
3	25.00	245.00	106.57	40.00	11.01	0.00	86.44	NONE
4	164.00	279.00	191.88	20.00	5.73	0.00	92.11	NONE

Fig. 10. Training data

4.3 Location Classifiers

To determine the accurate classifiers for inferring location types especially home and work places, we compared accuracy of 3 different classification models consisted of Naïve Bayes, K-Nearest Neighbors and Random Forest. From the training data, 70% was used for building classifier and 30% (128) was used for verification. Table 4 shows the results of all classifiers. The results indicate that Random Forest was the best among

evaluated models with an accuracy rate of 99.22%. We used Random Forest as a classifier for implementing in Hadoop/Hive.

Table 4. Classifier results

Classifier	Correct	Wrong	Accuracy (%)
K-Nearest neighbors	105	23	82.03
Naïve Bayes	72	56	56.25
Random forest	127	1	99.22

4.4 Performance

We compared processing time of a single process on a computer and Hadoop/Hive with 4 parallel nodes. For Hadoop/Hive, we had to preprocess data so that it could be used for extracting location efficiently. The preprocessing steps consisted of importing data, converting data structure to user-based schema. Processing time for preprocessing tasks with some comparisons is shown in Table 5. Originally, data files came in daily format such as day1, day2 and so on. In total, there were 365 files. We converted to user-based which was one file for a user resulting in about 1.5 million files. With only preparation time, Hadoop/Hive showed significant improvement compared with other methods. Hadoop/Hive took less than 2 days for processing the whole dataset compared with Java application run on a computer which took more than a month. This shows that Hadoop/Hive is much faster, especially when processing large dataset because all data can be processed in parallel rather than bottleneck in one computer. Also, we can increase performance easily by increasing number of cluster nodes.

Table 5. Processing time of preprocessing tasks

Method	Task	Time
Database (PostgreSQL)	Import data, create geometry, create spatial index	50 days (Estimated[a])
Java application	Convert from daily data file to user data	3 days
Hadoop/Hive (4 nodes)	Import data, convert data structure	14 h

[a]For database, one day data (20 million records) took 3.3 h

4.5 Spatial Population Census

We have attempted using spatial population census data which provided a real number of population over Japan with a detail down to grid level. Focusing on the results of home location, we counted number of home in each grid cell and compared them with population census at same grid level to find correlation. The result showed a good correlation between these two data with $R^2 = 0.8643$ as depicted in Fig. 10. Somehow, we could possibly use the extraction results for finding real population and this also expressed the benefit of processing large scale data over the small dataset.

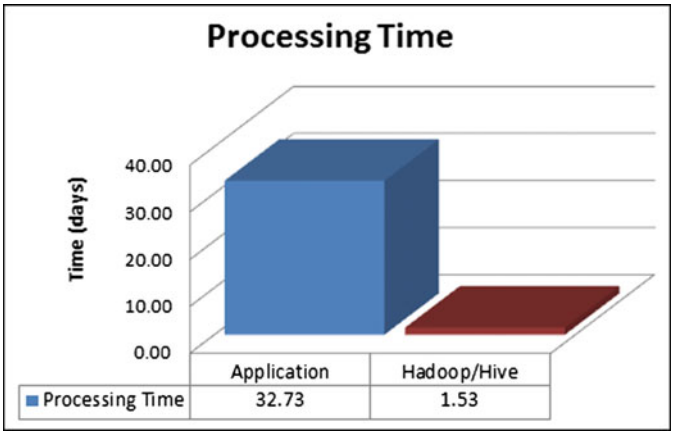


Fig. 11. Processing time comparison

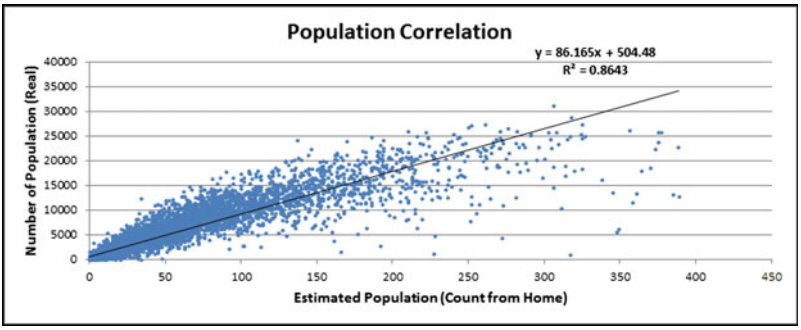


Fig. 12. Population correlation

5 Conclusion and Future Work

In this paper, we conducted the experiment on a large scale mobile phone dataset collected from a mobile operator side with aiming to extract meaningful places particularly home and work places of people from their daily trajectories. Stay point finding algorithm and DBSCAN cluster were employed to extract significant places. We also evaluated eight features including number of stay points, cluster rank, stay time related feature (min, max, avg), number of days, night and day time using feature selection algorithms in order to use as classification features. With all features, we considered three classifiers to infer home and work locations. Random Forest obtained the best accuracy at 99.22%. Using Hadoop/Hive, we showed that our technique is able to reduce processing time on the whole dataset from 32.73 days to 1.53 days with only four computation machines. Even though the system used 32 cores in total, the speed

up is not linear due to several factors such as disk I/O and number of tasks that if more than 32 tasks, all tasks cannot be run at the same time and will be run after some task finished. Regarding scalability, it is possible to add additional nodes without any changes in software development to increase performance both processing time and storage size. With limited number of ground truth data, we have compared home location results with real population census and it resulted in quite good correlation with $R^2 = 0.8643$. These results confirmed that the real large scale mobile phone dataset is a novel source for mining human behavior and even though there is very sparse in daily data, it is still possible for mining useful information. In the future, with the promising results so far, we have planned to process extraction of transportation modes and then combine results with significant places to create daily trip of each users. Finally, we will use all results to discover daily life patterns of people (Fig. 12).

In the future, with the promising results so far, we have planned to process extraction of transportation modes and then combine results with significant places to create daily trip of each users. Finally, we will use all results to discover daily life patterns of people.

Acknowledgements. The work described in this paper was conducted at Shibasaki Laboratory with an agreement from Zenrin Data Com to use mobile phone dataset of personal navigation service users for the research. This work was supported by GRENE (Environmental Information) project of MEXT (Ministry of Education, Culture, Sports, Science and Technology).

References

1. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. Pers. Ubiquitous Comput. **7**, 275–286 (2003). https://doi.org/10.1007/s00779-003-0240-0
2. Zhou, C., Frankowski, D., Ludford, P., et al.: Discovering personally meaningful places. ACM Trans. Inf. Syst. **25**, 12–es (2007). https://doi.org/10.1145/1247715.1247718
3. Liao, L., Patterson, D.J., Fox, D., Kautz, H.: Building personal maps from GPS data. Annal. New York Acad. Sci. **1093**, 249–65 (2006). https://doi.org/10.1196/annals.1382.017
4. Zheng, Y, et al.: Mining interesting location and travel sequences from GPS trajectories. In: Proceedings of WWW, Madrid, Spain, pp. 791–800. ACM Press (2009)
5. Montoliu, R., Blom, J., Gatica-Perez, D.: Discovering places of interest in everyday life from smartphone data. Multimed. Tools Appl. (2012). https://doi.org/10.1007/s11042-011-0982-z
6. Kirmse, A., Bellver, P., Shuma, J.: Extracting patterns from location history. In: Proceedings of ACM SIGSPATIAL, Chicago, USA, pp. 397–400 (2011)
7. Jinnan, Y., Sheng, W.: Studies on application of cloud computing techniques in GIS. In: Proceedings of IGASS 2010, China, pp. 492–495 (2010)
8. Buyya, R., Yeo, C., Venugopal, S., et al.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Future Gener. Comput. Syst. **25**, 599–616 (2009). https://doi.org/10.1016/j.future.2008.12.001
9. Hadoop Project: http://hadoop.apache.org/
10. Hive Project: http://hive.apache.org/
11. Thusoo, A., Sarma, J., Jain, N., et al.: Hive—a petabyte scale data warehouse using hadoop. In: Proceedings of ICDE 2010, pp. 996–1005 (2010)

12. Witayangkurn, A., Horanont, T., Shibasaki, R.: Performance comparisons of spatial data processing techniques for a large scale mobile phone dataset. In: Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications—COMGeo'12 1 (2012). https://doi.org/10.1145/2345316.2345346

Modeling Traffic Accidents Occurrences Based on Land Use and Road Factors Using Geographically Weighted Regression Models

Paweenuch Songpatanasilp[1(✉)], Teerayut Horanont[2], Harutoshi Yamada[3], and Ryosuke Shibasaki[3]

[1] Department of Civil Engineering, The University of Tokyo, Tokyo, Japan
jyajaz@gmail.com
[2] Sirindhorn International Institute of Technology, Thammasat University, Bangkok, Thailand
teerayut.horanont@gmail.com
[3] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan
{yamada.hal, shiba}@csis.u-tokyo.ac.jp

Abstract. Many researches regarding road traffic safety have been trying to identify the factors, which have significant effects on the probability of traffic accidents occurrences, and factors such as climate and road related factors have been proven to have substantial effects on the issue. On the contrary, in Japan, land use factors have not been mentioned by many of the researches because the positional information of traffic accidents was unavailable until recently; and, therefore, their momentous effect on traffic accidents occurrences unrevealed. In this paper, road related factors, such as traffic volume, lengths of various road types and number of intersections have been ana- lysed along with land use factors aggregated based on 1 × 1 km grid mesh of Tokyo to investigate the influence of these factors on traffic accidents occurrences. This study could shed the light for the future policies for traffic safety aiming at reducing collisions and improving traffic safety.

Keywords: Data mining · Modeling · GWPR · Land use · Road Traffic accident

1 Introduction

As science and technology have become more and more instrumental, people have moved towards modernity and expediency. As a result, we have witnessed a rise in number of motor vehicles used worldwide. According to the World Health Organization [1], from 2007 to 2010, the number of registered vehicles has seen a fifteen percent increase, directly impacting road traffic injuries and deaths. Furthermore, the rapid growth of modernized cities is negatively affecting safety level of not only non-motorized road users, notably cyclists and pedestrians, but also motorized road users. In 2010, 50% of road traffic deaths occurred among vulnerable road users. In particular, 27% of fatalities were non-motorized road users, 22% were pedestrians, and

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_18

5% were cyclists, with the remaining 23% were reported to occur amongst motorcyclists [1]. The 2013 global road safety status report indicated that the number of total road traffic deaths reached an unacceptably high total of 1.24 million. This ranked road traffics deaths as the eighth leading cause of death, and the number one cause of deaths amongst those aged 15–29 years [2] causing the global loss due to road traffic injuries as high as 518 billion US Dollars, and costs the world's governments between 1 to 4% of their gross national products (Integrated Conference of Better Air Quality; BAQ and Intergovernmental Eighth Regional Environmentally Sustainable Transport; EST, 2014).

As the impact of traffic accidents, both short term and long term, have been acknowledged by the world, there are many research results and papers regarding road safety trying to identify the elements, which have significant effects on the occurrence of vehicle collisions. The well-known factors, which have been proven to have substantial effects on traffic accidents, are weather conditions, road conditions and other road related factors, drivers' characteristics, etc. However, land use factors were the lacking element that has not been mentioned in many of the papers, leaving the potential of momentous effects on road accident occurrence undiscovered. Thus, the main objective of this research is to identify factors that have significant effects on the traffic accident occurrence.

2 Literature Review

In recent years, road safety has attracted more interests from people around the world. Many researchers and papers have gone into the direction of determining causes and improving road safety. Despite the interests and efforts, what is still lacking is a whole-encompassing and detailed data. It is not easy to obtain such data having potential factors; and, hence, most studies were conducted by small size samples or survey data with a few numbers of factors.

In addition, there have not been many of papers that investigated the effects of land use on traffic accidents occurrences while most of the interests were diverted to other factors, such as weather conditions, temporal factors, driver's characteristics, and road factors. Exceptionally, Dissanayake et al [3] investigated the suitability of using land use variables to predict the number of child pedestrian casualties with Generalized Linear Models (GLMs) in Newcastle upon Tyne, Great Britain. The results show secondary retail and high-density residential land use types are associated with all child pedestrian casualties. Furthermore, educational sites, junction density, primary retail, and low-density residential land use types are associated with child casualties at different time periods of the day and week.

Pahukula et al. [4] examined the correlations of temporal factors and road factors by separating crashes occurrence into five time periods, namely, early morning, morning, mid-day, afternoon, and evening. Their study was conducted in Texas, United States and showed that among many road-related variables, traffic flow, light conditions, surface conditions, time of the year and percentage of trucks on the road were found to be key differences among the time periods.

Furthermore, the combined use of land use and road-related factors in predicting traffic accident occurrances were conducted in Tokyo, Japan, using GLMs and

Zero-inflated Models. The result of Songpatanasilp et al. [5] showed that accidents occur more frequently in commercial areas than in other land use areas, and that accidents occur least frequently in residential areas. In addition, accidents occur more frequently around shops, hospitals, sport facilities and attractions on workdays. On holidays, accidents occur more frequently around shops and sport facilities. Regarding the models, Zero-inflated Negative Binomial Model showed great ability in capturing traffic accident characteristics in the data with excessive zero count.

One of the most widely used model for traffic accident analysis are Generalized Linear Models (GLMs), however, by using GLMs the spatial relationships of traffic accident occurrences and the risk factors are disregarded. Thus, there have been studies investigating the spatial relationship of count events using other models. For example, Carrel et al. [6] have investigated the impact of tubewell user density on cholera and shigellosis events in Bangladesh using a zero-inflated model and a geographically weighted model (GWM). It was found that the tubewell density effects are spatially stationary and the use of non-spatial statistical methods are recommended for that study.

On the other hand, Aguero-Valverde and Jovanis [7] found the existence of spatial correlation in the injury crashes analysis in Pennsylvania using full Bayes hierarchical models. However, no evidence of spatial correlation was found in the fatal crashes analysis of the same area.

In dealing with count data with overdispersion, geographically weighted negative binomial regression (GWNBR) has been proposed by da Silva et al. [8]. Spatial correlation is one of the causes of overdispersion. It is, therefore, necessary to investigate whether spatial correlation exists in Japanese traffic accidents data. This is one of our future research themes.

3 Data Description

3.1 Traffic Accident Data

The analysis in this paper is based on 1 × 1 km grid meshes that cover Tokyo city. All the data is aggregated mesh by mesh to create mesh based variables including two objective variables; one is the number of accidents on weekdays, and the other one is the number of accidents on weekends and holidays of each mesh occurred in 2013.

Road accidents data used in this study were collected by the National Police Agency of Japan in 2013. The data contain information such as the location (that is, longitude and latitude) and the time when accidents occurred, the gender and age of the primarily responsible party and the severity level of accidents.

3.2 Land Related Data

The land-related data in this study consists of four kinds of data, namely *Urban Map* provided by the Ministry of Land, Infrastructure, Transport and Tourism indicating urbanized area in Tokyo and has been aggregated based on 1 × 1 km grid meshes

using QGIS to determine whether each mesh belongs to an urbanized area (category 1), a non-urbanized area (category 0) or a partly-urbanized area (category 1).

Densely Inhabited Districts (DIDs) defined by the Ministry of Internal Affairs and Communications and has been determined using the same criteria as was used for the urbanized area.

Land Use prescribed in urban areas by local governments in Japan comprised of thirteen categories. Some of the 13 land use categories are analogous with each other and some of them have small areas. These facts cause significantly low number of meshes, occasionally zero, in some categories. Therefore, the categories which are akin with each other are integrated leaving eight land use categories remained.

Points Of Interest (POI) provided by ZENRIN cooperation, which indicate the purpose of buildings, constructions, and lands in Japan along with their location. The buildings with ten types of purpose, which are chosen for the study to investigate the relationship with traffic accidents, are banks and financial business, shops and convenience stores, department stores, entertainment places such as a theater, a museum, a café and a restaurant and others, hospitals, elementary schools, high schools and universities, sport facilities, railway stations, and tourist attractions such as a theme park, zoo etc.

3.3 Road Traffic Census 2010

The Road Traffic Census is conducted by road operators every five years in autumn when the traffic situation is considered to represent an average one.

Meshes with no road exist are excluded from this study since traffic accident cannot happen in those area. The road related variables of each mesh derived from the census data are number of intersections, length of national roads, arterial roads, municipal and prefectural road, and thin roads, average traffic volume and speed.

4 Methodology

4.1 Data Preparation

The accident data were divided into two sets; one is those accidents occurred in weekdays and the other one is those occurred in weekends and holidays. The traffic accidents data, land use data and road characteristics were imported to ArcGIS along with the 1 km by 1 km mesh data of Tokyo in order to calculate the values explanatory variables mesh by mesh. The basic statistics of objective and explanatory variables are summarized in Table 1. The candidate explanatory variables will be selected using an R function for model selection in the model calibration process.

4.2 Geographically Weighted Poisson Regression Model

The coefficients of Generalized Linear (GLM) Models are assumed to be constant over space. In this sense, they are global models. However, traffic accidents may have spatial correlation. For example, the auto usage of people might vary over space owing to

Table 1. Descriptive statistics of variables

Variable	Min	Mean	Max	S.D.
Objective variables				
Number of accident on workdays	0	15.29	127	16.73
Number of accident on holidays	0	5.83	58	6.51
Land use elements				
Number of elementary schools	0	0.83	5	0.94
Number of banks	0	1.75	98	5.86
Number of shops	0	80.29	2080	142.94
Number of department stores	0	6.44	324	18.96
Number of entertainment places	0	52.38	2257	147.13
Number of hospitals	0	0.28	28	1.15
Number of schools and universities	0	15.46	331	26.54
Number of sport facilities	0	2.49	54	4.63
Number of stations	0	0.40	28	1.43
Number of attractions	0	6.88	215	16.66
Road related variable				
Length of national roads (m)	0.0	457.7	7018.2	1028.76
Length of arterial roads (m)	0.0	1021.4	9072.1	1349.58
Length of municipal roads (m)	0.0	7118.0	36981.0	5439.96
Length of thin roads (m)	0.0	11656.0	39178.0	8964.44
Number of intersections	0	462.5	2140	370.21
Avg. traffic volume (veh/road/day)	0.0	138.9	688.6	117.24
Avg. traffic speed (km/h)	0.0	22.96	82.00	14.10
Categorical variables	# *of meshes*			
Land use				
Commercial	69			
Industrial	29			
Quasi-industrial	186			
Exclusive industrial	21			
Lower rise residential	710			
Middle rise residential	199			
Residential	112			
Mountainous and others	333			
Densely inhabited district (DID)				
DID	938			
Partly DID	289			
Non DID	432			
Urban area				
Urbanized	850			
Partly urbanized	438			
Non urbanized	371			

various factors and this variation could have influence on the occurrence of accidents. It is, however, difficult to identify this variation by one explanatory variable.

There are some models already developed to express spatial variation. Multi-level models or random effects models are able to consider spatial variation. However, these models 'aim to capture variances of parameter distributions rather than revealing their geographical patterns' [9].

In this paper geographically weighted Poisson regression (GWPR) models are selected to capture the spatial variation of traffic accidents. GWPR models can not only capture spatial correlations but estimate regression coefficient specific to spatial locations (that is, specific to each mesh); and it is expected that GWPR models produce a better estimate of a coefficient [10].

Let *Poisson(μ)* represent the Poisson distribution with mean μ. In the framework of GLM, *log(μ)* is expressed by a linear combination of parameters β_i. In GWPR models, the values of β_i are assumed to vary depending on the spatial location (u_i, v_i). The values of parameters in location (u_i, v_i) are estimated by maximizing the log-likelihood defined by the Eq. (1):

$$\max L(u_i, v_j) = \sum_{j=1}^{N} \left(-\hat{\mu}_j(\beta(u_i, v_i)) + y_j log\hat{\mu}_j(\beta(u_i, v_i))\right) w_{ij}\left(\left|u_i - u_j\right|\right) \tag{1}$$

where $\hat{\mu}_j(\beta(u_i, v_i))$ is the mean of the Poisson distribution at the location (u_i, v_i), y_j is the counts at the location j and w_{ij} is the weight for the location (u_i, v_i) used for the parameter estimation at the location (u_i, v_i). The values of weights w_{ij} are positive and their values become smaller as the distance from the position i gets larger.

For the mentioned purpose, the model development in this study were conducted using geographically weighted Poisson regression techniques that use log-link function of Eq. (2).

$$\ln(\mu_i) = \ln(\beta_0(u_i, v_i)) + \sum_k \beta_k(u_i, v_i) x_{ik} + \epsilon_i \tag{2}$$

where μ_i is the expected number of accidents at location i; in this study, the number of traffic accidents in weekday or those in weekend and holiday; x_{ik} is the kth explanatory variable. In this study, urban area, densely inhabited district, land use categories, number of banks, shops, department stores, entertainment places, hospitals, schools and universities, sport facilities, train stations, and tourist attractions, length of roads, number of intersections, average traffic volume, total traffic volume, and average traffic speed are used as explanatory variables; $\beta_k(u_i, v_i)x_{ik}$ is the unknown parameter or regression coefficient of the kth explanatory variable in the geographic coordinate (u_i, v_i); and ϵ_i is random error at location i.

As a geographically weight function, several kind of functions have been proposed. In this study, the Gaussian function, shown in Eq. (3) was chosen due to its continuity and applicability. Since the data are weighted geographically, it should be noted that the nearer observations have the more effects on the dependent variables than the ones

further away. Considering the variables used in this study, the equation can be rewritten as presented in Eq. (4).

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \tag{3}$$

where w_{ij} is the jth element of the diagonal of geographical weights $W(u_i, v_i)$ matrix; d_{ij} is the distance between observation i and j; b is the bandwidth determined by minimizing model fit diagnostic such as Akaike Information Criterion (AIC) in Eq. (5).

$$\begin{aligned}\ln(\mu_i) = {} & \ln(\beta_0(u_i, v_i)) + \beta_1(u_i, v_i)\mathrm{ur} + \beta_0(u_i, v_i)\mathrm{did} + \beta_0(u_i, v_i)\mathrm{lu} + \beta_0(u_i, v_i)\mathrm{b} \\ & + \beta_0(u_i, v_i)\mathrm{sh} + \beta_0(u_i, v_i)\mathrm{ds} + \beta_0(u_i, v_i)\mathrm{ep} + \beta_0(u_i, v_i)\mathrm{h} \\ & + \beta_0(u_i, v_i)\mathrm{su} + \beta_0(u_i, v_i)\mathrm{sf} + \beta_0(u_i, v_i)\mathrm{ts} + \beta_0(u_i, v_i)\mathrm{ta} \\ & + \beta_0(u_i, v_i)\mathrm{rln} + \beta_0(u_i, v_i)\mathrm{rla} + \beta_0(u_i, v_i)\mathrm{rlpm} + \beta_0(u_i, v_i)\mathrm{rlt} \\ & + \beta_0(u_i, v_i)\mathrm{int} + \beta_0(u_i, v_i)\mathrm{avgtv} + \beta_0(u_i, v_i)\mathrm{tottv} \\ & + \beta_0(u_i, v_i)\mathrm{avgts}\end{aligned} \tag{4}$$

$$AIC_c(b) = 2n\ \ln(\hat{\sigma}) + n\ \ln(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\} \tag{5}$$

where n is the sample size; $\hat{\sigma}$ is the estimated standard deviation of the error term; and tr (S) is the trace of the projected matrix S from the observed to fitted value.

5 Model Calibration

The traffic accident analysis in this chapter uses the same variables as described in the previous chapter. Unlike the previous chapter, however, the model calibration of geographically weighted Poisson regression models uses the R function 'model.selection.gwr()' for the selection of variables. This procedure start with calibrating all possible bivariate (that is, a constant term and one explanatory variable) GW regressions by sequentially regressing a single independent variable against the dependent variable; and, then, it will find out the best performing model with the lowest AICc and include the corresponding independent variable permanently one by one in subsequent models. This approach can be conducted in a rudimentary form, where the bandwidth of the weighting function is given by user beforehand and remains the same for each GW model regression fit. In this case, all variables are aggregated based on 1 × 1 km grid meshes and the bandwidth was chosen to be greater than 1000 meters at the beginning and ended up with 1600 m. Figures 1 and 2 show graphically the process of model calibration using model.selection.gwr() function for the traffic accidents on workdays model and for the traffic accidents on weekends and holidays model, respectively.

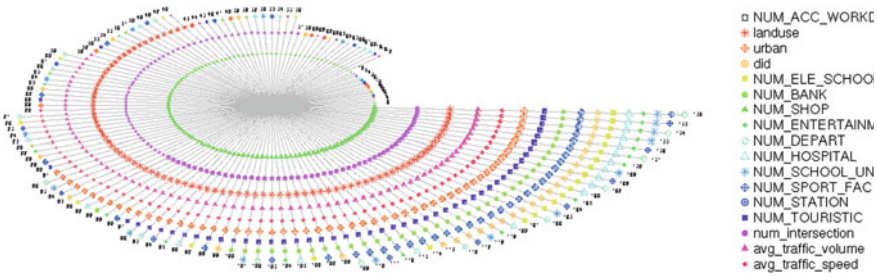


Fig. 1. GWPR model calibration process for traffic accidents on workdays

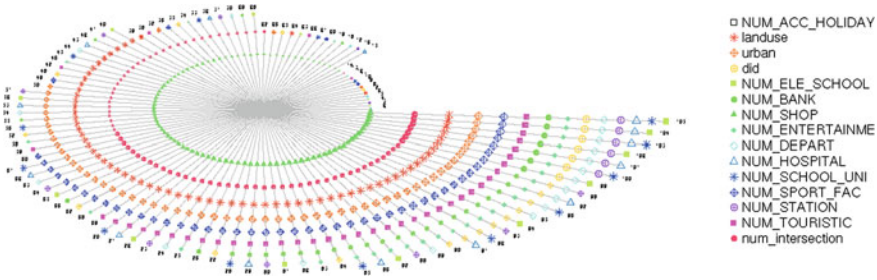


Fig. 2. GWPR model calibration process for traffic accidents on weekends and holidays

As shown in Fig. 1, the first variables that has been permanently included in the model is number of shops, which is located in the inmost ellipse, and is followed by number of intersections, land use, average traffic volume, average traffic speed, urban area, number of attractions, number of banks, number of sport facilities, DID, number of elementary schools, number of hospitals, number of entertainment places, number of high schools and universities, number of railway stations, and finally number of department stores.

While in Fig. 2 showing the calibration process of the weekends and holidays model, number of shops appears as the first permanently included variables in the model and is followed by number of intersections, land use, urban area, number of sport facilities, number of attractions, number of banks, number of entertainment places, DID, number of department stores, number of railway stations, number of hospitals, number of high schools and universities, and finally number of elementary schools.

The two models have similar order of permanently included variables especially in the first half. It must be noted that average traffic volume and average traffic speed are not included in the weekends and holidays model since these data were collected only on weekdays and, thus, have not included in the weekends and holidays model. In this model, number of high schools and universities and number of elementary schools have been included lastly among the variables; this is different from the workdays model. This indicates that the students and pupils commuting to schools have effect on

the traffic accidents occurrences. In particular, elementary school pupils who go to school on foot seem to have a greater effect.

On the other hand, in the workdays model, number of department stores is included lastly among the variables. This shows that most people prefer to go to department stores in the weekends and, hence, the effect of number of department stores is not influential in the workdays model.

This model calibration has been conducted based on the AICc value of each GW regression model. As more variables are permanently included in the model, the lower the AICc value becomes.

6 Model Estimates and Discussion

In geographically weighted regression models, the coefficient of each variable varies from mesh to mesh. Tables 2 and 3 show the minimum, the first quantile, the median, the third quantile and the maximum value of the estimated coefficients of each variable included in the workdays model and weekends and holidays model, respectively. The order of explanatory variables is the same as that of each variable is included permanently in the model. Most of the estimated coefficients are positive such as number of shops, intersections, attractions, sport facilities, hospitals, and traffic volume. This indicates that the increase of the number of these POIs will result in more traffic accidents occurrences, while the negative sign of the estimated coefficient of number of railway stations indicates that the increase of the number of railway stations in a mesh will result in a decrease of the number of traffic accidents.

The spatial variations of the estimated coefficients which were included in the model first in the workdays model and the weekends and holidays models are shown in Fig. 3. Those variables are common to both models and is number of shops.

Table 2. Estimated coefficients of traffic accidents models on workdays

	Min.	Median	Max.
Intercept	−0.4283	0.1888	0.4977
Number of shops	0.0004	0.0005	0.0005
Number of Intersection	0.0005	0.0006	0.0008
Land use			
Industrial area	−0.3312	−0.3076	−0.1917
Quasi industrial area	−0.9403	−0.6123	−0.4530
Exclusively industrial area	−0.5248	−0.5155	−0.4857
Low rise residential area	−0.3660	−0.3566	−0.3445
Middle rise residential area	−0.9967	−0.1210	−0.0252
Avg. traffic volume	0.0021	0.0023	0.0024
Avg. traffic speed	−0.0193	−0.0168	−0.0094

(*continued*)

Table 2. (*continued*)

	Min.	Median	Max.
Urban			
Urbanized area	1.4480	1.7310	1.9320
Partly urbanized area	1.1700	1.4490	1.7270
Number of attractions	−0.0003	0.0011	0.0015
Number of banks	−0.0051	−0.0005	0.0009
Number of sport facilities	0.0078	0.0097	0.0112
DID			
DID	1.0680	1.0960	1.2560
Partly DID	0.7667	0.8087	0.9732
Number of elementary schools	−0.0452	−0.0291	−0.0102
Number of hospitals	0.0095	0.0117	0.0166
Number of entertainment places	−0.0001	−0.0001	0.0003
Number of schools and universities	−0.0004	−0.0002	0.0001
Number of railway stations	−0.0373	−0.0118	−0.0048
Number of department stores	−0.0003	−0.0001	0.0004
		Effective degrees of freedom: 1633.95	
		AICc: 12035.25	
		Residual sum of squares: 138225.2	
		R-square value: 0.7021398	
		Adjusted R-square value: 0.6975705	

7 Conclusions and Future Work

The estimated geographically weighted Poisson regression (GWPR) models show number of shops, number of intersections, and land use variables are highly significant to the traffic accidents occurrences, while number of department stores and number of elementary and high schools and number of universities are appeared to be the least significant variables for the workdays model and the weekends and holidays model. However, the coefficient estimates did not change drastically, which shows the insignificant of the spatial relationship in this data. This could be due to the annual data used in this study, which the seasonal, monthly, or daily variation is not reflected.

The GWPR models work well in an urbanized prefecture such as Tokyo, however, in rural areas where numbers of meshes with no traffic accidents are excessively high, the number of traffic accidents cannot be successfully estimated using a GWPR model.

Table 3. Estimated coefficients of the traffic accidents model on weekends and holidays

	Min.	Median	Max.
Intercept	−1.0700	−0.7927	−0.5180
Number of shops	0.0003	0.0004	0.0008
Number of Intersection	0.0006	0.0007	0.0009
Landuse			
Industrial area	−0.3687	−0.3197	−0.3008
Quasi industrial area	−0.9006	−0.5847	−0.4448
Exclusively industrial area	−0.6101	−0.5157	−0.4544
Low rise residential area	-0.4734	−0.3166	−0.2738
Middle rise residential area	−1.2140	−0.2611	−0.0232
Urban			
Urbanized area	1.2620	1.5110	1.6940
Partly urbanized area	0.9440	1.2090	1.4140
Number of sport facilities	0.0141	0.0155	0.0177
Number of attractions	−0.0004	0.0005	0.0009
Number of banks	−0.0094	−0.0057	−0.0039
Number of entertainment places	−0.0001	0.0000	0.0002
DID			
DID	1.1720	1.2970	1.3090
Partly DID	0.8372	0.9117	0.9539
Number of department stores	0.0002	0.0009	0.0013
Number of railway stations	−0.0247	−0.0128	−0.0095
Number of hospitals	0.0032	0.0068	0.0125
Number of schools and universities	−0.0005	−0.0002	0.0000
Number of elementary schools	−0.0121	0.0025	0.0075
		Effective degrees of freedom: 1632.054	
		AICc: 8136.859	
		Residual sum of squares: 29368.27	
		R-square value: 0.5820062	
		Adjusted R-square value: 0.5751006	

For the future work, it is necessary to investigate temporal change of the traffic accident related variables. Daily or hourly variation of traffic volume or climate is required in order to realize timely evaluation of the traffic accident risk. Also, according to the excessive number of zero count in this data, the model in which zero count data and count data are considered separately such as Zero-inflated Models are highly recommended.

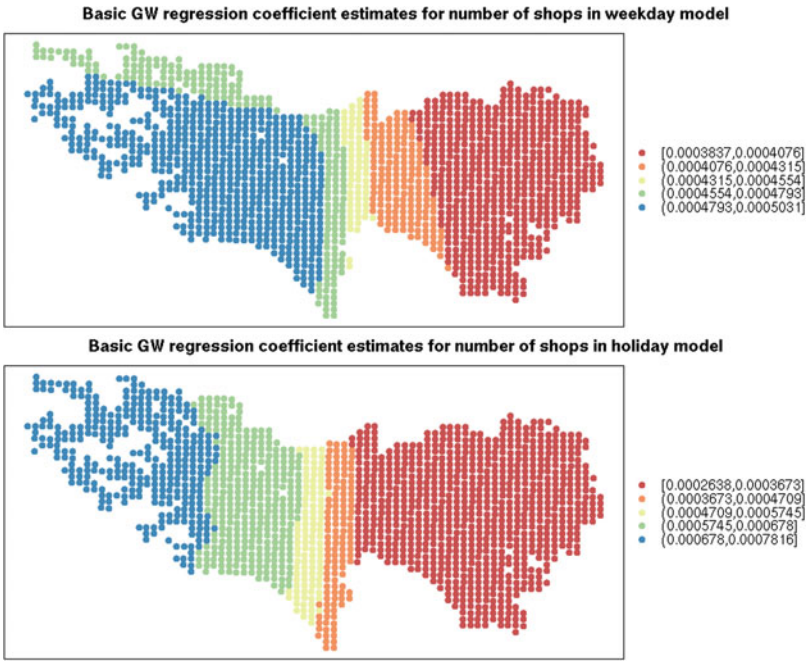


Fig. 3. Estimated coefficient variation of number of shops

Acknowledgements. This paper was supported in part by JSPS KAKENHI grant No.26540166 and SIIT Young Researcher Grant under contract No.SIIT-2014-YRG2. Special thanks for the provision of traffic accidents data to the Institute for Traffic Accident Research and Data Analysis.

References

1. WHO: Global Status Report on Road Safety 2010 (2010)
2. WHO: Global Status Report on Road Safety 2013 (2013)
3. Dissanayake, D., Aryaija, J., Wedagama, D.M.P.: Modelling the effects of land use and temporal factors on child pedestrian casualties. Accid. Anal. Prev. **41**(5), 1016–1024 (2009)
4. Pahukula, J., Hernandez, S., Unnikrishnan, A.: A time of day analysis of crashes involving large trucks in urban areas. Accid. Anal. Prev. **75C**, 155–163 (2015)
5. Songpatanasilp, P., Yamada, H., Horanont, T., Shibasaki, R.: Traffic accidents risk analysis based on road and land use factors using GLMs and zero-inflated models. In: Proceedings of 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM 2015), 7–10 July 2015, Massachusetts, USA, pp. 320-1–320-26 (2015)
6. Carrel, M., Escamilla, V., Messina, J., Giebultowicz, S., Winston, J., Yunus, M., Streatfield, P., Emch, M. (2011). Diarrheal disease risk in rural Bangladesh decreases as tubewell density increases: a zero-inflated and geographically weighted analysis. Int. J. Health Geograph. **10**(1), 41

7. Aguero-Valverde, J., Jovanis, P.P.: Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. **38**(3), 618–625 (2006)
8. Da Silva, A.R., Rodrigues, T.C.V.: Geographically weighted negative binomial regression—incorporating overdispersion. Stat. Comput., 769–783 (2013)
9. Nakaya, T., Fotheringham, A S., Brunsdon, C., Charlton, M.: Geographically weighted poisson regression for disease association mapping. Stat. Med. **24**(17), 2695–717 (2005)
10. Matkan, A.A.: Explorative spatial analysis of traffic accidents using gwpr model for urban safety planning, pp. 1–16 (2011)
11. Note, F.C., Lanka, S.: Next generation solutions for clean air and sustainable transport—towards a livable society. In: Integrated Conference of Better Air Quality (BAQ) 2014 Intergovernmental Eighth Regional Environmentally Sustainable Transport (EST) Forum in Asia Theme, (2010) (2014)

An Innovative Outdoor IoT System Architecture for Service Oriented Things

Takanobu Otsuka(✉), Yoshitaka Torii, and Takayuki Ito

Nagoya Institute of Technology, Nagoya, Japan
otsuka.takanobu@nitech.ac.jp
http://www.itolab.nitech.ac.jp

Abstract. In recent years, Internet of Things (IoT) has widely studied as a technology using the Wireless Sensor Networks (WSNs). The essential purpose of WSN is is the availability of the sensor stream. In addition, IoT is a research networking of the present in the real world "Physical device". In the Smart Home and Smart City has attracted attention as an IoT application and are intended to provide the best service for each residents and citizens. Therefore, as a study of the application layer, which provides the information and the environment necessary for the user, Things-Oriented (TO), such as Service-Oriented-Things (SoT), research is underway that is based on the IoT technology. However, it is assumed that a lot of research is installed in the power supply available location such as an indoor space, it is less premised to design changes of the sensors to be used. It has been the implementation of the many outdoor for WSN system. For example, research and for detecting a tornado early with a pressure sensor that is installed outdoors, it is an implementation of an ad hoc simple water gauge. This paper proposes the outdoor autonomous distributed IoT system for building SoT system reconfigurable, depending on the purpose in the outdoors.

Keywords: Internet of things · Service oriented things · Wireless sensor network · Field informatics

1 Introduction

In recent years, Internet of Things (IoT) has widely studied as a technology using the Wireless Sensor Networks (WSNs). The essential purpose of WSN is is the availability of the sensor stream. In addition, IoT is a research networking of the present in the real world "Physical device". In the Smart Home and Smart City has attracted attention as an IoT application and are intended to provide the best service for each residents and citizens. SoT is to provide a service desired by the user, if necessary, has been promoted research for the purpose of corresponding to a flexible application.

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_19

Therefore, as a study of the application layer, which provides the information and the environment necessary for the user, Things-Oriented (TO) [1,2], such as Service-Oriented-Things (SoT) [3], research is underway that is based on the IoT technology. However, it is assumed that a lot of research is installed in the power supply available location such as an indoor space, it is less premised to design changes of the sensors to be used.

It has been the implementation of the many outdoor for WSN system. For example, research and for detecting a tornado early with a pressure sensor that is installed outdoors [5], it is an implementation of an ad hoc simple water gauge [6,7]. This paper proposes the outdoor autonomous distributed IoT system for building SoT system reconfigurable, depending on the purpose in the outdoors.

The rest of our paper is structured as follows. Section 2 describes the structure and challenges of outdoor SoT system. Then, we propose an autonomous distributed IoT system for SoT in Sect. 3 and describe the our previous work in Sect. 4. Section 5 describes related previous studies and introduces our present study.Finally, we show future challenges and summarize our paper in Sect. 6.

2 Challenges to Realize the Outdoor SoT

Examples of outdoor SoT services, disaster prevention and disaster mitigation and livestock, and it is the agricultural sector. Especially in the disaster prevention and disaster mitigation areas are operated in a disaster area, as has been mentioned as a problem in the prior studies, high reliability is required. WSN which is operated outdoors is difficult supply of the external power supply. Therefore, it is required to autonomously operate using an environmental power generation typified by a solar battery. However, in the solar power generation, dirt and of solar panels, power generation amount is not stable due to the influence of the weather, there is a problem that the communication node stop frequently occurs.

In addition, in the agriculture and livestock sector, depending on the circumstances of the harvest season and livestock animals, sensors that require different. However, in some cases sensors required by expensive, not practical it is to keep connecting all sensors to be required for all nodes. Moreover, requires a different voltage supplied to each sensor, there are differences in connection methods, such as analog or digital. Therefore, to allow connecting the different sensors, it is necessary to design changes of the power supply circuit, changing the communication program for each sensor.

Therefore we propose a system for solving the problems of the following two points outdoor SoT needs.

- Controlling of the sensor power supply by the state of the node, easy system change of sampling interval
- Systems that do not require circuit modifications or program changes according to the connection sensor.

By solving the above problems, reconstruction will propose easy SoT system in accordance with the purpose.

3 Outdoor Autonomous Distributed IoT System for SoT

3.1 Realization of SoT by Middleware

In many WSN systems and IoT systems, sensors for use in development is determined. Design and implementation of hardware and software is performed. Therefore, for applications changes that occur after the implementation is necessary design changes for circuit board, and software. It these problems, is achieved by incorporating the middleware to the external MPU connected to the communication module.

We shown system outline in Fig. 1. According to the present system, we will solve the problem points shown in Sect. 2.

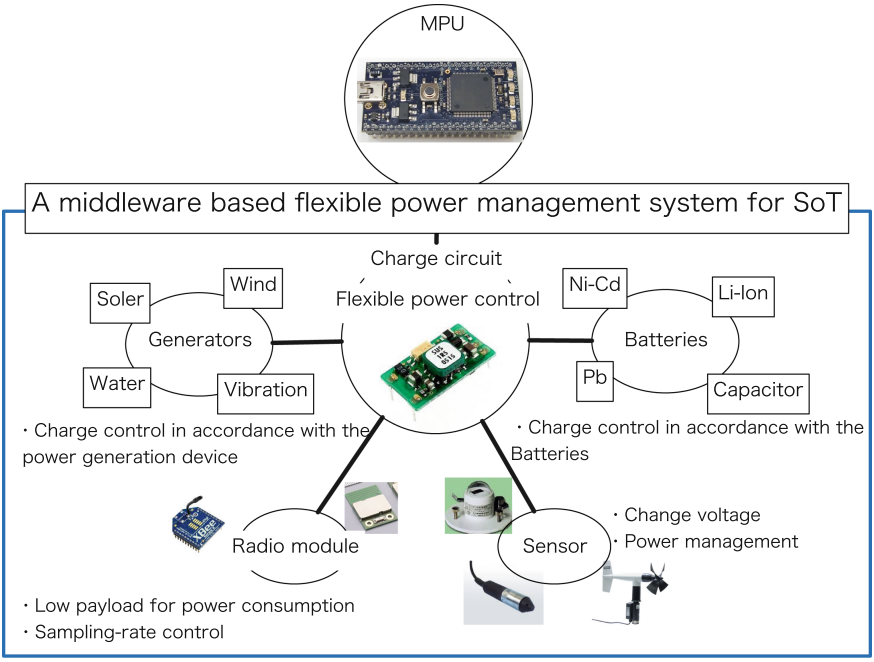


Fig. 1. System outline

3.2 Sensor Resource Allocation According to the Power Status

In this section, it is shown how to control the sensor resources in accordance with the status of a node. In energy harvesting, the charging voltage is not stabilized, the power supply to the sensor there are cases be difficult. In our proposed middleware, and performs monitoring of the charging voltage and current and the battery voltage node. Therefore, if there is the same sensor in the peripheral

nodes when the power supply voltage becomes unstable. It will stop the power supply of the sensor an alternative exists. Then, by transmitting only the sensor information that is not held by the neighbor nodes, and performs allocation of sensor resources. In addition, only the resource allocation if the power supply is insufficient, to change the sampling interval. We propose the method for performing continuous sensing using middleware.

Figure 2 shows an overview of sensor resource allocation. Thus, by performing the sensor resource allocation in accordance with the status of a node which is the subject of outdoor type SoT system can be made continuous sensing.

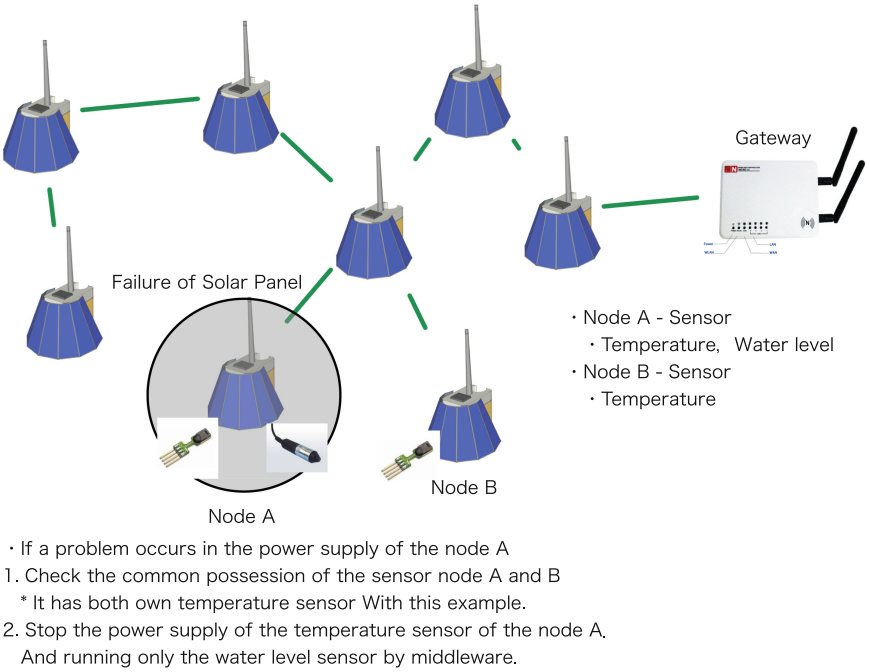


Fig. 2. Example of sensor resource allocation

3.3 Flexible Sensor Connection System

@ In this chapter, we show how to allow easy connection of various sensors. Typically, the sensor can and the supply voltage defined for each sensor, is operated by an interface such as a digital or analog, to obtain the sensor information. However, in general IoT systems, it is necessary to change the command control in digital the signal such as I2C and SPI by the class file Regarding the analog connection, there can be connected research to the analog input of the communication node [4]. However, it is not possible to change the supply voltage itself,

the connection of the sensor must be connected manually for the GPIO port to be specified, it is hard to say to be easily handle.

For our proposed middleware, a sensor connector in common, and it is possible to built a 1-wire [8] identification chip, thus manages the sensor id, it is possible to expand the server a sensor profile that requires through middleware, wide it is possible to accommodate a variety of sensors. Specifically, it wants to get the sensor profile by 1-wire device id that is built into the connector. Then, it becomes possible to control the supply voltage by an external power source control circuit, the sensor interface is also controlled by the middleware. From the above, when using only the time required for the sensor to be required, it is possible to obtain only the sensor information to the connector.

4 Development and Technical Challenges of Portable WSN

4.1 Required to Portable Ad-Hoc Water Gauge Capability

In recent years, and landslides caused by flooding of the river due to heavy rain, the number of such flooding of houses has increased. With regard to internal water damage, the structure and of the river, above sea level and, such as where a landfill, how completely prevent flood damage by internal water is difficult. At present, it is only possible addressing of Operating the drainage devices typified by car to pump submerged location. However, in order to quickly drained work as possible, it is necessary to understand the full flooding water by the water level data of the flooding point throughout. A mobile ad-hoc level gauge is mounted to the sink node to the pump truck to perform the drainage of flooded areas, and to aggregate the water level information from the measurement nodes installed anywhere. By their information, the water level conditions in the flooded area was visualized, of and calculation time, the development for the purpose of most water pump truck to high point to shorten the time required for waste water by moving conducted required for drainage to have. In addition, the cooperation with the pump future vehicle integrated management system is intended to improve the efficiency of waste water work using a pump wheel with a small number of operators. We show system outline in Fig. 3. The system, for the system to be operational in reality, high reliability unlike sensor node of research applications are required. Further, because it differs with the agricultural WSN are WSN and outdoor operation fixed, the following three problems are present.

Challenge of communication reliability Since the sink node, the measurement nodes are installed anywhere, it is difficult to consider the pre-position. Therefore, not only the communication distance required there to check the redundancy of the ad hoc network.

Challenges of the systems Since the measurement node itself moves, the current position in real time and it is necessary to clearly display the water level change.

Challenges in operation Because it is operated at disaster sites, ease of installation in the field, and maintainability of the time of failure is essential.

Also, unlike the stationary WSN, likely to external factors affecting the signal quality is changed dynamically, as in the fixed WSN, the advance radio wave condition is measured, and to optimize the placement is difficult. Therefore, that all communication nodes act as a relay station by the ad hoc network is required. Thus, rather than the hardware side, and identifying the measurement nodes and the relay node by setting the connection whether the water level gauge on the server software. Moreover, since the measurement nodes, the task of mounting the router nodes located disaster site, it is necessary to reduce the effort of the operator. Therefore, of course ease of installation, visualization of communication nodes location and, by performing the water level information visualization, it is required it is easy to use system for the site worker.

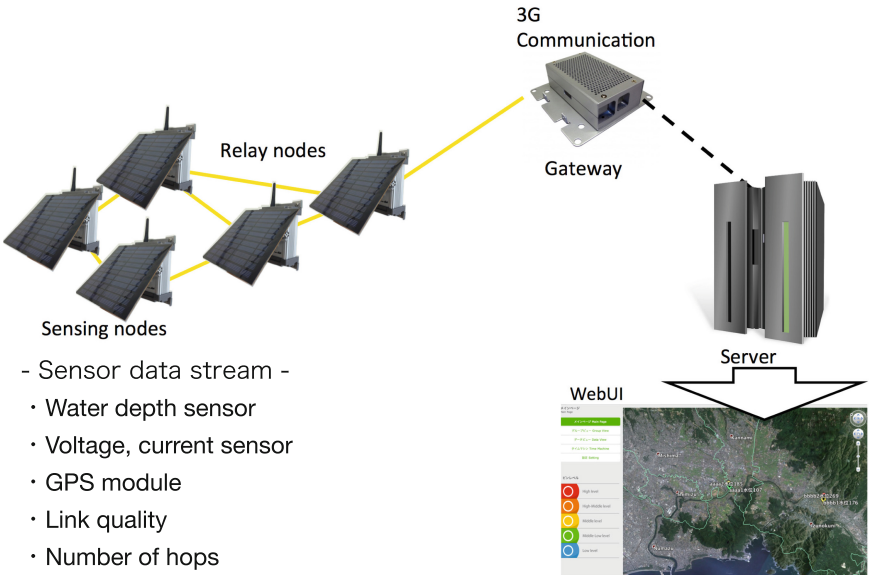


Fig. 3. System overview

4.2 The Operation Experimental Results with Actual Field

In this section, we show the results of operational experiments were performed at Mishima, Shizuoka on February 12, 2015. This experiment fields, are present in the branch line and the main line to the confluence position of Kano, it has experienced three times of flooding over the past 10 years, is designated as a flood hotspots in hazard map provided by Mishima that has been is a region. This experiment was carried out under the supervision of which is the main purchaser Ministry of Land, Infrastructure and Transport Chubu Maintenance

Division. The equipment layout in the experimental field, and for the test item is shown in Fig. 4. Experiments, assuming the inland water damage occurs, we observed the actual level due to installation and immersion depth sensor type portable ad hoc water gauge using the actual river. In this experiment it was confirmed whether satisfies the following performance.

- After local arrival, immediately confirmed that it is possible to operate.
- Confirmation of redundancy due to the ad hoc network when a part of the communication path becomes disconnected.
- Easiness of tablet viewing by webUI.

In addition to check these items, experiments were carried out at actual river watershed. Portable ad hoc water gauge is mounted in the following procedure.

1. Fixed the communication node by tube clamp on a tripod.
2. Writes tow anchor rivers, to irrigation canals.
3. And turned on by the mounted pilot wire water depth sensor to the anchor.
4. Connect the water depth sensor to communication node.

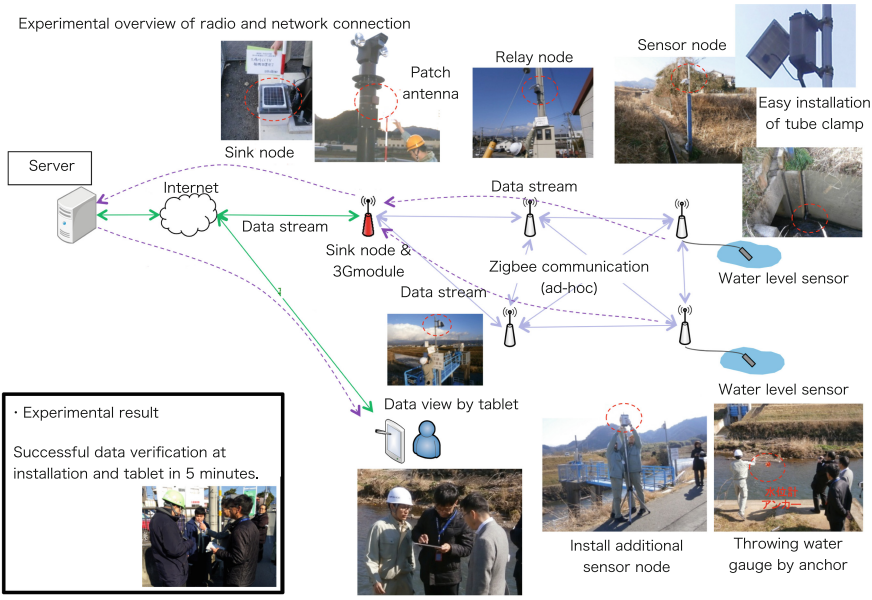


Fig. 4. Actual overview of test fields

the above procedure, it is possible to measure water level anywhere. Actual installation landscape and the steps Fig. 5. Established a gateway that is connected to the 3G communication network that assumes the pump car mounted on the site, and transmits the data to the web server. Transmitted data is processed

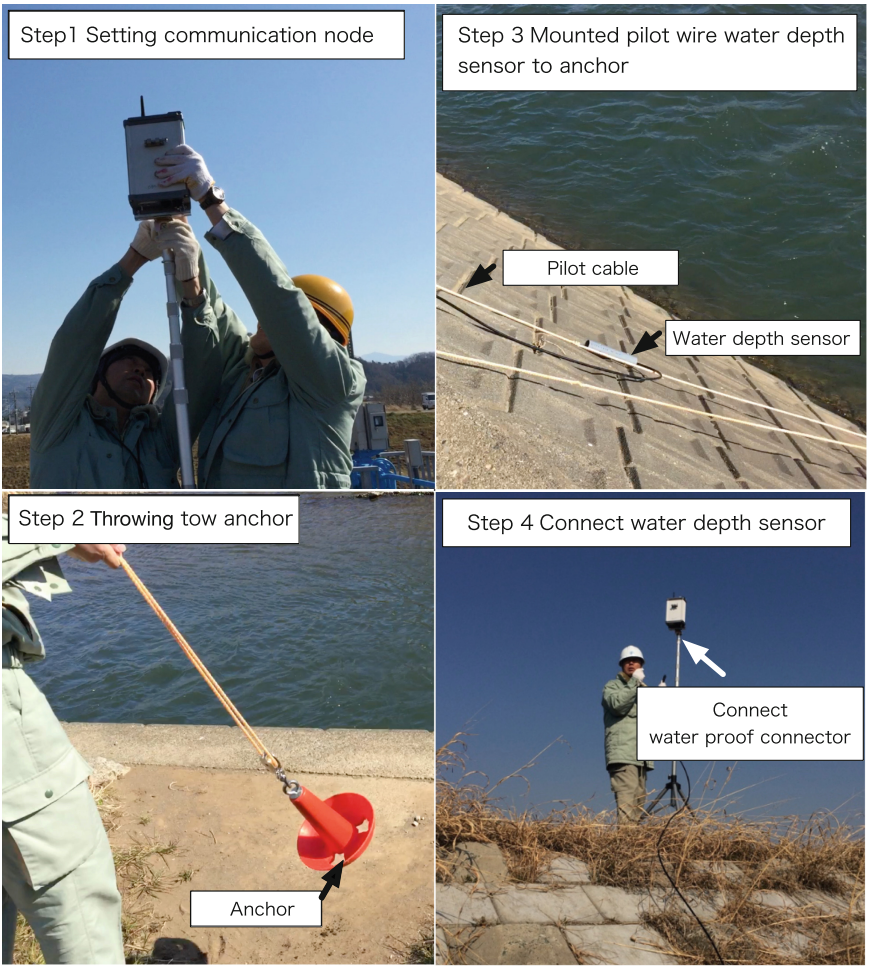


Fig. 5. Actual installation landscape and procedures

on the server, the position of the measurement nodes by performing a visualization and are visualized water level information. Visualized data by the server software can be viewed by using a tablet, it is possible to check the status of the internal water damage to real time. Therefore, efficient operation of the pump car is made possible. We have confirmed that it satisfies the following capability by the experimental results.

1. After unloading equipment, can be measured water level in 5 min
2. Even when it is intentionally cut off a portion communication path 30 seconds (one sampling period) within it communicatively detour to other routes.
3. About webUI, it can be viewed real data without having to explain many of the participants.

As shown above, the result of the operation experiment of this system, from the Ministry of Land, Infrastructure and Transport is the ordering party, has been developed by the present study, a portable ad hoc water gauge for operational possible internal water damage grasped at a disaster site is useful were evaluated as is. In addition, through the work in the field, or a need to change the wing nut so that it can be tightened by hand screwing part of the tripod, etc. becomes impossible communication in theory on the communication where possible, also a problem with the actual field operation can be extracted. In particular radio propagation in urban areas is often to determine the communication availability in concept sight, and the impact of the type waves reflected by the outer wall of the building, also been found that there are many parameters to determine the communication availability. Currently, we conduct ongoing operational test of the system, but some in the state has been fixed since March 2015, in Mishima, Shizuoka Misono district, using the node that was developed in this study, continuously water level I have continued to measure the data. By continuing test results, we have been continuously verify the reliability of long-term behavior of the hardware and software.

5 Related Work

5.1 Research Trends in WSN

The technologies of wireless sensor networks (WSNs), Internet of Things (IoT), and Machine to Machine (M2M) form this fields core [9]. Nodes, which constitute WSNs, temperature, and illumination, acquire such sensor data as acceleration and constitute a multi-hop ad-hoc network to transfer the obtained data by a bucket relay system using radio waves [10,11]. WSNs can simply configure autonomous networks by setting nodes and thus, reduce the installation work in field use. We can also obtain sensor data to capture real-world dynamics, and tracking and monitoring an objects natural environment are another promising WSN application. Wireless sensor networks, which are composed of a number of slave units and a sink node for aggregating information from a sensing node and a sink mode, of a multi-hop network in the information of relay capable router node. The observation environment information is beginning to be widely used in fields such as smartphones. In research applications, we give the example of MOTE [12], which was developed at UC Berkeley in 1998. In recent years, it has been standardized as IEEE802.15.4 and actively studied with a ZigBee network. However, compared to other countries, for its licensed sales in Japan, the construction authentication of unnecessary communication modules, the output has been severely limited, which is a large limitation in communication distance. In recent years, individual interoperable environment information acquisition devices are widely sold for research applications and homes.

One wireless sensor system is MEMSICs “Eko system” [13]. Eko is widerly using agricultural fields. Example research [14] has also been performed with poor scalability. As an example of a consumer product, there is Davis Vantage rice series [15], which is only capable of operating in serious applications; it

is not suitable for large-scale data collection. As an example of environmental information settings in an experimental WSN, we can install a large number of observation nodes in an active volcano, and research was also performed on data collection to predict eruptions [16]. We developed a sensor network device and a server application that collects high density information to gather environmental data on a large-scale. In addition, in wireless sensor network fields, other research [17] has optimized the sensor arrangement to reduce power consumption and to construct sensor node applications to apply agent technology [18]. Furthermore, to manage the operational aspects, we can prepare program nodes according to purpose, and research has constructed a sensor network according to applications [19].

Depending on the circumstances, autonomous study focuses on network topology [20]. Thus, much research has focused on WSNs as a framework for performing flexible measurements.

5.2 Studies on Radio Wave Quality Stabilization of WSNs

The research aim of managing operational aspects can prepare program nodes to facilitate the construction of a sensor network according to applications [19]. Depending on the circumstances, autonomous study can constitute network topology [20].

However, the most desired WSN performance is a stable network. By optimizing the placement of fixed sensor networks, continuous data collection studies constantly maintain network quality. For example, studies use the radio wave level (RSSI values) of the topology formed by the nodes to optimize the node arrangement in a predefined spatial design [21]. To capture the node as a cluster, research has optimized the communication route [22].

By optimizing the relay stations position for multi-hop communication among nodes, it's possible to reduce communication errors [23]. By arranging a plurality of sink nodes, the position of the sink node can be optimized [24]. However, other research optimized closed environments to secure the communication quality of mobile-type WSNs when the ambient environment changes dynamically. Directly applying portable sensor networks that operate in real fields is difficult.

Sensor networks are generally constructed in a wide range to obtain environmental information. In preliminary surveys, we identified the need to communicate distance, and in agriculture and river management operations, we measured the radio wave state around installation sites based on installation techniques for the placement of measurement nodes. However, pre-installation of the above portable sensor networks to measure the radio wave state is an unrealistic field operations.

In recent years in large outdoor sensing, nodes have used many techniques to extend the communication distance by acting as relay nodes [25,26]. However, little consideration has been given to speed up the reconnection of timeline failures; if real-time data acquisition is required, the reconnection time will continue to be a problem.

When sensor nodes are operating as relay stations operating a sensor network outdoors become, due to power consumption increases. We solved this problem by optimizing both the hardware and software architecture for full-mesh networks. In this paper, we proposed an autonomous reconnection, high-speed full-mesh network called a Rapid Re-routing Full-Mesh Network for failure times of both the communication quality and communication distance performances.

6 Conclusion

In this paper, we showed the architecture of outdoor autonomous distributed IoT system for implementing the SoT in outdoor. This study has already been collected in the various knowledge of WSN [5–7] of the outdoors is complete. We are also conducting joint research of WuKong [4] that have developed at the University of California, Irvine for middleware, and is promoting the implementation. In the future, the sensor and resource allocation when power is reduced, we will perform the actual implementation of the system of the sensor identification by 1-wire.

References

1. Gubbi, J., et al.: Internet of Things (IoT): a vision, architectural elements, and future directions. Fut. Gen. Comput. Syst. **29.7**, 1645–1660 (2013)
2. Atzori, Luigi, Antonio, I., Giacomo, M.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
3. Teixeira, Thiago, et al. : Service oriented middleware for the internet of things: a perspective. In: Towards a Service-Based Internet. Springer, Berlin, Heidelberg, pp. 220–229 (2011)
4. Lin, K.-J., et al.: Building smart m2m applications using the wukong profile framework. Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom). In: IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE (2013)
5. Takanobu, O., Yoshitaka, T., Takayuki, I.: gAnomaly Weather Information Detection using Wireless Pressure-Sensor GridhC. J. Informat. Process. (JIP), **23**(6), Information Processing Society of Japan (IPSJ) (2015)
6. Otsuka, T., Yoshitaka, T., Takayuki, I.: Challenges and implementation of ad-hoc water gauge system for the grasp of internal water damage. In: 14th International Conference on Computer and Information Science (ICIS), 2015 IEEE/ACIS. IEEE (2015)
7. Otsuka, T., Yoshitaka, T., Takayuki, I.: A proto-type of a portable ad hoc simple water gauge and real world evaluation. In: The Proceedings of the 2nd International Workshop on Smart Simulation and Modelling for Complex Systems
8. Sepetka, Ivan, Erik, T.E.: Guide wire device. U.S. Patent No. 4,832,047. 23 May 1989
9. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**, 393–422, Elsevier Science B.V (2002)
10. Chong, Chee-Yee, Srikanta, P.K.: Sensor networks: evolution, opportunities, and challenges. Proc. IEEE **91**(8), 247–1256 (2003)

11. Kuorilehto, M., Hannikainen, M., Hamalainen, T.D.: A survey of application distribution in wireless sensor networks. EURASIP J. Wirel. Commun. Netw. **5**(5), 774–788 (2005)
12. Wireless Sensor Network MOTE. http://www.xbow.jp/01products/index.html
13. MEMSIC. Inc, Wireless Sensor Network, Eko. http://www.memsic.com/wireless-sensor-networks/
14. Massage Ito, Environmental Monitoring Project with Sensor Network Technology. IEEE Japan. **2010**(66) (2010)
15. Davis Instruments, Weather Products. http://www.davisnet.com/weather/
16. Geoffrey, W.-A., Konrad L., Matt W., et al.: Deploying a wireless sensor network on an active volcano. IEEE Int. Comput. March April, 18–25 (2006)
17. Dhananjay, L., Arati, M., Falk, H., Elif Uysal-Biyikoglu, Abtin, K.: Measurement and characterization of link quality metrics in energy constrained wireless sensor networks. In: Proceedings of IEEE International Conference on Global Communications Conference, pp. 446–452 (2003)
18. Cho Kenta, Ohsuga, A., Honiden, S.: Intelligent mobile agent framework for multipurpose wireless sensor network application. (Mobile Applications, Special Issue, Mobile Communications and Intelligent Transportation Systems towards the Ubiquitous Era) IPSJ Journal Information Processing Society of Japan (IPSJ)
19. Hill, J., Szewczyk, R., Woo, A., Hoilar, S., Culler, D.E., Kristofer, S., Pister, J.: System Architecture Directions for Networked Sensors. Arehitect. Support. Prog. Lang Operat. System **3**, 93–104 (2000)
20. Fbk, CL., Roman, G.-C., Lu, C.: Rapid development and flexible deployment of adaptive wireless sensor network applications. In: Proceedings of 25th IEEE International Conference on Distributed Computing Systems Vbl, OO, pp. 653–662 (ICDes, 05). (2000)
21. Romer, K., Friedemann, M.: The design space of wireless sensor networks. Wireless Communications. IEEE 11.6 (2004) 54–61
22. Younis, Ossama, Krunz, Marwan, Srinivasan, R.: Node clustering in wireless sensor networks: recent developments and deployment challenges. Network, IEEE 20.3 pp. 20–25 (2006)
23. Hongyang, C.H.E.N., et al.: An improved DV-Hop localization algorithm with reduced node location error for wireless sensor networks. IEICE Trans. Fundament. Elect. Commun. Comput. Sci. **91.8**, 2232–2236 (2008)
24. Akihide, U., Hisao, Y.: Particle swarm optimization for computing plural acceptable solutions and its application to sink node allocation problem in wireless sensor networks with multiple sinks. The IEICE Transactions on Information and Systems, The Institute of Electronics, Information and Communication Engineers (in Japanese)
25. Keigo Koizumi, Yukishige Fujita, Kenji Hirata, Kazuhiro ODA, Sadayuki KAMIDE, Experimental research toward practical use of wireless sensor network for slope disaster monitoring, Journal of Japan Society of Civil Engineers, Ser. C Geosph. Eng. **69**(1), 46–57 (2013)
26. Liang, Nia-Chiang, et al.: Impact of node heterogeneity in ZigBee mesh network routing. Systems, Man and Cybernetics. SMC'06. IEEE International Conference on. vol. 1. IEEE (2006)

Ontology-Based Architecture to Improve Driving Performance Using Sensor Information for Intelligent Transportation Systems

Susel Fernandez[1,2(✉)] and Takayuki Ito[2]

[1] University of Alcala, Alcalá de Henares, Madrid, Spain
susel.fernandez@uah.es
[2] Nagoya Institute of Technology, Nagoya, Japan
ito.takayuki@nitech.ac.jp

Abstract. Intelligent transportation systems are advanced applications with aim to provide innovative services relating to road transport management and enable the users to be better informed and make safer and coordinated use of transport networks. A crucial element for the success of these systems is that vehicles can exchange information not only among themselves but with other elements in the road infrastructure through different applications. One of the most important information sources in this kind of systems is sensors. Sensors can be located into vehicles or as part of an infrastructure element, such as bridges or traffic signs. The sensor can provide information related to the weather conditions and the traffic situation, which is useful to improve the driving process. In this paper a multiagent system using ontologies to improve the driving environment is proposed. The system performs different tasks in automatic way to increase the driver safety and comfort using sensor information.

Keywords: Intelligent transportation systems · Ontology · Reasoning
Agents · Sensors

1 Introduction

Today, it is known as Intelligent Transportation Systems, the set of applications and technological systems created with the aim of improving safety and efficiency in road transport. These systems allow to control, manage and monitoring the different elements of roads.

The continuing evolution of intelligent transportation systems has ushered in a new era of interconnected intelligent systems, which certainly has been a quantitative leap in safety of road transport. These systems enable the exchange of information between different applications, and the subsequent analysis to improving the safety of drivers and eases travel and comfort in road travel.

One of the most important information sources in this kind of systems is sensors. Sensors can be located into vehicles or as part of an infrastructure element, such as bridges or traffic signals.

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_20

The increasing miniaturization of computers raises the idea of developing extremely small, inexpensive computers that communicate wirelessly and are organized independently. A sensor network is a network of tiny computers (nodes) equipped with sensors which collaborate on a common task [1]. These nodes have certain sensory capabilities and wireless communications that enable ad hoc networking without any preset physical infrastructure or central administration. Sensor networks are a relatively new concept in data acquisition and processing for multiple applications in different fields such as industry, medicine, home automation, military environments, environmental detection, etc. Their main features (to be small, cheap, autonomous, easy to deploy, self-configurable and able to perform efficient energy management) have made sensor networks a very active research field, in which systems as diverse as Berkeley Motes [2], Pico-Radio [3], Smart-Dust [4] and WINS [5] have been developed.

Due to its high degree of expressiveness, the use of ontologies is crucial to ensure greater interoperability among software agents and different applications involved in intelligent transportation systems. Ontologies provide a common vocabulary in a given domain and allow defining, with different levels of formality, the meaning of terms and the relationships between them.

This paper presents a multiagent architecture using ontologies to improve the driving environment. The system performs different tasks in automatic way to increase the driver safety and comfort using sensor information.

The paper is organized as follows. Section 2 is a review of the state of the art in ontologies for road transportation systems, Sect. 3 presents the architecture of the proposed multiagent system. In Sect. 4 the cases of study are explained. Finally the conclusions and lines of future work are summarized in Sect. 5.

2 Related Works

There are some previous works focused on ontology for road transportation systems. In [6] an ontology to represent traffic in highways has been developed. Its aim was the construction of reliable Traffic Information System providing information about roads, traffic, and scenarios related to vehicles in the roads. It also provides ways to the Traffic Information System to analyze how critical a specific situation is. For example, an ambulance may need to know about the congestion status of a toll plaza. Requesting this information is critical if the ambulance is moving to the scene of an accident. On the other hand, if a common vehicle is moving through a road without hurry, then its information requested is not critical.

In [7] they proposed a high-level representation of an automated vehicle, other vehicles and their environment, which can assist drivers in taking “illegal” but practical relaxation decisions (for example when a car damaged does not allow the circulation, take the decision to move to another lane crossing a solid line and overtake the stopped car, if the other lane is clear). This high-level representation includes topological knowledge and inference rules, in order to compute the next high-level motion an automated vehicle should take, as assistance to a driver.

In [8] an ontology-based spatial context model was proposed. The work takes a combined approach to modeling context information utilized by pervasive transportation

services: the Primary-Context Model facilitates interoperation across independent Intelligent Transportation Systems, whereas the Primary-Context Ontology enables pervasive transportation services to reason about shared context information and to react accordingly. The independently defined, distributed information is correlated based on its primary-context: location, time, identity, and quality of service. The Primary-Context Model and Ontology have been evaluated by modelling a car park system for a smart parking space locator service.

The work proposed in [9] is an approach to create a generic situation description for advanced driver assistance systems using logic reasoning on a traffic situation knowledge base. It contains multiple objects of different type such as vehicles and infrastructure elements like roads, lanes, intersections, traffic signs, traffic lights and relations among them. Logic inference is performed to check and extend the situation description and interpret the situation e.g. by reasoning about traffic rules. The capabilities of this ontological situation description approach are shown at the example of complex intersections with several roads, lanes, vehicles and different combinations of traffic signs and traffic lights.

In the work in [10] an ontology for traffic management is proposed, adding certain concepts of traffic to general sensor ontology A3ME [11]. The added concepts are specializations of position, distance and acceleration sensor classes, and the different actions that take place in the car motions. The ontology is developed in OWL, using the JESS reasoner with SWLR [12] rules.

In [13] an ontology-based Knowledge Base, which contains maps and traffic regulations, was introduced. They can aware over speed situations and make decisions at intersections in comply with traffic regulations, but they did not consider important elements such as traffic signals and weather conditions.

Most of the works found in the literature focus on describing very specific traffic situations such as finding parking, actions of emergency vehicles and intersection situations. But none of them is general and expressive enough to encompass a wide variety of traffic situations. Therefore it's necessary to develop ontologies in the domain of road traffic expressive enough to describe any traffic situation. The ontologies should be richer with respect to the various sensor inputs, and use it to drastically improve the general routing mechanism.

3 System Architecture

The proposed system is composed of four interconnected layers, as shown in Fig. 1. At the bottom is the sensor layer, in which the different sensors collect the information related to the traffic environment. That information is stored in the Database. The next one is the ontological layer, where we have developed a general ontology of the domain of road traffic in OWL. The ontology has many SWRL rules, which are used for the reasoner to infer new knowledge. Finally in the upper layer are the agents, who perform their tasks accessing the information stored in the ontology through SPARQL queries.

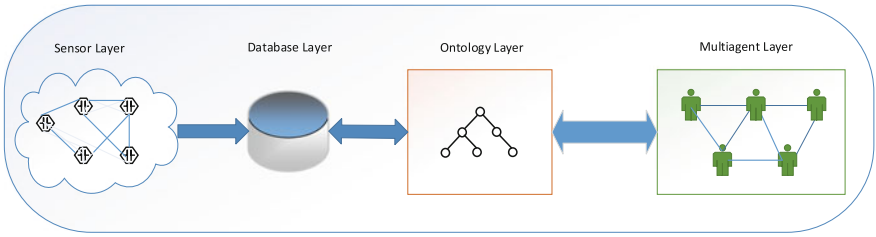


Fig. 1 System architecture

3.1 The Ontology

In the ontology layer of the system, an ontology that relates the different road traffic entities has been developed. The ontology was implemented in OWL-RDF language [14] using the protégé tool [15]. In this work we used the reasoner Pellet [16], which is implemented in Java; it is freely available and allows checking the consistency of the ontology.

For better understanding, we divided the knowledge in the traffic ontology divided in different groups of interrelated concepts.

The main group is related to vehicles. The taxonomy of vehicles is composed of: commercial vehicles, public vehicles (bus and taxi), private vehicles (car, bicycle and motorbike) and priority vehicles (ambulances, fire trucks and police cars). Different relationships between vehicles and other entities are defined also in this group. Some of these entities are: location, showing the exact location (latitude and longitude) of a vehicle, route point or infrastructure item; information about drivers and the vehicle's types which they can drives by license.

The second group organizes the elements related to road infrastructure. In this group the most important concept represents the roads, which in Japan are classified as local roads, prefectural roads, national highways and national expressways.

For better management of traffic situations the roads have been divided into segments, connected through intersections. Each segment contains lanes, and different signs such as stop signs or speed control, traffic lights or road markings are in each lane. Each signal has an action associated following the Japanese traffic regulations.

Here we focused in the third group, which is concern with the concepts related with the different types of sensors used in intelligent transportation system scenario. Figure 2 shows the sensor taxonomy and the relations between the sensors with other concepts of the ontology. The sensors can be located in both the vehicles as in different infrastructure elements, for example as part of bridges, roads, traffic signs, etc.

The principal sensors we use in this work are related to environmental measures and to flow detection. As environmental sensor we have snow and fog detectors, temperature and humidity sensors. Regarding to flow detection we use crowd, pedestrians and car sensors.

As we can see in the figure, sensors can be classified taking into account their energy requirements into active and passive; regarding the nature of the measures into magnetic, chemical, mechanic and thermic; and regarding the nature of the output into digital, analogic or temporal.

We have defined the relation *has_output_nature* (brown color) to specify the nature of each sensor output; the relation *has_measure_nature* (yellow color) to specify the nature of the physical variable measured; and finally the relation *has_energy_requirements* (gray color) is used to differentiate between the sensors that generate the output signal without the need for external power supply, of those that require power to perform this function.

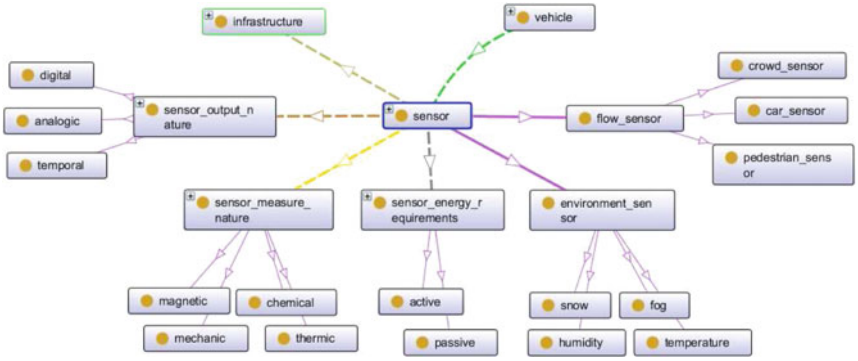


Fig. 2 Concepts related to sensors

3.2 The Multi-agent System

At the top layer is the multiagent system. Here the different agents perform their tasks using the information stored in the ontology through different queries. The multiagent platform was developed using de *Java Agent Development Framework* (*JADE*). As ontology query language, SPARQL [17] has been used.

The agents that we have defined in the system are:

- *Driver Personal Agent*: This agent is in charge of performing the driver personal tasks, taking into account the driver preferences and the driver behavior.
- *Environment Agent*: This agent is responsible of providing the information related to the environment, measured by the vehicle sensors, for example the temperature and the humidity values.
- *Air Conditioner Agent*: The principal task of this agent is the air conditioner settings, taking into account the values of temperature and humidity measured inside and outside the vehicle and the driver's preferences.
- *Car Agent*: This agent was defined to perform in automatic way, the different actions related to the car movements following the specific route, the traffic signs and traffic regulations.
- *Road Agent*: This agent is responsible of performing the tasks related to the road, for example, determine if the road is congested taking into account the information of the traffic flow collected by the sensors.
- *Traffic Light Agent*: This agent perform the tasks related with the traffic light settings, such as: adjust the duration of the traffic light taking into account the traffic flow and the weather conditions.

- *Weather Agent*: This agent performs the tasks related to the weather conditions, for example provide the weather information to the rest of the agents of the system and make predictions related to the weather.

4 Cases of Study

Two different cases of study are presented in this section. The first one is concerned with the air conditioner settings depending on the temperature and humidity values measured inside and outside the car, and also the user's preferences regarding the environment parameters. The second one is related to the traffic light duration time adjustment taking into account different traffic elements such as the level of congestion of the roads and the weather conditions. With the accomplishment of these tasks the system allows the improvement of the safety and comfort of the driving process.

4.1 The Air-Conditioner Setting Scenario

This is a simple example scenario and consist in regulate the air conditioning of the car for better comfort while driving. As we can see in Fig. 3, three agents are involved in this task, the air-conditioner agent, the environment agent and the driver personal agent. To perform this task, the temperature and humidity sensors located in the vehicles are needed.

The air conditioning setting task consists in compare the temperature and humidity measured inside the car with the temperature and humidity outside. With these values and the values of the driver's preferences (managed for the driver personal agent), the air-conditioner agent is able to make the decision of adjust the temperature and humidity of the car. The accomplishment of this task allows the comfort of the driving ambient.

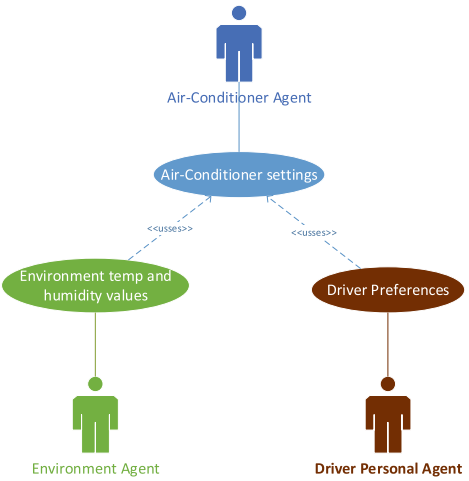


Fig. 3 Use case diagram of the air-conditioner settings task

Figure 4 shows a sequence diagram of the air conditioner setting use case. As we can see in the figure, first, the air conditioner agent sends two request message regarding temperature and humidity, one asking for the temperature and humidity values to the environment agent and the other asking for the driver preferences to the driver personal agent. A person may have different preferences in terms of temperature depending on where is located.

Therefore, once the personal agent receives the request, he consults with the ontology to know the location of the person. Then, knowing the location, the agent queries the database for the user preferences for that particular location and sends that values to the air conditioner agent. Finally, knowing the driver preferences and the environment temperature and humidity values, the air conditioner sets the right combination of parameter values to ensure user comfort.

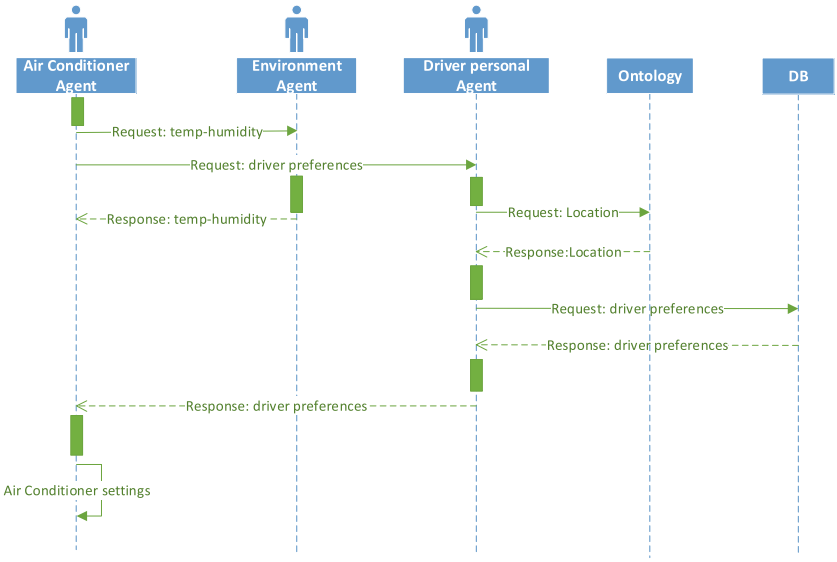


Fig. 4 Sequence diagram of the air-conditioner settings task

4.2 Traffic Light Adjustment Scenario

The second scenario is related to the traffic light adjustment, taking into account the traffic flow and the weather conditions. Figure 5 shows the use case diagram of the traffic light adjustment scenario. Two use cases are involved in this task. These use cases are: the flow detection and the weather events detection in the specific road.

Figure 6 shows the sequence diagram to accomplish the traffic flow detection task. First, the agent queries to the ontology to obtain the information about the segment and

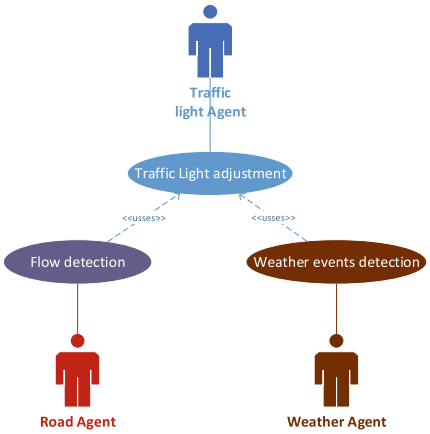


Fig. 5 Use case diagram of the traffic light adjustment task

lane where the traffic light is located. The ontology contains detailed information on each segment and lane, such as its location (latitude and longitude) and the maximum density of the lane. From the coordinates of the beginning and end of each lane segment, using the SWRL rules, the reasoner can infer the length of the lane.

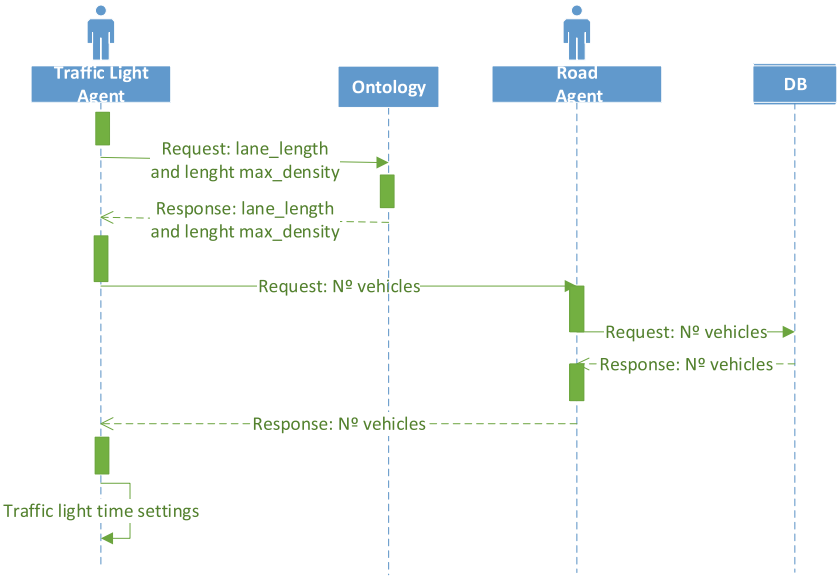


Fig. 6 Sequence diagram of the traffic light duration time task

Taking into account the length of the lane and the number of vehicles detected by the road sensor, the traffic light Agent can compute the real density of the lane. The density is defined as the number of vehicles per unit length of the roadway at a specific time [18]. The equation of the density is the following:

$$density = \frac{N}{L},$$

where N is the number of vehicles detected by the road sensor in the specific lane, and L is the length of the lane. Once the lane density is computed, the value is compared to the maximum density. If the actual density is greater or equal than the maximum density, it is assumed that lane is congested.

Finally to decide the duration of the traffic light state we use the following rules: if the traffic light is in red, and the road segment is congested, then the duration time of the light is decreased; however, if the traffic light is in red and the road segment is not congested, then the duration time of the light is maintained in the default value. If the traffic light is green and the road segment is congested, then the duration time of the light is increased, while if it is not congested, the duration time is maintained.

5 Conclusions

In this paper a multiagent system using ontologies to improve the driving environment is proposed. The system performs different tasks in automatic way to increase the driver safety and comfort using sensor information.

The architecture of the proposed system consists of four interconnected layers. At the bottom is the sensor layer, in which the different sensors collect the information related to the traffic environment. The next one is the ontological layer, where we have developed a general ontology of the domain of road traffic in OWL. Finally in the upper layer are the agents, who perform their tasks accessing the information stored in the ontology through SPARQL queries. Examples of these tasks are the air conditioner parameters' setting and the adjustment of the traffic light duration time, taking into account the traffic flow and the weather conditions. This system allows us to give higher priority to road segments that are congested over others.

Several experiments have been performed, simulating actual road traffic situations. It has been shown that the automatic adjustment of the duration of the traffic lights contributes to optimize the traffic flow, allowing drivers to gain time along the route. To carry out this process, the system only takes into account the level of congestion and the weather conditions of the corresponding road segment. However there are many other factors involved in this process which have not been taken into account, such as the congestion level of other road segments along the route. Taking this into account, we propose as future work, to carry out different negotiation algorithms between the agents involved in deciding the status and duration of traffic lights along the route.

Acknowledgements. This work was partially supported by Research and Development on Utilization and Fundamental Technologies for Social Big Data by NICT (National Institute of Information and Communications Technology), and the Fund for Strengthening and Facilitating the National University Reformations by Ministry of Education, Culture, Sports, Science, and Technology, Japan.

References

1. Verdone, R., Dardari, D., Mazzini, G., Conti, A.: Wireless Sensor and Actuator Networks. Elsevier, London, UK (2008)
2. Hill, J., Culler, D.: A wireless embedded sensor architecture for system level optimization, UC Berkeley Technical Report, University of California: Berkeley, CA, USA (2001)
3. Rabaey, J., Ammer, J., Da Silva, J.L., Jr., Patel, D.: PicoRadio: ad hoc wireless networking of ubiquitous low-energy sensor/monitor nodes. In: Proceeding of IEEE Computer Society Workshop on VLSI, Orlando, FL, USA, 27–28 April 2000, pp. 9-12
4. Kahn, J.M., Katz, R.H., Pister, K.S.J.: Next century challenges: mobile networking for "Smart Dust". In: Proceeding of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, New York, NY, USA, Aug 1999, pp. 271–278
5. Asada, G., Dong, M., Lin, T.S., Newberg, F., Pottie, G., Kaiser, W.J.: Wireless integrated network sensors: low power systems on a chip. In: Proceeding of the 24th European Solid-State Circuits Conference (ESSCIRC 1998), The Hague, Netherlands, 22–24 Sep 1998, pp. 9–16
6. Gorender, S., Silva, Í.: An ontology for a fault tolerant traffic information system.In: 22nd International Congress of Mechanical Engineering (COBEM 2013), Ribeirão Preto, SP, Brazil, 3–7 Nov 2013
7. Pollard, E., Morignot, P., Nashashibi, F.: An ontology-based model to determine the automation level of an automated vehicle for co-driving. In: Fusion pp. 596–603 (2013)
8. Lee, D., Meier, R.: Primary-context model and ontology: a combined approach for pervasive transportation services. In: First International Workshop on Pervasive Transportation Systems (PerTrans 2007), with the Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, pp. 419–424. PerCom 07 (2007)
9. Hülsen, M., Zöllner, J.M., Weiss, C.: Traffic intersection situation description ontology for advanced driver assistance. In: 2011 IEEE Intelligent Vehicles Symposium (IV) Baden-Baden, Germany, 5–9 June 2011
10. Bermejo, A.J., Villadangos, J., Astrain, J.J., Cordoba, A.: Ontology based road traffic management. In: Fortino, G et al. (eds) Intelligent Distributed Computing VI, pp. 103–108. SCI 446
11. Herzog, A.; Jacobi, D.; Buchmann, A.: A3ME-An agent-based middleware approach for mixed mode environments. In: Proceeding of Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2008), Valencia, Spain, 29 Sep–4 Oct 2008; pp. 191–196
12. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML, Submission to W3C, May 2004. http://www.w3.org/Submission/SWRL/
13. Zhao, L., Ichise, R., Mita, S., Sasaki, Y.: Ontologies for Advanced Driver Assistance Systems
14. Dean, M., Schreiber, G.: OWL Web Ontology Language Reference. http://www.w3.org/TR/2004/REC-owl-ref-20040210/ (2004)

15. Protégé. http://protege.stanford.edu/
16. Pellet. http://clarkparsia.com/pellet/
17. SPARQL. http://sparql.org/
18. Lieu, H.: Traffic-flow theory. Public Roads (US Dep of Transportation) **62**(4)
19. Fernandez, S., Hadfi, R., Ito, T., Marsa-Maestre, I., Velasco, J. R.: Ontology-based architecture for intelligent transportation systems using a traffic sensor network. Sensors **16**(8), 1287 (2016)

An Analysis of Allocation Stability on Approximation-Based Pricing for Multi-unit Combinatorial Auctions

Naoki Fukuta(✉)

Shizuoka University, Hamamatsu, Shizuoka 4328011, Japan
fukuta@cs.inf.shizuoka.ac.jp
http://whitebear.cs.inf.shizuoka.ac.jp/

Abstract. In this paper, a discussion and an analysis about the *stability* on pricing and allocation of resources are presented. On the discussion, an approximate auction which has VCG-like pricing mechanism is used when cancellation of winner bid(s) after its winner determination is considered. An analysis about *stable* approximate pricing mechanisms against cancellation of a winner after its winner determination is also presented. In there, a single-unit non-combinatorial reserve price bidding on a multi-unit combinatorial auction could also be employed as well. The pricing algorithm employs an approximate allocation and pricing algorithm that is capable of handling multi-unit auctions with reserve price biddings. This type of auction is expected to be applied to a situation when we consider an allocation of electricity while considering electricity generation costs on the power suppliers in more realistic configurations, i.e., some bidders might be untrustful in their ability. Based on the experimental analysis, the algorithm effectively produces approximation allocations that are necessary in the pricing phase, as well as yielding better stability in the case of single-winner cancellation. It also behaves as an approximation of VCG(Vickrey-Clarke-Groves) mechanism satisfying budget balance condition and bidders' individual rationality without enforcing the single-minded bidders assumption.

Keywords: Multi-unit auctions · Combinatorial auctions
Approximation

1 Introduction

There are numerous types of problems on making consensus among entities in difficult situations, e.g., some conflicts existed among those entities [22]. Assigning limited amount of resources to self-interested entities (e.g., agents) [31] is one of important research fields that is aiming to provide solutions for such kind of problems. There are emerging needs for sophisticated approaches for dynamic allocation of limited resources, considering a situation with rational and self-interested attendees [1,21]. There are several criteria to stably assign those

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_21

limited resources, e.g., fair allocations without money [2], or efficient allocations using money [4].

Auction-based resource allocation is one of the ideas to allocate resources efficiently and stably, i.e., once they are allocated, no attendee will have incentives to try to break the assignments to obtain further allocations of goods when the auction is *well designed* as a mechanism [4]. There has been proposed a number of variants of auctions to capture various conditions in the resource allocation problems [4]. For example, multi-unit auctions can handle amount of resources to be allocated for each bidder in an auction [27], and combinatorial auctions can handle complex allocations among different types of resources [4,28]. There exist numerous works on giving important theoretical properties, i.e., *incentive compatibility* [23,25], *strategy proofness* and other related properties [35], or considering *locally envy free equilibrium* [5] or similar properties [8] for more realistic scenarios.

Most of those approaches assumed that once winners had their allocations of auctioned items, all those winners always have capability to pay for them, or have rights for the allocations [4]. However, there might be cases that one or more winners have been cancelled for allocating items that were auctioned [3], because of keeping revenue of the auctioneer, or just some winners did not have the rights to have them on a regulation. Such cancellations could be done by both buyers, sellers, and even by the auctioneers, due to some misconduct of the buyers or sellers. Although there are some ideas to prevent buyers from doing misconduct such as false name biddings, these mechanisms could not easily be realized when an auction would have other important characteristics [32].

Because of the nature of combinatorial optimization problem, only a small change of its condition may affect to large parts of the allocation when we recalculate its (sub)optimal allocations from scratch. In the above-mentioned context, we would say the allocation may not be *stable* for cancellations of bids.

In this paper, a discussion is presented that are concerning to the *stability* on pricing and allocation of resources using an approximate auction which has VCG-like pricing mechanism when it considers cancellation of winner bid(s) after its winner determination. An idea about more *stable* approximate pricing mechanism against cancellation of a winner after its winner determination is also presented. In there, a single-unit non-combinatorial reserve price biddings on multi-unit combinatorial auction can also be employed to consider the costs of producing resources to be allocated.

2 Preliminary

2.1 Multi-unit Combinatorial Auctions

Combinatorial auction is an auction that allows bidders to place bids for a combination of items rather than a single item [4].

The winner determination problem on single unit combinatorial auctions is defined as follows [4]: The set of bidders is denoted by $N = 1, \ldots, n$, and the set of items by $M = \{m_1, \ldots, m_k\}$. $|M| = k$. Bundle S is a set of items: $S \subseteq M$.

We denote by $v_i(S)$, bidder i's valuation of the combinatorial bid for bundle S. An allocation of the items is described by variables $x_i(S) \in \{0, 1\}$, where $x_i(S) = 1$ if and only if bidder i wins bundle S. An allocation, $x_i(S)$, is feasible if it allocates no item more than once, for all $j \in M$.

$$\forall j \in M \quad \sum_{i \in N} \sum_{S \ni j} x_i(S) \leq 1$$

The winner determination problem is the problem to maximize total revenue for feasible allocations $X \ni x_i(S)$.

$$\max_X \sum_{i \in N} \sum_{S \subseteq M} v_i(S) x_i(S)$$

Note that I used simple *OR-bid* representation as the bidding language. Substitutability can be represented by a set of atomic *OR-bids* with *dummy items* [4]. On *OR-bidding*, each bid can be a winner unless there is a *conflict* with other bids (e.g., the situation in which two bids are placed for the same item but the item can only be assigned to one). To represent substitutability (e.g., a situation that "a bidder wants to have only one of two items"), the bidder can use a virtual item called *dummy item*, which can only be assigned to one bid and can place two *OR-bids* both of which include the same dummy item in their bundle of items.

When some items in an auction can be replaceable each other, i.e., they are indistinguishable, the auction is called multi-unit auction. Multi-unit combinatorial auction is the case when some items are indistinguishable in a combinatorial auction [4]. Multi-unit combinatorial auction can be applied to electricity allocation problems, environmental exhausting right assignment problems, and other problems that consider quantitative or countable items in them [27].

Since winner determination on a combinatorial auction is a kind of combinatorial optimization problem, there are really many approaches for calculating approximate allocation on combinatorial auctions. Zurel and Nisan [36] proposed a very high performance approximate winner determination algorithm for combinatorial auctions. The main idea is a combination of approximated positive linear program algorithm for determining initial allocation and stepwise random updates of allocations. Also Lehmann et al., proposed a very practical approximation approach [26]. There have been a lot of works on obtaining optimal solutions for winner determination in combinatorial auctions [34]. For example, CABOB [30] and CASS [6] have been proposed by aiming to get the optimal allocations. They also mentioned that their approach can also be used for obtaining approximate allocations.

In [17–19], it has been shown that the presented hill-climbing approach outperforms SA [17], SAT-based algorithms [20], LP-based heuristic approximation approach [36], and a recent LP solver product in the setting when an auction has a massively large number of bids but the given time constraint is very hard.

2.2 Winner Approximation and Pricing

It is crucial for an auction mechanism to have a proper pricing mechanism to incentivize bidders to reveal their true valuations of items appropriately [4]. In VCG (Vickery-Clarke-Groves) mechanism [33,34], prices that winners will pay will be given as follows [29]. A payment p_n for a winner n is calculated by

$$p_n = \alpha_n - \sum_{i \neq n, S \subseteq M} v_i(S) x_i(S)$$

Here, the right part of the right side of the equation denotes the sum of all bidding prices of won bids, excluding the bids that are placed by the bidder n. The left part of the right side of the equation, α_n is defined by

$$\alpha_n = max \sum_{i \neq n, S \subseteq M} v_i(S) x_i(S)$$

for a feasible allocation $X \ni x_i(S)$. This means that the α_n is the sum of all bidding prices of won bids when the allocation is determined as if a bidder n does not place any bids for the auction.

In [29], Nisan et al. showed that optimal allocations should be used for VCG-based pricing to make the auction incentive compatible (i.e., revealing true valuations is the best strategy for each bidders). Also, Lehmann et al. showed that VCG-based pricing with approximate winner determination will not make the auction incentive compatible even when it is assumed that all bidders are single-minded(i.e., each bidder can only place single bid at each auction) [26].

To overcome this issue, Lehmann et al. prepared a special pricing mechanism that can be applied for their approximate greedy winner determination [26]. However, since this pricing mechanism deeply depends on their allocation algorithm, it cannot be easily applied to other approximation allocation algorithms. Also the mechanism is incentive compatible only when single-minded bidders are assumed [26].

The main problem in which VCG-based pricing is applied to approximation allocation of items is that there are the cases that: (1) the price for a won bid is rather higher than the bid price, and (2) the price for a won bid is less than zero, it means the bidder will win the items and also will obtain some money rather than paying for it [29]. In the situation of (1), it breaks individual rationality (i.e., the one will not pay a higher price than the placed bid when the one won the bundle of items). Also the situation of (2) is not preferable for both auctioneers and sellers. Although other approaches have been proposed (e.g., [24]), they could work only on a limited situation.

2.3 Approximation for Multi-unit Combinatorial Auctions

Here, as a starting point of overviewing this topic, I would briefly describe the approximation allocation algorithm for multi-unit combinatorial auctions proposed in [7]. The inputs are *Alloc*, L, and *Stocks*. L is the bid list of an auction. *Stocks* is the list of the number of auctioned units for each distinguishable item type. *Alloc* is the initial greedy allocation of items for the bid list.

```
1: function LocalSearch(Alloc, L, Stocks)
2:   RemainBids:= L − Alloc;
3:   sortByLehmannC(RemainBids);
4:   for each b ∈ RemainBids
5:     RestStocks:=getRestStocks({b}, Stocks);
6:     AllocFromWinners:=greedyAlloc(RestStocks, Alloc);
7:     RestStocks:=
8:       getRestStocks(AllocFromWinners + {b}, RestStocks);
9:     AllocFromRest:=
10:      greedyAlloc(RestStocks, RemainBids − {b});
11:    NewAlloc:=
12:      {b} + AllocFromWinners + AllocFromRest;
13:    if price(Alloc) < price(NewAlloc) then
14:      return LocalSearch(NewAlloc,L,Stocks);
15:  end for each
16:  return Alloc
```

As explained in [7], the algorithm employs the following sub-functions. The function $sortByLehmannC(Bids)$ has an argument $Bids$. This function sorts the list of bids $Bids$ by descending order of Lehmann's weighted bid price. The results are directly stored (overwritten) to the argument $Bids$. The function $getRestStocks(Bids, Stocks)$ has two arguments: $Bids$ and $Stocks$. This function returns how many unit of items will remain after allocating the items in $Stocks$ to the list of bids $Bids$. The function $greedyAlloc(Stocks, Bids)$ has two arguments: $Stocks$ and $Bids$. This function allocates the items in $Stocks$ to the list of bids $Bids$ by using Lehmann's greedy allocation, and then the winner bids are returned as the return value. The function $price$ calculates the sum of bidding prices for bids specified in the argument.

In [17–19], I have shown that the proposed hill-climbing approach outperforms SA [17], SAT-based algorithms [20], LP-based heuristic approximation approach [36], and a recent LP solver product in the setting when an auction has a massively large number of bids but the given time constraint is very hard. Note that, those algorithms were designed for single-unit combinatorial auction problems so they cannot be applied for multi-unit problems directly. Thus, in [7], it was shown that the algorithm MHC also produces better performance than LP-based approximation approach and others even when it is applied to single-unit auction problems.

To realize a pricing mechanism that receives little effect from the winners bid prices, the following algorithm [9] has been presented.

```
1: function transformToSWPM(Alloc, L, Stocks)
2:   RemainBids:= L − Alloc;
3:   sortByLehmannC(RemainBids);
4:   clear(payment);
5:   for each b ∈ Alloc
```

6: $RestStocks$:=$getRestStocks(Alloc - \{b\}, Stocks)$;
7: $AllocForB$:=$greedyAlloc(RestStocks, RemainBids)$;
8: $NewAlloc$:=$Alloc - \{b\} + AllocForB$;
9: **if** $price(Alloc) < price(NewAlloc)$ **then**
10: **return** transformToSWPM($NewAlloc$,L,$Stocks$);
11: **else** $payment_b = price(NewAlloc) - price(Alloc - \{b\})$
12: **end for each**
13: **return** ($Alloc$,$payment$)

Here, the function *transformToSWPM* calculates the prices and associated updated winners as well, to keep a condition called *strong winner price monotonicity* [7]. The inputs are $Alloc$, L, and $Stocks$. L is the bid list of an auction. $Stocks$ is the list of the number of auctioned units for each distinguishable item type. $Alloc$ is the initial allocation of items for the bid list that is obtained by the previously defined $LocalSearch$ function.

The above algorithm computes the price to be paid for each winner bid. The payment price for a winner bid b is denoted by $payment_b$, and it's value is obtained by $price(NewAlloc) - price(Alloc - \{b\})$. When the obtained payment price is higher than the bidding price of the winner bid, the algorithm discards the winner bid and place the items to $AllocForB$. To the end, the algorithm produces modified allocations $Alloc$ and their payment prices $payment$ that satisfies budget constraints for bidders.

For simplicity of description, the above algorithm is written with single-minded bidders assumption. To extend the algorithm without the assumption can be realized by just replacing $\{b\}$ with the all bids that come from the bidder of $\{b\}$. This mechanism can actually be implemented by using dummy items that are assigned to corresponding agents [9].

2.4 Approximate Pricing with Reserve Prices

When approximately solving a combinatorial auction problem with reserve price biddings, a naive approach may produce a winner bid which includes a set of items whose prices are higher than the sum of reserve prices. For example, when we naively apply the pricing mechanism *transformToSWPM* introduced in [9], the mechanism does not satisfy *reserve price conditions* (i.e., each winner places the price which is higher than the best combination of reserve price bids for the items in the bundle), since the mechanism used Lehmann's approximation allocation [26].

Lehmann's greedy algorithm [26] is a very simple but powerful linear algorithm for winner determination in combinatorial auctions. Here, a bidder declaring $< s, a >$, with $s \subseteq M$ and $a \in \mathcal{R}_+$ will be said to put out a bid $b = < s, a >$. Two bids $b = < s, a >$ and $b' = < s', a' >$ conflict if $s \cap s' \neq \emptyset$. The greedy algorithm can be described as follows. (1) The bids are sorted by some criterion. In [26], Lehmann et al. proposed sorting list L by descending average amount per item. More generally, they proposed sorting L by a criterion of the form $a/|s|^c$

for some number c, $c \geq 0$, possibly depending on the number of items, k. (2) A greedy algorithm generates an allocation. L is the sorted list in the first phase. Walk down the list L, allocates items to bids whose items are still unallocated.

Definition Let $w_i =< s_i, a_i > \in W$ be a winner of an auction, $rp_j =< s_j a_j > \in RP$ be a reserve price bid. When a pricing mechanism satisfies *reserve price condition*, for any $w_i \in W$, the a_i is greater than the best combination of $rp_j \in RP$ for a bundle of item s_i.

Proposition *The pricing mechanism* **transformToSWPM** *introduced in [9], does not satisfy reserve price conditions.*

A sketch of proof for the above proposition has been shown in [12].

To overcome the issue shown in the previous part of this section, following algorithm *transformToSWPMRP* that addresses the issue mentioned was presented [12,14].

1: **function** transformToSWPMRP(*Alloc*, *L*, *Stocks*)
2: *RemainedBids*:= $L - Alloc$;
3: $sortByLehmannC(RemainedBids)$;
4: $clear(payment)$;
5: **for each** $b \in Alloc$
6: $RestStocks$:=$getRestStocks(Alloc - \{b\}, Stocks)$;
7: $AllocForB$:=$greedyAlloc(RestStocks, RemainedBids)$;
8: $NewAlloc$:=$Alloc - \{b\} + AllocForB$;
9: **if** $price(Alloc) < price(NewAlloc)$ **then**
10: **return** transformToSWPMRP($NewAlloc$,L,$Stocks$);
11: **else**
12: $RemainedReserveBids$:=$getReservedBids(RemainedBids)$
13: $AllocForR$:=$greedyAlloc(RestStocks, RemainedReserveBids)$;
14: $NewAllocR$:=$Alloc - \{b\} + AllocForR$;
15: **if** $price(Alloc) < price(NewAllocR)$ **then**
16: **return** transformToSWPMRP($NewAlloc$,L,$Stocks$);
17: **else** $payment_b = price(NewAlloc) - price(Alloc - \{b\})$
18: **end for each**
19: **return** (*Alloc*, *payment*)

Here, the function *getReservedBids(RemainedBids)* filters the specified set of bids to obtain only reserved bids in the set. The algorithm *transformToSWPMRP* avoids the situation that the winner bid has less value than the sum of reserve prices for the items in the bundle, when all reserve price bids are given as single-item bids, rather than combinatorial ones.

As shown in [9], since the algorithm seeks higher efficiency on each step, the algorithm is deterministic when there is an optimal allocation and the price does not diverge to infinity [14].

There have been investigations about applying multi-unit combinatorial auction-based resource allocation approach to smart-grid and dynamic grid resource allocation problem [10,11]. Here, the basic models for generating artificial resource allocation problems are borrowed from [10,11] as example of serve resource allocation problems and then used for the following numerical analysis.

3 Improving Stability for Allocation and Pricing Algorithm on Single Winner Cancellations

Here, I consider a way to have a more relaxed condition than *Strong Winner Price Monotonicity* for better stabilization of pricing and re-allocation on bid cancellations. I define a weaker condition, *Locally-strong Winner Price Monotonicity*,[1] as follows:

Definition 1 (*Locally-strong Winner Price Monotonicity: LWPM*) For any non-empty bundle $s' \subseteq S$ for a bidder $j \in N, X_j(s') = 1$ (i.e., a winner) and whose bid price is $v_j(s')$, the Lehmann's greedy allocation of bids for the bundle of items s' by bids from bidders $N' = \{i | i \in N, i \neq j\}$ that satisfies $X_{i \neq j}(s'') = 0$ (i.e., not winners) for any $s'' \supseteq s'$, always produces non-efficient allocation for the bundle s' (i.e., the total price of greedy allocation is less than $v_j(s')$).

Here, to satisfy this *Locally-strong Winner Price Monotonicity*, I modify the algorithm transformToSWPMRP and introduce an algorithm transformToLWPMRP[2] as follows:

```
1: function transformToLWPMRP(Alloc, L, Stocks)
2:   RemainedBids:= L − Alloc;
3:   sortByLehmannC(RemainedBids);
4:   clear(payment);
5:   for each b ∈ Alloc
6:     RestStocks:=getPlacedStocksInBid(b);
7:     AllocForB:=greedyAlloc(RestStocks, RemainedBids);
8:     NewAlloc:=Alloc − {b} + AllocForB;
9:     if price(Alloc) < price(NewAlloc) then
10:      return transformToLWPMRP(NewAlloc,L,Stocks);
11:    else
12:     RemainedReserveBids:=getReservedBids(RemainedBids)
13:     AllocForR:=greedyAlloc(RestStocks, RemainedReserveBids);
14:     NewAllocR:=Alloc-{b} + AllocForR;
15:     if price(Alloc) < price(NewAllocR) then
16:      return transformToLWPMRP(NewAllocR,L,Stocks);
```

[1] The idea has been initially presented in [13].

[2] The initial idea has been presented in [13]. A faster algorithm has also been proposed in [16] and analyzed in [15].

17: **else** $payment_b = price(NewAlloc) - price(Alloc - \{b\})$
18: **end for each**
19: **return** ($Alloc$,$payment$)

The algorithm transformToLWPMRP applies its pricing and respective re-allocations of winners in a similar way that the algorithm transformToSWPMRP does. The major difference between them is the conditions applied to justify and calculate price for each winner. On the algorithm transformToSWPMRP, it applied *Strong Winner Price Monotonicity*. However, the algorithm transformToLWPMRP applied *Locally Strong Winner Price Monotonicity*, instead, as both algorithms also applied *Reserve Price Condition*. For instance, the line 6 in the above algorithm is different. Here, a function $getPlacedStocksInBid(b)$ returns the bundle of items that the bid b placed to, as a set of pairs of the items and their units to be won.

Proposition 1 *The worst case computational complexity of transformToLWPMRP is $O(k^3N)$ for the number of items k and the number of bids N.*

Proof (Sketch) The worst case computational complexity of transformToLWPMRP should be equal to its base algorithm transformToSWPM, $O(k^3N)$ for the number of items k and the number of bids N, as shown in [9], since the difference of those computation is within a constant order. □

Proposition 2 *The algorithm transformToLWPMRP is decidable.*

Proof (Sketch) The algorithm transformToLWPMRP can be said **decidable** since, as mentioned in [9], each re-allocation of goods in the algorithm will not decrease the total number of winners, the number of items to be allocated should be a discrete value, and finite bids are included in an auction. □

The two algorithms, transformToLWPMRP and transformToSWPMRP were compared to see the differences on its allocation stability over pricing in the condition that any single winner has been cancelled on each auction. Here, the comparison has been done on how many winners have been removed from winners due to such a single winner cancellation had been made and re-pricing had been applied. The auction problems that were used are the same datasets used in [14], one of which includes 1000 bidders and 100 percent of electricity production ratio, with 515024 bids and 67392 reserve-price bids, and another of which includes 3566 bidders and 100 percent of electricity production ratio, with 1710967 bids and 238584 reserve-price bids.

We observed that the both algorithms produced the same number of winners for those auction problems. Table 1 shows the comparison result on the case of 1000 bidders. Here, we would compare how those winners will be shifted when single winner has been cancelled and the respective pricing and allocation adjustment algorithm were applied. In this case, we can observe that the algorithm transformToSWPMRP produced more winners that had been shifted to be

Table 1. Winner stability on Single Winner Cancellation Scenario (1000 bidders) [13]

	transformToLWPMRP	transformToSWPMRP
Winners	898	898
Changed(total)	750	857
Changed(average)	0.835	0.954

Table 2. Winner stability on Single Winner Cancellation Scenario (3566 bidders)

	transformToLWPMRP	transformToSWPMRP
Winners	3210	3210
Changed(total)	2684	3012
Changed(average)	0.836	0.938

losers than that in the algorithm transformToLWPMRP. On the algorithm transformToLWPMRP, it produced approximately 87.5 percent[3] of such winner-shifts compared with transformToSWPMRP.

Table 2 shows the comparison results on the case of 3566 bidders. Here, we can see that the very similar results have been obtained. In this case, the algorithm transformToSWPMRP produced slightly less percentage of winners that have been altered by the cancellation of a winner, when compared to the case of 1000 bidders shown in the Table 1. In contrast, the algorithm transformToLWPMRP produced almost equal percentage of winners that have been altered by the cancellation of a winner, compared to the case of 1000 bidders shown in the Table 1. Although it deeply depends on the problem structure, when a very similar scenario has been applied, relatively a smaller percentage of winners will be altered by cancellation of single winner in an auction when the number of bidders is larger.

Then, the effect of giving reserve prices for the items in this scenario has also been evaluated. Table 3 shows a result without any bids that represented reserve prices for the items. In this case, we can see there is no difference between transformToSWPMRP and transformToLWPMRP. An interesting point of this result is that, giving proper reserve prices to the items could help the mechanism allocate items to the buyers in a more stable manner, although it should deeply depend on the structure of the problem.

Even in those cases, average number of affected winners on each single winner cancellation was below 1 winner. In those cases, average differences of winners on each auction was below 0.1 percent. Further analysis should be done to show how those values are feasible enough in more comprehensive scenarios.

[3] Decreased from 857 to 750.

Table 3. Winner stability on Single Winner Cancellation Scenario (3566 bidders/no reserves)

	transformToLWPMRP	transformToSWPMRP
Winners	3564	3564
Changed (total)	3192	3192
Changed (average)	0.896	0.896

4 Conclusions

In this paper, I discussed about the *stability* of pricing and allocation of resources using an approximate auction which has VCG-like pricing mechanism when cancellation of winner bid(s) after its winner determination is considered. An idea about more *stable* approximate pricing mechanism against cancellation of a winner after its winner determination was presented. In there, a single-unit non-combinatorial reserve price biddings on multi-unit combinatorial auction could also be employed.

The pricing algorithm employed an approximate allocation and pricing algorithm that is capable of handling multi-unit auctions with reserve price biddings, as its predecessors. Those auctions can be applied to a situation when we consider an allocation of electricity, considering electricity generation costs on the power suppliers in more realistic configurations, i.e., some bidders might be untrustful in their ability.

As shown in the experimental results, the algorithm effectively produced approximation allocations that are necessary in the pricing phase, as well as yielding better stability in the case of single-winner cancellation. It also behaves as an approximation of VCG(Vickrey-Clarke-Groves) mechanism satisfying budget balance condition and bidders' individual rationality without enforcing the single-minded bidders assumption.

Although this paper did not give a clear comparison in the sense of performance in computation costs, since its computational complexity is comparable (i.e., in its worst case, equal) to its predecessors, the proposed mechanism could be applied for a large-scale auction problem to overcome computational intractability on naive approaches.

In this paper, the presented analysis has been done only for a limited number of allocation scenarios for electricity usage in industries, although the mechanism and the algorithm themselves could be applied to other problems such as bandwidth and channel allocation problem in wireless networks [28]. Applying the mechanism and the algorithm to such problems is one of our future work. Also further analysis on theoretical characteristics of the mechanism is future work.

Acknowledgements. The work was partly supported by Grants-in-Aid for Young Scientists(B) 22700142, Grants-in-Aid for Challenging Exploratory Research 26540162, and JST CREST JPMJCR15E1.

References

1. An, B., Lesser, V., Irwin, D., Zink, M.: Automated negotiation with decommitment for dynamic resource allocation in cloud computing. In: Proceedings International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), vol. 1, pp. 981–988 (2010)
2. Cavallo, R.: Incentive compatible two-tiered resource allocation without money. In: Proceedings the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014). pp. 1313–1320 (2014)
3. Constantin, F., Feldman, J., Muthukrishnan, S., Pal, M.: An online mechanism for ad slot reservations with cancellations. In: Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA 2009). pp. 1265–1274 (2009)
4. Cramton, P., Shoham, Y., Steinberg, R.: Combinatorial Auctions. The MIT Press (2006)
5. Edelman, B., Ostorvsky, M., Schwarz, M.: Internet advertising and the generalized second price auction: selling billions of dollars worth of keywords. Am. Econ. Rev. **97**(1), 242–259 (2007)
6. Fujishima, Y., Leyton-Brown, K., Shoham, Y.: Taming the computational complexity of combinatorial auctions: optimal and approximate approarches. In: Proceeding of the 16th International Joint Conference on Artificial Intelligence (IJCAI 1999). pp. 548–553 (1999)
7. Fukuta, N.: Toward a VCG-like approximate mechanism for large-scale multi-unit combinatorial auctions. In: Proceeding IEEE/ACM/WIC International Conference on Intelligent Agent Technology(IAT 2011). pp. 317–322 (2011)
8. Fukuta, N.: A mobile agent approach for p2p-based semantic file retrieval. J. Inf Process. **20**(3), 607–613 (2012)
9. Fukuta, N.: An approach to VCG-like approximate allocation and pricing for large-scale multi-unit combinatorial auctions. J. Inf. Process. **21**(1), 9–15 (2013)
10. Fukuta, N.: A market-based agent-mediated resource control framework for middle-scale smart grids. In: Proceeding the 2013 IEEE/WIC/ACM International Conference on Web intelligence/Intelligent Agent Technology (WI-IAT 2013). vol. 3, pp. 292–293 (2013)
11. Fukuta, N.: A preliminary implementation on pricing mechanisms for electric resource control markets. In: Proceedings the 3rd International Workshop on Knowledge and Service Technology for Life, Environment, and Sustainability (KASTLES 2013). pp. 338–342 (2013)
12. Fukuta, N.: An approximation approach for large-scale multi-unit combinatorial auctions with reserve-price biddings. In: Proceeding the 2nd International Conference on Smart Computing and Artificial Intelligence(ICSCAI 2014). pp. 487–492 (2014)
13. Fukuta, N.: A preliminary analysis of allocation stability on approximation-based pricing for multi-unit combinatorial auctions–a single-winner cancellation scenario. In: Proceedings KICSS 2015 International Workshop on Collective Intelligence and Crowd/Social Computing. pp. 236–246. Phuket, Thailand (2015)
14. Fukuta, N.: Toward efficient approximation for large-scale multi-unit combinatorial auctions with reserve-price bidding. IEICE Trans. Inf. Syst. J98-D(6), 948–961 (Jun 2015), (In Japanese.)
15. Fukuta, N.: A pre-processing approach for fast and stable allocations on approximation-based pricing for multi-unit combinatorial auctions. Inf. Eng. Express **2**(4), 21–30 (2016)

16. Fukuta, N.: Toward fast approximation of stable allocation and pricing on combinatorial auctions. In: Proceedings of 1st IEEE International Conference on Agents(ICA 2016). pp. 74–77 (Sep 2016)
17. Fukuta, N., Ito, T.: Towards better approximation of winner determination for combinatorial auctions with large number of bids. In: Proceedings of The 2006 WIC/IEEE/ACM International Conference on Intelligent Agent Technology(IAT 2006). pp. 618–621 (2006)
18. Fukuta, N., Ito, T.: Fine-grained efficient resource allocation using approximated combinatorial auctions-a parallel greedy winner approximation for large-scale problems. Web Intel. Agent Syst. Int. J. **7**(1), 43–63 (2009)
19. Fukuta, N., Ito, T.: An experimental analysis of biased parallel greedy approximation for combinatorial auctions. Int. J. Intel. Inf. Database Syst. **4**(5), 487–508 (2010)
20. Hoos, H.H., Boutilier, C.: Solving combinatorial auctions using stochastic local search. In: Proceedings of 17th National Conference on Artificial Intelligence (AAAI 2000). pp. 22–29 (2000)
21. Ishikawa, T., Fukuta, N.: A prototype system for federated cloud-based resource allocation by automated negotiations using strategy changes. In: Proceedings the 6th International Workshop on Agent-based Complex Automated Negotiations (ACAN 2013) (2013)
22. Ito, T., Imi, Y., Ito, T., Hideshima, E.: COLLAGREE: A faciliator-mediated large-scale consensus support system. In: Proceedings of the International Conference on Collective Intelligence (2014)
23. Ito, T., Parkes, D.C.: Instantiating the contingent bids model of truthful interdependent value auctions. In: Proceedings of the International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS 2006). pp. 1151–1158 (2006)
24. Krysta, P., orestis Teleis, Ventre, C.: Mechanisms for multi-unit combinatorial auctions with a few distinct goods. In: Proceedings the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013). pp. 691–698 (2013)
25. Lavi, R., Nisan, N.: Competitive analysis of incentive compatible on-line auctions. In: Proceedings of the 2nd ACM conference on Electronic commerce(EC 2000). pp. 233–241. ACM (2000)
26. Lehmann, D., O'Callaghan, L.I., Shoham, Y.: Truth revelation in rapid, approximately efficient combinatorial auctions. J. ACM **49**, 577–602 (2002)
27. Leyton-Brown, K., Pearson, M., Shoham, Y.: Towards a universal test suite for combinatorial auction algorithms. In: Proceedings of ACM Conference on Electronic Commerce (EC 2000). pp. 66–76 (2000)
28. McMillan, J.: Selling spectrum rights. J. Econ. Perspect. **8**, 145–162 (1994)
29. Nisan, N., Ronen, A.: Computationally feasible VCG mechanisms. In: Proceeding of ACM Conference on Electronic Commerce. pp. 242–252 (2000)
30. Sandholm, T., Suri, S., Gilpin, A., Levine, D.: Cabob: a fast optimal algorithm for winner determination in combinatorial auctions. Manag. Sci. **51**(3), 374–390 (2005)
31. Shoham, Y., Leyton-Brown, K.: Multiagent Systems: algorithmic, game-theoretic, and logical foundations. Cambridge University Press (2008)
32. Todo, T., Iwasaki, A., Yokoo, M., Sakurai, Y.: Characterizing false-name-proof allocation rules in combinatorial auctions. In: Proceedings 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009) (2009)

33. Vickrey, W.: Counterspeculation, auctions, and competitive sealed tenders. J. Finance **16**(1), 8–37 (1961)
34. de Vries, S., Vohra, R.V.: Combinatorial auctions: a survey. Informs J. Comput. **15**(3), 284–309 (2003)
35. Yokoo, M.: The characterization of strategy/false-name proof combinatorial auction protocols: Price-oriented, rationing-free protocol. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence. pp. 733–739 (2003)
36. Zurel, E., Nisan, N.: An efficient approximate allocation algorithm for combinatorial auctions. In: Proceeding of the 3rd ACM Conference on Electronic Commerce (EC 2001). pp. 125–136 (2001)

Revenue Improvement Mechanisms Based on Advertisement's Values and Layouts

Satoshi Takahashi[1(✉)] and Tokuro Matsuo[2]

[1] The University of Electro-Communications, Chofu, Japan
stakahashi@uec.ac.jp
[2] Advanced Institute of Industrial Technology, Tokyo, Japan
tokuro@tokuro.net

Abstract. We study an advertisement transaction market which is considering a value of advertisement in addition to monetary value. This paper proposes a novel model for advertisement transactions which employs two stages auctions. To improve an advertisement value for advertisers, it is important to consider not only advertisement sizes and its allocations but also its combinations for the advertisers' marketing strategies. The two stages auction aims to improve the advertisement value and seller's revenue. The concept of our model is that the advertisers can consider own marketing strategies for the auction. The first auction winner can choose some preferred other advertisers to make a second auction, and also he/she can reject inconvenient advertisers. The first auction winner can consider the marketing strategy by making a convenient bidders set in the second auction.

Keywords: Advertisement auction · Combinatorial optimization
Mechanism design · Market strategy · Packing problem

1 Introduction

Recently, the Internet advertisement market is growing up significantly, since the Internet becomes one of the primal media in real world. Particularly, a keyword auction market, which is managed by some search engine company, is the best successful advertisement market in the advertisement history. People can get some advertisement information when he/she searches keywords associated with the advertisement. Almost Internet advertising mechanisms are considering a value of advertisement in addition to monetary value. For example, [1] proposes the generalized second-price auction, employed by Google Adwords [2], which is focus on the credit history of the advertisement. In this paper, we focus on a value of advertisement on the paper media advertisement market. The advertisement value, in a business marketing context, consists of the following elements:

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_22

(a) *Discovery*: It is important for customers to find the advertisement on the media. This factor contains a slogan and picture.
(b) *Perception*: Customers get impression about the advertisement. This factor contains a coloring and a font of contents.
(c) *Recognition*: It is important for customer to understand the contents of advertisement. This factor contains a readability and a number of words.
(d) *Reading time*: Customer spends own time to reading the advertisement. This factor contains a size of the advertisement and medias.
(e) *Unforgettability*: Customer memorizes the advertisement. It is important to distribute "Buzz Word" about the advertisement to others.

These advertisement values is changed by not only the distance between advertisements but also the combination and layout. The layout consists of (1) adjacent two advertisements, (2) two advertisements place on the same page, (3) two advertisements place on the different page, (4) two advertisements place on the different media. For example, we consider two car factories' advertisements adjacent each other. In this case, the value (a) grows and to increasing the probability of finding the advertisement encourages the customers to see the advertisements. Other example, there is a golf advertisement adjacent to the car factory's advertisement. In this case, the combination of these advertisement encourages the customer to growing the value (b) and (c). If the car advertisement is One box car or RV type, the car advertisement is more desirable for golf player customers. On the other hand, if the clothes iron's advertisement adjacent to the car advertisement, these relationship is not deep. In this case, the probability of making the value (a) to (e) is low and the advertiser loses the business chance. It is important problem to consider the layout and allocation of the advertisements which changes the values of advertisements for not only the advertisers but also the medias.

This paper discusses a novel model for advertisement transactions which employs two stages auctions. The two stages auction aims to improve the advertisement value and seller's revenue. The concept of our model is that the advertisers can consider own marketing strategies for the auction. In the advertisement auction, the value of advertisement is not decided by own advertisement but it is decided by combination and layout of other advertisement. This problem has been discussed in the customer mix problem which is general fact in the real world. Our research's novelty is that we model these phenomena as the advertisement auction model. The first auction winner can choose some preferred other advertisers to make a second auction, and also he/she can reject inconvenient advertisers. The first auction winner can consider the marketing strategy by making a convenient bidders set in the second auction.

2 Preliminaries

2.1 Customers Mix Problem

In the advertisement space allocation, advisers mixing is one of important issues. Advertisement space business is one of service business and this kind of business

exists the problem on the factor of unsuccessful service [3–5]. It is very easy to explain the actual situation where the service quality reduces. For example, when people go to have a special dinner in the restaurant in their wedding anniversary, most of people must choose an appropriate restaurants like the restaurant serving a full-course meal. If they choose the fast food restaurant, the must be disappointed because other customers visit there on their regular diet. As the result, they do not have a happy anniversary time due to noisy and non-special meals. In the formal restaurant, because most of other customers do not have their daily dinner, the atmosphere and situation is better than the fast food restaurant. If all of customers use their anniversary, each of them makes his/her invisible spaces in the restaurant and enjoys his/her anniversary.

2.2 Traditional Advertisement Allocation on the Media

Traditional advertisement market is one on one negotiation transaction, it is not competitive market. On the other hand, the internet advertisement on the search engine company, for example Google [2], is when users search for some words on the search engine, an advertisement related with the searched keywords is displayed with result of search [1]. In this transaction, the auctioneer records the credit history of transaction results and apply it to the result of allocation. Compared with traditional transaction, the internet advertisement auction considers the value of the advertisement without the monetary value.

As related work of our research, there is a GlobalAd system which advertisement transaction system on the news paper [6,7]. We show the system flow of GlobalAd system on the Fig. 1. The GlobalAd system employs an auction mechanism for the advertisement determination. Also, they discuss an interdependence value model. In general auction based advertisement model, the advertisement value which includes monetary value is independent. However these papers are discussing about each advertisement value influences other advertisement. This concept is the same as our proposing model. However, the auction is one shot, they did not discuss a revenue improvement.

3 Revenue and Advertisement Value Improvement Auction

We propose a transaction model that makes additional revenue by layout of the advertisement. The additional revenue is that the first winner of the page designates other advertisers to participate second auction. This entitlement is that designated advertisers can rebid to get the adjacent or the same page of the first winner's advertisement. Also for the first winner can reject the advertiser from second auction if the first winner pays the bid value of rejected advertiser. For the auctioneer the balance of first auction and second auction is the additional revenue. The protocol is the follows.

1. The advertisers bid for the advertisement space according to the size and page. For example, we consider 16 pages, the advertisers bid as the Table 1.

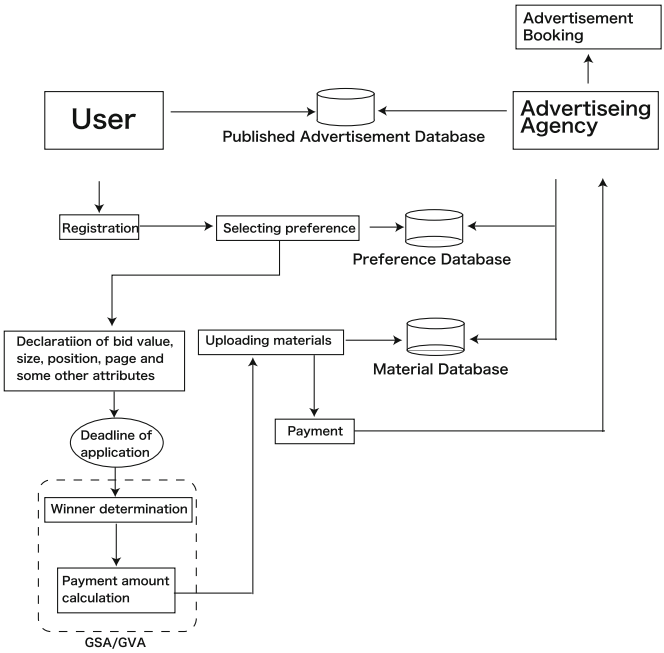


Fig. 1. System flow of GlobalAd system

2. The auctioneer calculates the allocation which maximizes the revenue.
3. The first winner of each page is the fixed advertiser. Then the auctioneer announce the result only to the fixed advertiser. Other advertisers cannot get any information about the result of the auction.
4. The fixed advertiser announce the rejected advertisement to the auctioneer if he/she want.
5. The fixed advertiser can designate other advertisers to the second auction.
6. The designated advertisers bid if they wish to bid. The Table 2 is an example of second auction's bid.
7. After the second auction, the auctioneer calculates final revenue and allocation of the advertisements.

Figure 2 illustrates the auction protocol. The 1st winner of the first auction can invite and remove some advertisers based on the result of the first auction.

We show an example of proposed protocol. Suppose that the auctioneer prepare 2 pages of the advertisement space on the media. We define three types of space size of advertisement. For Large size, Medium size is half of Large, Small is quarter of Large. Assume that there are 20 advertisers. We can allocates two Large size advertisements for each page. Hence eight Small size advertisements can place on the page. The Table 3 is a bid table of each bidder. On this example, we denote the advertiser by Ai. Also each size of jth page denote PjL, PjM and PjS. For example, the Small size advertisement of the second page is $P2S$.

Table 1. Combination of bidding

	Large	Mid. Large	Medium	Mid. Small	Small
Page 1	$50,000	$40,000	$25,000	$20,000	$10,000
Page 2	$30,000	$20,000	$15,000	$10,000	$5,000
⋮	⋮	⋮	⋮	⋮	⋮
Page 16	$4,000	$3,500	$2,000	$1,500	$1,000

Table 2. Bid table of 2nd auction

Page 2	Large	Mid. Large	Medium	Mid. Small	Small
Adjacent	$50,000	$30,000	$18,000	$15,000	$10,000
Very near	$35,000	$25,000	$17,000	$12,000	$9,000
Near	$32,000	$22,000	$16,000	$11,000	$5,000
Same page	$30,000	$20,000	$15,000	$10,000	$5,000

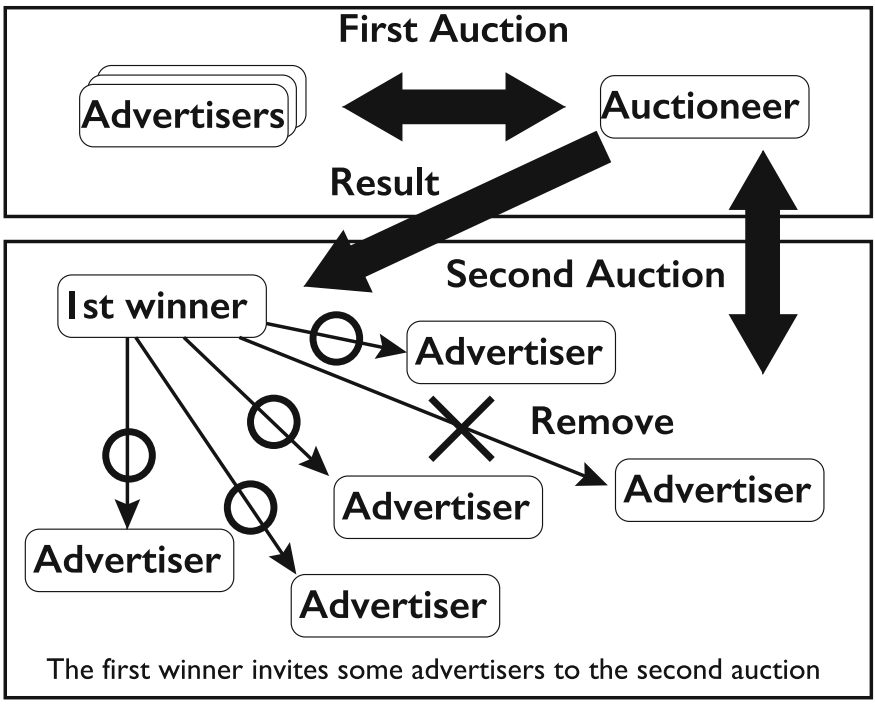


Fig. 2. Proposed auction protocol

Hence the optimal combination is $P1L$ to $A8$, $P1L$ to $A19$, $P2L$ to $A14$, $P2M$ to $A11$ and $P2S$ to $A18$ and $A4$. The fixed advertiser of page 1 is $A8$ and page 2 is $A14$. Then the auctioneer gets the revenue 4702. The fixed advertiser $A8$ rejects $A19$ and pays $A19$'s bid value and $A14$ designates $A4$, $A11$, $A18$ and $A20$. On the page 2, designated bidders $A11$, $A18$ and $A20$ bid for the second auction. This bid table is shown in the Table 4. In this case, the optimal allocation is $P1L$ to $A8$, $P1M$ to $A10$, $P1M$ to $A20$, $P2L$ to $A14$, $P2M$ to $A11$, and $P2S$ to $A18$ and $A4$. The final revenue is 5925. The revenue is increase 1223 from the first auction. Also the advertisement value is growing up since the fixed advertiser can operate own marketing strategy. By our protocol, qualitative individual rationality and Palate efficiency can be improved.

Table 3. Bid table

	P1L	P1M	P1S	P2L	P2M	P2S
A1	623	429	234	602	426	60
A2	618	504	269	1035	512	139
A3	967	562	167	618	315	283
A4	674	322	185	846	491	296
A5	1000	428	135	1005	438	205
A6	800	379	178	715	490	204
A7	871	406	101	1105	423	86
A8	1180	347	144	951	485	291
A9	900	486	131	1052	400	257
A10	604	572	243	718	397	91
A11	692	507	107	728	591	191
A12	883	362	249	980	381	268
A13	678	315	153	701	453	51
A14	962	399	187	1172	336	280
A15	985	434	248	1151	579	267
A16	1130	419	207	744	407	245
A17	956	528	73	969	352	97
A18	1149	383	253	790	339	293
A19	1170	426	144	938	544	52
A20	761	585	265	689	429	101

3.1 Mathematical Model

We describe a mathematical model of above discussion. Let N be a set of bidders (advertisers) and S be a set of advertisement spaces. For each $s \in S$, $\ell_s \in \Re$

Table 4. Bid values of designated bidders

	P2L	P2M	P2S
A11	760	650	250
A18	790	500	300
A20	1000	550	240

is a size. Suppose that $S_p \subseteq S$ is a subset of advertisement spaces on a page p, $\sum_{s \in S_p} \ell_s = 1$. For each bidder $i \in N$ has a preference $\theta_i : N \to \{0, 0.5, 1\}$ for other bidders. If $\theta_i(j) = 1$, the bidder i wish to place own advertisement on the same page of the bidder j's advertisement. If $\theta_i(j) = 0$, the bidder i wants to reject the bidder j. If $\theta_i(j) = 0.5$, the bidder i's preference is neutral. Each bidder $i \in N$ bids a value $v_i(s)$ for each space $s \in S$. Then we assume the following monotonicity for v_i:

$$\forall i \in N \ (\forall \hat{s}, \tilde{s} \in S \text{ such that } \ell_{\hat{s}} \leq \ell_{\tilde{s}})(v_i(\hat{s}) \leq v_i(\tilde{s})).$$

Also we assume $v_i(0) = 0$. Hence, the auctioneer solve the following optimization problem [8,9] and decide the allocation.

$$\begin{array}{ll} \text{Maximize} & \displaystyle\sum_{i \in N} \sum_{s \in S} v_i(s) \cdot x_{is} \\ \text{Subject to} & \displaystyle\sum_{i \in N} \sum_{s \in S} x_{is} \leq 1 \\ & \displaystyle\sum_{i \in N} x_{is} \leq 1, \qquad \forall s \in S \\ & \displaystyle\sum_{i \in S} \sum_{s \in S_p} x_{is} \leq 1.0, \forall p \\ & x_{is} \in \{0, 1\}, \qquad \forall i \in N, \ \forall s \in S. \end{array} \tag{1}$$

We define a payment of the bidder i p_i as follows:

$$p_i = \begin{cases} v_i(s) & (x_{is} = 1) \\ 0 & (\text{otherwise}) \end{cases}.$$

Suppose that $\boldsymbol{x}^*$ is an optimal solution of the problem (1). Then we define a first winner of each page p as follows:

$$i_p^* = \text{argmax}_{i \in N, s \in S_p} \{v_i(s) \cdot x_{is}^*\}.$$

We describe a proposed protocol by mathematical model.

1. Each bidder $i \in N$ bids a value $v_i(s)$ for each advertisement space $s \in S$.
2. The auctioneer solves the problem (1) and decides the allocation which maximizes the revenue.

3. For each page p the auctioneer decides the first winner i_p^* and designates the fixed advertiser. The auctioneer announces the other bidders' information and the result of first auction $\boldsymbol{x}^*(p)$.
4. The fixed advertiser i_p^* announces the bidders to the auctioneer if the fixed advertiser wants. Then the auctioneer inform the bid values of rejected bidders to the fixed advertiser i_p^*. Let $\hat{N}$ be a set of rejected bidders.
5. The fixed advertiser i_p^* announces designated bidders $\tilde{N}$ who can bid the same page or adjacent space. $\tilde{N}$ is decided by the preference $\theta_{i_p^*}$.
6. Each designated bidder $i \in \tilde{N}$ bids if they wish. In this case, they have to bid $\tilde{v}_i(s)(\geq v_i(s))$. Then we use the function $\gamma(\theta_i)$ denoted by

$$\tilde{v}_i(s) = v_i(s) + \gamma(\theta_i)$$

7. The auctioneer solves the problem (1) based on a bidding set $\{v_i(s) \mid N \setminus \{i_p^*\} \setminus \hat{N} \setminus \tilde{N}\} \cup \{\tilde{v}_i(s) \mid i \in \tilde{N}\}$ and calculates the new allocation $\tilde{\boldsymbol{x}}^*$ and revenue.
8. The auctioneer announce the payment to each bidder.

For rejected bidders set $\hat{N}$, we define a rejected cost of the fixed advertiser i_p^* as follows.

$$\text{cost}_{i_p^*} = \sum_{i \in \hat{N}} \sum_{s \in S_p} v_i(s) \cdot x_{is}^*. \tag{2}$$

According to the Eq. (2), the final revenue of the auctioneer is

$$\text{revenue} = \sum_{i \in \tilde{N}} \sum_{s \in S_p} \tilde{v}_i(s) \cdot x_{is}^* + \sum_p (\text{cost}_{i_p^*} + \sum_{s \in S_p} v_{i_p^*}(s) \cdot x_{i_p^*,s}^*). \tag{3}$$

We can evaluate the improvement of the revenue as follows:

$$\text{revenue\_improvement} = \text{revenue} - \sum_{i \in N} \sum_{s \in S} v_i(s) \cdot x_{is}^*.$$

Since the minimum bid value of the second auction is at least the first auction's bid value, the revenue_improvement is nonnegative. we cannot specify the monetary Palate improvement on this model, however, for the fixed advertiser this protocol investigates the Palate improvement of other advertisement value, since the fixed advertiser operates the advertisement strategy.

4 Discussion

4.1 Possibility of Shill Bids

In the proposed protocol, selected bidders may shills of first winner. The shill bids for auctions are discussed in many papers [10–13]. The following Table 5 shows an example of auction where shills exist. Bidder 1 becomes a first winner of the auction. Then, she chooses bidders 2, 3, and 4. In second auction, each

designated bidder bids the same amount of price in the advertisement space B, C, and D. The revenue of auctioneer is 1,900 USD. If the bidder 5 becomes a winner and select bidders 6, 7, and 8 in second stage auction, the auctioneer can get totally 2,400 USD revenues at least even though these designated bidders bid same valuation of first stage auction. This means the auctioneer loses its potential 500 USD revenues.

Table 5. Bidding table

	A	B	C	D
Bidder 1	1000	100	100	100
Bidder 2	200	200	200	200
Bidder 3	300	300	300	300
Bidder 4	400	400	400	400
Bidder 5	900	800	700	600
Bidder 6	800	700	600	500
Bidder 7	700	600	500	400
Bidder 8	600	500	400	300

To solve the problem, the proposed mechanism should be improved to avoid this shill bids issue. Thus, we propose an improved protocol shown as follows.

1. The auctioneer declares the several candidates of winners based on bid value. Each winner is not informed as the final winner.
2. Each candidate of winners removes unappropriated bidders and selects bidders in the second stage auction.
3. Each designated bidder bid the valuation in second stage auction.
4. Auctioneer chooses the final winner group based on total revenue of the auction.

Using above bid table, bidders 1 and 5 are candidate of winner if the auctioneer defined. Because bidders 2, 3, and 4 are shills for bidder 1, bidder 1 selects them as the designated bidders in the second stage auction. On the other hands, we consider a case where bidder 5 chooses all bidders. Even though each bidder does not increase the bid value from the first stage auction in the second stage, the winners for advertisement space B, C, and D become respectively bidders 6, 7, and 4. In this case, total revenue is $2,500.

As the possibility, bidder 1 and 5 may be untruthful bidders. If these bidders are created by one advertiser, the bidder 5 does not have any incentive to choose bidders 6, 7, and 8. For example, bidder 5 select 1, 2, and 3 and designated bidders do not change the valuation in the second stage auction, the revenue can be calculated as $1,500. However, in the case where bidder 5 selects three bidders from bidders 1, 2, 3 and 4, it is easy for auctioneer to know these bidders 1 to 5 are created by one advertisers. When this kind of situation is detected,

auctioneer can add other candidate of winners from bidders 6, 7, or 8. If the auctioneer selects bidder 6 as the additional candidate of the winner and she selects all bidders in second stage auction, the winners for advertisement space B, C, and D become respectively bidders 5, 7, and 4. The total revenue is calculated as $2,600.

It is known that there is no auction protocols which satisfy Pareto efficiency, incentive compatibility, individual rationality without Vickrey-Clarke-Groves (VCG) mechanism [13,14]. In particular [13] has been saying the impossibility theorem. (1) the VCG mechanism, which is strategyproofness and Pareto efficient when there exists no shill bid, is not false-name-proof; (2) there exists no false-name-proof combinatorial auction protocol that satisfies Pareto efficiency; (3) one sufficient condition where the VCG mechanism is false-name-proof is identified, i.e. the concavity of a surplus function over bidders. Also in [15], they said the auction protocol is satisfied false-name-proof if and only if its allocation rule satisfies weak-monotonicity and sub-additivity. These results are only for VCG mechanism and combinatorial auctions. However, these auctions are difficult for participants in real auction markets. It is important to specify the conditions or properties for simply auction mechanism as same as the VCG mechanism and the combinatorial auctions.

5 Conclusion

In this research, we proposed the revenue improvement auction model and protocol in which the buyer designates other buyers to improves the auctioneer's revenue in the advertisement auction. This paper discussed a novel model for advertisement transactions which employs two stages auctions. In the advertisement auction, the value of advertisement is not decided by own advertisement but it is decided by combination and layout of other advertisement. This problem has been discussed in the customer mix problem which is general fact in the real world. Our research's novelty is that we model these phenomena as the advertisement auction model. The first auction winner can choose some preferred other advertisers to make a second auction, and also he/she can reject inconvenient advertisers. The first auction winner can consider the marketing strategy by making a convenient bidders set in the second auction. Also we considered a detecting method for shill biddings in our auction. Furthermore we invested a successful probability of shilled bids in the auction by computational simulations.

There are three future work for our research. Since the advertisement allocation problem employed by our research is formulated as the knapsack problem, it is hard to solve efficiently. We have to consider an effective algorithm for solving our allocation problem. The second is, from the mechanism design, to theoretical solving for our protocol's economical properties. We need to specify the

allocation efficiency, strategyproofness and individual rationality. Particularly, since the reject cost is not small amount for the advertisers, the advertisers not necessarily rejecting if he/she has the preference. In this case we have to analyze based on the probability. The strategy analysis of the cost is a bid future work. The last future work is the evaluation by human experiment. It is necessary to specify a human behavior in this auction model for applying this to real world.

Acknowledgements. This research is supported by Grant-in-Aid for Scientific Research (KAKENHI 26870200, 15H02972).

References

1. Edelman, B., Ostrovsky, M., Schwarz, M.: Internet advertising and the generalized second price auction: selling billions of dollars worth of keywords. Am. Econ. Rev. **9**(1), 242–259 (2007)
2. http://google.com
3. Arshad, A.M., Su, Q.: Interlinking service delivery innovation and service quality: a conceptual framework. J. Appl. Bus. Res. **31**(5), 1807–1822 (2015)
4. Gan, C., Cohen, D., Clemes, M.: A Survey of Customer Retention in the New Zealand Banking Industry. Banks and Banks Syst. **1**(4), 83–99 (2006)
5. Sridhar, M.S.: Customer participation in service production and delivery system. Libr. Sci. Slant Doc. Inf. Stud. **35**(3), 157–163 (1998)
6. Takahashi, S., Matsuo, T., LEE, R.Y.: Advertisement pricing and layout mechanism in GlobalAd. J. Inf. **16**(2B), 1546–1554 (2013)
7. Takahashi, S., Matsuo, T., Lee, R.Y., Ito, T.: A co-dependent value-based mechanism for the internet advertisement auction. In: Cranefield, S., Song, I., (eds.) Agent Based Simulation for a Sustainable Society and Multi-agent Smart Computing LNAI 7580, pp. 64–77. Springer (2011)
8. Schrijver, A.: Combinatorial Optimization: polyhedra and efficiency. Springer-Verlag, Berlin Heidelberg (2003)
9. Takahashi, S., Shigeno, M.: Approximation algorithms for a winner determination problem of single-item multi-unit auctions. JSIAM lett. **3**, 29–32 (2011)
10. Conitzer, V., Sandholm, T.: Failures of the VCG mechanism in combinatorial auctions and exchanges. In: Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 521–528 (2006)
11. Sher, I.: Optimal shill bidding in the VCG mechanism. Econ.Theor. **50**(2), 341–387 (2012)
12. Matsuo, T., Ito, T., Shintani, T.: A new approach to detecting shill bids in e-auctions. Int. J. Intell. Inf. Database Syst. **1**, 3–17 (2007)
13. Yokoo, M., Sakurai, Y., Matsubara, S.: The effect of false-name bids in combinatorial auctions: new fraud in internet auctions. Game Econ. Behav. **46**(1), 174–188 (2004)
14. Milgrom, P.: Putting Auction Theory to Work. Cambridge University Press (2004)
15. Todo, T., Iwasaki, A., Yokoo, M., Sakurai, Y.: Characterizing false-name-proof allocation rules in combinatorial auctions. Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems, pp. 265–272 (2009)

A Model-Based Method for Modeling and Verifying Event-Based Aspect-Oriented Applications

Hong Anh Le[1(✉)], Ninh Thuan Truong[2], and Aziz Nanthaamornphong[3]

[1] Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam
lehonganh@humg.edu.vn

[2] VNU - University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam
thuantn@vnu.edu.vn

[3] Faculty of Technology and Environment, Prince of Songkla University, Phuket Campus, Songkhla, Thailand
aziz.n@phuket.psu.ac.th

Abstract. Event-based aspect-oriented programming (EAOP) model allows to systematically treat relationship between point-cuts and to execute aspects whenever receiving events emitted by the base program. However, this model neither comes with formal specification nor provides any mechanism to verify its properties formally. In this paper, we propose a model-based method based on Event-B formal method to analyze an EAOP application. First, we specify its components in Event-B notations where we make use of Event-B refinement mechanism to model the base and the monitoring programs. After that, we exploit Event-B generated proof obligations to check if the application constraints are affected by aspect cross-cuts. Finally, the proposed method is illustrated in detail with an example.

Keywords: Event-based · Aspect-oriented · Verification · Event-B

1 Introduction

Aspect-Oriented Programming (AOP) [5] extends object-oriented programming by introducing the term "aspect" to represent key unit of modularity. It also introduces some more new terminologies such as *join point*(a point during the execution method or error handling of a program), *advice*(action taken by an aspect at a particular join point), and *advice*(a predicate that matches join points), etc.

Extending the AOP approaches (e.g., AspectJ), which restrict crosscuts to individual points during program execution, Dounce et al. presented an Event-based Aspect-Oriented Programming (EAOP) approach, determines aspect

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_23

weaving by detecting a sequence of events. The approach makes it possible to use aspects to change events instead of changing each class separately. It is desirable to check if the program still preserves certain defined properties after weaving the program. However, the approach does not provide any mechanism to verify in such perspectives. There are a numerous number of research works that have focused on modeling and verifying aspect-oriented programs using various techniques such as UML [10], model checking [9], Petri-net [4], and B [7] but they seem are not suitable for an event-based system. In this paper, we propose to use Event-B formal method to verify such applications.

Event-B [1] is a formal method that is more suitable for developing large reactive and distributed systems. Software development in Event-B begins by abstractly specifying the requirements of the whole system and then refining them through several steps to reach a description of the system in such a detail that can be translated into code. The consistency of each model and the relationship between an abstract model and its refinements are obtained by formal proofs. Support tools also have been provided for Event-B specification and verification, e.g. the Rodin platform.

The contributions of this paper are (1) present rules to translate main components of an event-based aspect-oriented application to Event-B notations, (2) make use of Event-B refinement and its generated proof obligations to verify if aspect weaving of the base program violates invariant properties, and (3) illustrate the proposed method in a program.

The rest of the paper is structured as follows. Section 2 provides some background of EAOP and Event-B. Section 3 is our main work where we present translation rules and how to verify an EAOP application. A running example is given in Sect. 4. Next, Sect. 5 summarizes related work. Finally, we give some conclusions and present future works in Sect. 6.

2 Backgrounds

2.1 Event-Based Aspect-Oriented Programming

Aspect-Oriented Programming (AOP) is an emerging programming paradigm that separates interested concerns at the implementation level. Note that a concern is something that we want to be able to treat as a single conceptual unit. In software development perspective, AOP allows the developers apply aspects that alter behavior to classes or objects independent of any inheritance hierarchy. The developers then can apply these aspects either during runtime or compile time. Here, we will describe the four key elements of AOP.

- Join point—It is a place that can be defined between logical steps of the execution of the program.
- Pointcut—It is a set of join points (or an expression that describes a set of join points). Pointcuts can be simple, such as “before every constructor in a class” or complex, such as “before every method in a class that begins with get”

- Advice—It is the code that performs the cross-cutting concern.
- Mixing—An instance of a class to be mixed in with the target instance of a class to introduce new behavior.

Event-based AOP is a general framework that is presented in [2]. This model abstracts the points of interest as events emitted during program execution independently from any specific programming languages. Key feature of the model is that an aspect can define an action for a sequences of events instead of individual points as described in the AOP model. It has the following characteristics :

- Aspects are defined in terms of events emitted during program execution.
- Cross-cuts relate sequences of events, possibly including state modifications. They are defined by event patterns which are matched during program execution.
- Once a cross-cut has been matched, an associated action is executed.

2.2 Event-B

Event-B [1,3] is a formal method for system-level modeling and analysis. The main features of Event-B are three-fold: (1) the use of the Zermelo-Fraenkel set theory as a modeling notation; (2) the use of refinement to represent systems at various level of abstraction; and (3) the use of mathematical proof to verify consistency between refinement levels. A basic structure of an Event-B model consists of CONTEXT and MACHINE.

An Event B CONTEXT describes a static part where all the relevant properties and hypotheses are defined. A CONTEXT consists of carrier sets, constants, axioms. Carrier sets, denoted by s, are represented by their names, and are non-empty. Different carrier sets are completely independent. The constants c are defined by means of a number of axioms $P(s, c)$ also depending on the carrier sets s.

A MACHINE is defined by a set of clauses. A machine is composed of variables, invariants, theorems and events. Variables v are representing states of the model. Invariants $I(v)$ yield the laws that state variables v must always be satisfied. These laws are formalized by means of predicates expressed within the language of First Order Predicate Calculus with Equality extended by Set Theory. Events $E(v)$ present transitions between states. Each event has the form evt = any x where $G(x, v)$ then $A(x, v, v')$ end where x are local variables of the event, $G(x, v)$ is a guard condition and $A(x, v, v')$ is an action. An event is enabled when its guard condition is satisfied. The event action consists of one or more assignments. We have three kinds of assignments for expressing the actions associated with an event: (1) a deterministic multiple assignment ($x := E(t, v)$), (2) an empty assignment (skip), or (3) a non-deterministic multiple assignment ($x : |P(t, v, x')$).

To deal with complexity in modeling systems, Event-B provides a refinement mechanism that allows us to build the system gradually by adding more details to get more precise model. A concrete Event-B machine can refine at

most one abstract machine. A refined machine usually has more variables than its abstraction as we have new variables to represent more details of the model. In superposition refinement, the abstract variables are retained in the concrete machine, with possibly some additional variables. In vertical refinement such as data refinement, the abstract variables v are replaced by concrete ones w. Subsequently, the connections between them are represented by the relationship between v and w, i.e. gluing invariant $J(v, w)$.

Rodin, an extension of the Eclipse platform, allows to create Event-B models with an editor. It also automatically generates the proof obligations of a model that can be discharged automatically or interactively. The architecture of the tool is illustrated in Fig. 1. Event-B UI provides users interfaces to edit Event-B models. Event-B Core has three components: static checker (checking the syntax of Event-B models), the proof obligation generator (producing simplified proof obligations that make them easier to discharge automatically), and the proof obligation manager (managing proof obligations and the associated proofs). The Rodin Core consists of two components: the Rodin repository (managing persistence of data elements) and the Rodin builder (scheduling jobs depending on changes made to files in the Rodin repository).

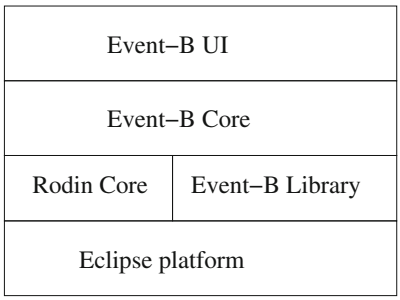


Fig. 1. Rodin tool architecture.

3 Modeling and Verifying EAOP Systems

In this section, we introduce a method for modeling and verifying EAOP applications. First, we introduce new definitions of the application components. Based upon these definitions, we specify the application and its properties in Event-B notations. Finally, we are able to verify these properties based on generated proof obligations.

3.1 Representation of EAOP in Event-B

Recall that EAOP is a model for aspect oriented programming based on execution events. If a base program emitted an event or a sequence of events, then

the monitoring component executes associated aspects. We introduce some definitions of its components as follows.

Definition 1. (*base program*) A base program is a 4-tuple $BP = \langle E, A, V, C \rangle$, where E is a set of events, A is sequences of actions, V is a set of attributes of the program, and C is a set of attribute constraints.

Since we consider the base program as an event-based program, it simply consists of events, actions, variables, and constraints. Attribute constraints are considered as the desired properties of the application because it is required to satisfy several certain conditions.

Definition 2. (*cross-cut*) A crosscut, denoted by $CC \subseteq E$, defines a sequence of events representing defined points in the base program execution.

An aspect in the EOAP model consists of new variables and a cross-cut where it weaves code to the base program.

Definition 3. (*aspect*) An aspect in EOAP model denoted $A = \langle Vr, CC \times S \rangle$ where S is a set of advices associated with the cross-cut CC and Vr states new variables.

Example: An EOAP application containing an aspect that commits a source code file to a server whenever it is modified within a session. This program can be defined as $A = \langle \{\}, \{login \mapsto do\_login, modify \mapsto commit\_svn, logout \mapsto do\_logout\} \rangle$, where login, modify and logout are three events; *do_login*, *commit_svn*, and *do_logout* are three corresponding advices respectively.

3.2 Modeling EAOP Systems Using Event-B

In this Section, based on the definitions in Sect. 3.1, we present translation rules to translate an event-based aspect-oriented application to Event-B notations.

Rule 1: Base program $P = \langle E, A, V, C \rangle$ is translated to an abstract Event-B machine M such that $e \in E$ is mapped to an event of machine M, actions of the base program are modeled as the body of the Event-B events, program variables are translated to variables of machine M, and program constraints are described by Event-B invariant clauses.

Rule 2: Aspects are performed when a sequence of events emitted by the base program. We model aspects using refinement mechanism in Event-B. More specifically, each aspect is transformed to a concrete machine, which refines the abstract machine (representing for the base program).

Rule 3: Advices associated with events in an aspect are translated to actions of the corresponding Event-B events.

3.3 Verifying System Properties

After specifying the application in Event-B, we exploit generated proof obligations to verify its constraints.

Since an aspect can change variables of the base program, it may cause violation to the base program constraint. Therefore, we need to make sure that weaving of aspect does not change these constraints.

Proposition 1. *With the translation rules proposed in Sect. 3.2, an aspect preserves the constraints of the base program.*

Proof. Let $P = \langle E, A, V, C \rangle$ be a base program and $a = \langle v, e \mapsto s \rangle$ be an aspect, where v is new variable, $e \in E$, and s is an advice. Let v is a variable, is required to satisfy constraint $c(v)$, is modified by the aspect (let v' be v after executing the aspect). We need to prove that $c(v')$ still holds.

With rule 1, e is translated to ev of a machine M, let $g(ev)$ is guard of this Event-B event, variable v is transformed to Event-B variable vb. We also have an Event-B invariant I that denotes the constraint of the base program (i).

With rule 2, we have a refined machine M' containing a refine event ev_r with guard $g(ev\_r)$ (ii).

With rule 3, associated advice of the event ev_r, is translated to the action of the event, assigns the variable vb to vb' (denoted by $A(vb, vb')$) (iii).

INV proof obligation of refine machine M' is generated as follows.

$$g(ev\_r) \wedge I(vb) \vdash I(vb') \tag{1}$$

If Eq. 1 holds, with translations (i), (ii), and (iii), we can conclude that $c(v')$ holds.

4 A Running Example: A Bank Account Program

In this section, we illustrate the proposed method on a small program. The program is designed for processing transactions of user credit card that has three events *withdraw*, *deposit*, and $transfer$ emitted when users withdraw, deposit, or transfer from their bank account. The users bank account need to satisfy their balance is always greater than zero.

With EAOP apporach, we design the base program $P =< E, A, V, C >$, where $E = \{withdraw, deposit, transfer\}$, $V = \{balance, amount\}$ represent account balance and amount that users want to withdraw or deposit, $C = balance > 0$ states the constraint, $A = \{withdraw\_act, deposit\_act\}$. Following Rule 1, we achieve the abstract Event-B model which is illustrated in Fig. 2 ($inv1$ and $inv2$ not only define types of two variables but also make sure the constraint of the program always satisfied). There are three events of the abstract machine that correspond to three events of the base program such as *withdraw*, *deposit*, and $transfer$.

We create an aspect for withdrawal where it charges a fee and offers a bonus if users withdraw the first time. We need to add three more new variables fee, $bonus$, and $bFirst$ to define these values respectively. In the refined event $withdraw\_c$, advice that modify the *balance* is translated to the assignment $a1$.

Using ProB [6], a model checking tool for Event-B, we find a counter-example when proving invariant $balance \in \mathbb{N}1$. It means that this aspect violates the

MACHINE M
VARIABLES
balance
amount
action
targetAcount
INVARIANTS
inv1 : $balance \in \mathbb{N}1$
inv2 : $amount \in \mathbb{N}1$
inv3 : $targetAccount \in \mathbb{N}1$

Event *withdraw* $\widehat{=}$
Status ordinary
when
g1 : $action = withdraw$
g2 : $amount < balance$
then
a1 : $balance :=$
$balance - amount$
end

Event *deposit* $\widehat{=}$
Status ordinary
any
amt
where
grd1 : $amt \in \mathbb{N}$
then
a1 : $balance :=$
$balance + amt$
end

Event *transfer* $\widehat{=}$
Status ordinary
any
amt
accountNo
where
grd1 : $amt \in \mathbb{N}$
grd2 : $accountNo \in \mathbb{N}$
then
a1 : $balance :=$
$balance - amt$
a2 : $targetAccountNo :=$
$accountNo$
end

Fig. 2. Event-B specification of the base program.

MACHINE M'
REFINES M
VARIABLES
fee
bonus
bFirst
INVARIANTS
inv1 : $fee \in \mathbb{N}1$
inv2 : $bonus \in \mathbb{N}1$
inv3 : $bFirst \in BOOL$

Event *withdraw_c* $\widehat{=}$
extends *withdraw*
when
g1 : $bFirst = TRUE$
then
a1 : $balance := balance - amount - fee + bonus$
end

Fig. 3. Event-B specification of the aspect.

constraint of the base program. To make the model precise, we need to strengthen the guards of the event $withdraw\_c$ by adding one more clause $g2 : bonus > fee$ indicating $bonus$ is greater than fee, i.e., the aspect needs to check this condition in its advice (Fig. 3).

5 Related Work

Some research work exploited the meta-level notation of UML or extends the UML notation for specifying crosscutting concerns. However, they have not dealt with aspect verification due to the informal or semi-formal nature of UML.
Ubayashi and Tamail [8] proposed a method to verify AOP-based program using model checking. This method targets on the programming phase and applies model checking to the result of weaving code of classes and aspects. It ensures the verification correctness however ignores the modular verification issue. It means that it is difficult to use this method to verify large software.
Dianxiang Xu et al. [9] have proposed to use finite state machines to model and verify aspect oriented programs. They have transformed the woven models and the class models not affected by the aspects into FSP processes, which are to be checked by the LTSA model checker against the desired system properties. However, this method needs to take one more immediate step to transform the base program and aspects to state models before generating FSP processes.
Thuan and Ha [7] used B method to verify weaving of aspects. The authors expressed the base class and some related aspects of AspectJ model in B notations. It aims to benefit from proof obligations generated by B tools to ensure the correctness of AspectJ component composition.

6 Conclusion and Future Work

EAOP is an extended approach for aspect-oriented programming. It combines advantages of both aspect-oriented programming and event-based architecture. In this paper, we present a new method for modeling and verifying an event-based aspect-oriented application using Event-B formal method. The proposed method, based on the translation of an EAOP program to Event-B notations, makes use of refinement mechanism to verify program constraints in each aspect. The method, however, needs to extend to work with more sophisticated crosscuts.

Acknowledgement. This work is partly supported by the project no. 102.03–2014.40 granted by Vietnam National Foundation for Science and Technology Development (Nafosted).

References

1. Abrial, J.-R.: Modeling in Event-B: System and Software Engineering, 1st edn. Cambridge University Press, New York, NY, USA (2010)
2. Douence, R., Südholt, M.: A model and a tool for event-based aspect-oriented programming (EAOP). Technical report TR 02/11/INFO, École des Mines de Nantes (2002)
3. Event-b and the Rodin platform. http://www.event-b.org (2012)
4. Guan, L., Li, X., Hu, H.: A petri net-based approach for supporting aspect-oriented modeling. In: Theoretical Aspects of Software Engineeering, 2008, pp. 83–90 (2008)
5. Holzer, A., Ziarek, L., Jayaram, K., Eugster, P.: Putting events in context: aspects for event-based distributed programming. In: Proceedings of the Tenth International Conference on Aspect-oriented Software Development, AOSD '11, pp. 241–252. ACM, New York, NY, USA (2011)
6. Ligot, O., Bendisposto, J., Leuschel, M.: Debugging event-b models using the prob disprover plug-in. In: Proceedings AFADL'07, June 2007
7. Thuan, T.N., Ha, N.V.: Using b to verify the weaving of aspects. In: 14th Asia-Pacific Software Engineering Conference, APSEC 2007, pp. 199–205, Dec 2007
8. Ubayashi, N., Tamai, T.: Aspect-oriented programming with model checking. In: Proceedings of the 1st International Conference On Aspect-oriented software development, AOSD '02, pp. 148–154. ACM, New York, NY, USA (2002)
9. Xu, D.-X., El-Ariss, O., Xu, W.-F., Wang, L.-Z.: Aspect-oriented modeling and verification with finite state machines. J. Comput. Sci. Technol. **24**(5), 949–961 (2009)
10. Zhang, J., Chen, Y., Liu, G.: Modeling aspect-oriented programming with UML profile. In: First International Workshop on Education Technology and Computer Science, 2009. ETCS '09. vol. 2, pp. 242–245, March 2009

A Retrospective on Constructing a Personal Narrative on Agile Development

Roy Morien(✉)

Naresuan University Language Centre, Phitsanulok, Thailand
roym@nu.ac.th

Abstract. The underlying principles of Agile Development had been stated and established for 30 years before the Agile Manifesto was proposed. Software Prototyping, Rapid Application Development, Rapid Development, Iterative Development, even Help Desks and SWAT teams must be seen as antecedents of Agile Development, as we understand it today. The original proponents of 'agile development' attempted to develop a new narrative for software system development. However, notwithstanding the length of this history, now some 45 years in the making, many developers, and education institutions, still question the approach, still refuse to include it in course curriculum, at least partly due to it being ignored in IEEE curriculum structures, still demand 'proof that it works', and still adhere to the traditional linear phased thinking of the Waterfall Approaches. Responding to the need for 'proof' and attempting to debate the adherents of the long-standing Waterfall Approaches has proven futile, but a new narrative has been evolved which attempts to describe Agile Development as being a normal, usual approach to human activity, reflecting the way people inevitably act when intending to undertake, and in the process of undertaking, any purposeful activity with a planned outcome. Uncertainty is the norm, planning is essential but empirical and adaptive, leadership is preferred over management. Other research disciplines have much to contribute to agile thinking and action.

Keywords: Agile development · Agile projects management
Empirical planning · Experiential planning · Project learning
Just-in-time learning · Just-in-case planning

1 Introduction

In 2001 a group of seventeen software developers and consultants proclaimed a new software development manifesto, the Agile Manifesto [1]. A certain amount of mythology seems to have been created since then as to who they were, why they did this, where their ideas came from. What can be said without much dissent is that this was a group of experienced developers frustrated with the conventional wisdoms of the time, in regard to systems development methodologies, and were seeking to apply their experience and knowledge to the creation of a different way of thinking and a different way of behaving in a software systems development project. This also implied a different way of managing such projects, but this is not specified in the Manifesto, merely implied.

T. Theeramunkong et al. (eds.), *Recent Advances and Future Prospects in Knowledge, Information and Creativity Support Systems*, Advances in Intelligent Systems and Computing 685, https://doi.org/10.1007/978-3-319-70019-9_24

Pre-2001, many pre-"agile development" development methods were published and researched. These were variously termed iterative methods, evolutionary development approaches, rapid development approaches and software prototyping methods. While usually misquoted and misinterpreted, Royce's original paper in 1970 [2] discussed iterative development, and subsequently Gilb in 1988 clearly proposed an iterative, 'agile' approach [3]. By the mid 1980s extensive literature was available on Software Prototyping [4–8], Rapid Application Development [9] and Rapid Development [10], the prominent 'agile' approaches at that time. Since then a plethora of books have been published describing a variety of different 'agile' approaches, including Scrum [11], Extreme Programming [12], DSDM [13], FDD [Feature Driven Development] [14], Crystal Clear [15], Lean Software Development [16–18] and others.

As well, many Agile User Groups have been set up around the world. How active or influential these user groups are, and how many people actually have joined and actively become involved is not immediately known. However, their geographical distribution clearly implies the level of interest in Agile Development world-wide. This does however suggest a significant discrepancy in interest in and use of Agile Development in different parts of the world. This list, available at [19], shows the geographical distribution. By continent, North America (the USA and Canada) has 95 groups, Europe has 62 (extending from England to Ukraine, and includes Russia), Asia (extending from Japan to Iran) has 25, South America has 11, Oceania (Australia and New Zealand) has 7 and Africa has 4. Overall 204 groups of which 76% are in North America and Europe. It is also interesting to note that there is an annual Agile Conference in the USA which in excess of 1500 practitioners attend. This must be the best attended conference of all time; at least, this author's experience of attending more than 30 conferences, including the Agile Conference in Denver, Colorado, in 2005 certainly suggests that. Overall, there were 105 Conferences, in 26 countries, with a total of around 200 conference days, and all the official Twitter accounts put together have more than 100,000 followers, in 2014. The USA leads with 18 Agile conferences, Germany had 13 and the UK had 11. These were the top countries by way of holding agile conferences and events. There can be no doubt that the topic of Agile Development, and the various antecedent approaches have been well published, and appear to be well supported in North America and Europe, principally. So why is there a need for "a new narrative for Agile Development"?

2 A Retrospective Review

This paper is a principally a retrospective on previous publications by this author. Over the last 20 years, this author has presented some 35 or so papers at conferences, or published in journals, or as a book chapter, and in a published, and an unpublished, textbook, on the topic of Agile Development methodology and Information Systems pedagogy. These papers are referred to here not as prior evidence supporting a current proposition; such self-referencing using previously published papers as being the authority or proof of a current proposition is not acceptable. The purpose of this self-referencing is to construct a coherent and hopefully cogent narrative from the

various fragments of research, experience, opinion and viewpoint proffered in the past. This publication history has its genesis in 1991 [20] and 1994 [21, 22]. In support of this narrative, a software system that so recently spectacularly failed is discussed. This failed system was developed by well-known international consultancies for a State Government Service in Australia.

3 The Matter of Certainty

The proposition stated in [23] was *"In the traditional Waterfall Approach, every attempt is made, at the start, to impose order and certainty on the process. ... This approach to systems development was favoured because it apparently provided an appropriate level of certainty at the outset; certainty of scope, certainty of cost and certainty of time. This desire for certainty was inevitably a vain quest. ... To create 'certainty' at the start of a project is to imply that the future can be controlled, which is a fallacy. Uncertainty is the hallmark of the future."*. As the author of this paper, and this quote, I was able to draw on personal experience and observation of project management practices during the late 1970s and early 1980s. I refer to a major development project failure that is, at the present time (2015) subject to litigation in Australia [24]. A number of different media reports refer to this disaster, which was also previously referred to by this author in a conference [25]. Quotes of interest and importance from the various media reports indicate the obviously futile attempt at introducing certainty into the development project; *"The Barrister acting for IBM ... argued any failures with the scheme were because the Government was 'not able to define and stick to a scope'"* and *"*(referring to the) *QC for the Queensland government ... there were up to 20 drafts of the supplementary agreement between lawyers, to show it was very carefully worded ... the release is phrased very specifically.",* yet this agreement is in dispute. It appears that, in this situation, a software development project that cost in excess of $1.2 billion, the seeking for certainty did indeed prove futile, and an obviously arduous attempt at ensuring the exact details of an agreement fundamentally failed.

So, this is the first paragraph in the construction of the narrative supporting Agile Development; Agile Development is a development approach that acknowledges that the future is uncertain, and attempting to eradicate uncertainty is futile.

4 We Always Plan, Don't We?

It is a dismal observation that attempting to control the future is a futile endeavour. If we have no way of controlling the future, or at least being able to meet and overcome the obstacles that we may face, then we have chaos. So we must make every effort to identify the future risks and to have a plan of action to overcome them. There is no doubt that planning has always been an essential activity in any purposeful human endeavour. According to the greatest planners in history, planning is essential, but this comes with the caveat that once the work begins the plan becomes useless, unless constantly revised according to contemporary circumstances. So, this develops the

narrative; First, seeking certainty is futile, but normal human action and personal behaviour encompasses planning the activity when the activity is purposeful. Plans are intuitively, that is normally, modified depending on circumstances experienced and arising during the planned activity, new information is learned, and unforeseen obstacles or opportunities are encountered and overcome or embraced. Publilius Syrus, a 1st Century Roman writer of apparently lasting reputation, said *"It's a bad plan that admits of no modification"*. General George S. Patton, commanding general of one of the most successful armies in World War II stated *"If you tell people where to go, but not how to get there, you'll be amazed at the results*" and *"Never tell people how to do things. Tell them what to do and they will surprise you with their ingenuity."* Patton certainly did not eschew planning, which allowed Patton's 3rd Army to swiftly and effectively meet the surprise attack by German forces in what became known as the Battle of the Bulge, in 1944. Patton's 3rd Army was able to change the direction of its line of march and proceed rapidly, almost at an hour's notice, to join the battle, marching over a considerable distance to do so. How was this achieved? By Planning for such a contingency! We might call this 'just-in-case' planning. Dwight D. Eisenhower, 34th President of the USA and the General commanding the greatest military operation in human history, put it succinctly; '*Plans are nothing; planning is everything.'* On another occasion he is quoted as saying *"In preparing for battle I have always found that plans are useless, but planning is indispensable"*. Benjamin Franklin [1706–1790] American statesman, scientist and philosopher is quoted as saying "*By failing to prepare, you are preparing to fail.".* However, to use the words of Isaac Asimov *"It is change, continuing change, inevitable change that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is, but the world as it will be.", and,* last but not least, to quote Sir Winston Churchill, Britain's war time Prime Minister, "*Plans are of little importance, but planning is essential.*"

This, then, is the next stanza of the conversation on agile development. Without plans we are almost inevitably likely to fail, but when we do plan, carefully and meticulously, we are still subject to the Law of Uncertainty, and we must take an adaptive approach, based on empirical observation, to our on-going adherence to the plan. Is there a problem of concept and practice here? We need only look at the oft-repeated mantra of *'Plan the Work and Work the Plan'* seen as the guiding principle of 'good project management' to see the problem. This saying disavows the possibility of changing circumstances in the future, and indeed ignores the possibility that comprehensive and correct information at the time of planning may not be 100% correct or exact. It can be said that by tying your on-going actions to the initial plan, assiduously 'working the plan', you are ignoring and eschewing the possibility of learning more and other information about the system requirements, leading to the delivery of a system that is no longer required, or does not meet the requirements at the time of delivery.

5 Agile Development—It's Normal

Why do we need substantial and well structured software development methodologies, even agile methodologies, to guide our development activities? Is planning and adaptation so foreign to people that a strict regime governing the practice must always be in place? Is it essential to have methodology adherence audits to ensure the methodology is being adhered to? In [26] the author argued that every activity undertaken by any person can be described as being done in an agile manner. *"Long and hard consideration of these particular notions* (about the requirement for planning and methodology) *has arrived at the conclusion that these encompass almost all aspects of human endeavour, and are essential to the maximisation of the success of any human activity. They can therefore, surely, be seen as normal, natural actions and the way things are usually done by people generally, in their normal and daily lives and activities… ".* This helps us define the concept of a project, and to answer the question What is the nature of a "project"? An acceptable answer to this question is *"A project is a purposeful activity being undertaken by, or intended to be undertaken by, one person, or a group of people acting together as a team, with the intention to achieve a useful outcome". The terms "purposeful" and "thoughtful" are used to imply activities that are not just spur-of-the-moment, ad hoc and impulsive.".* A question in tandem with this is then "What is a project plan?" Again, a reasonable, brief definition would be *"A thoughtful view of a future project sufficient to guide the activities of the project participants".* Plans are always about the future. The terms "*thoughtful view*" encompasses the identification of possible risks and an assessment of the probability of the risks occurring, the seriousness of the risks, and the cost of avoiding the risks. Going forward should not be "going into the night" without a plan.

The proposition in [26] that every activity undertaken by any person is actually an agile project was illustrated by considering an activity undertaken by probably every urbanite in any large city around the world, that of going to the supermarket to buy groceries. Nothing could be more commonplace or normal than this, in most urban communities. This normal and possibly daily activity of going to the supermarket can readily be seen as a project (a purposeful activity to be undertaken with a useful outcome). Even for such a simple activity the actor will have a plan and a checklist, mentally or written, or both. Without a plan the shopper probably would never arrive at the supermarket, or, if they did, would probably not know what it is that they came to purchase, or they might not have any money to buy anything because they didn't think to check their wallet or purse before leaving the house. The plan for such a simple activity, especially if it is one that has been carried out on previous occasions, would be mostly unconscious planning; they will go by car, and the unspoken assumption is that the car will have enough fuel to go and return, or perhaps a visit to a known service station nearby to refuel is called for. The way will probably be known, or a map must be consulted prior to leaving the house or a GPS used along the way. The potential

shopper will recognise the store when they arrive there, and will know where to park the car, based on prior experience. The conscious planning will include; preparation of a list of items to be purchased, a check to see if they have sufficient money to pay for the purchases, the time they will leave the house, the probable time that they will return. So, when the shopper sets off on this endeavour, to carry out this plan, they are confident that the plan is feasible (technically, financially, geographically…), doable, and that they will be able to achieve the outcomes intended. What cannot be controlled to a considerable extent are the variables inherent in the future. Along the way, road conditions are observed; stopping at red lights, stopping at pedestrian crossings to allow pedestrians to cross, driving slowly if the other traffic is slow (empirical observations), and, if for some reason the road is found to be blocked (an empirical observation) the plan must be deviated from and another route found (change of plan). It may even be necessary to terminate the project and return home {cancellation of project due to circumstances beyond control of the participants). Once at the supermarket it might be found that one or other of the articles that were on the shopping list are not available (an empirical observation) and again there will be a deviation from the plan to either forgo the purchase or purchase a substitute. Another similar product may be found that is "on special" at an attractive, cheap price (an empirical observation) and can be purchased instead of the intended product (an unplanned action based on new knowledge—learning!). Should the slogan *"... work the plan"* be assiduously applied and the more expensive item originally identified in the plan be purchased, because it was in the plan? This seems a foolish notion, the best outcome from the project would not be achieved if the shopper remained totally constrained by the original plan.

What is being described here are the risks, and indeed the opportunities, that may be inherent in the project; all projects have risks, and all projects potentially present opportunities not foreseen. Alternative actions that may be taken, or be forced to be taken, are being described. What is clearly being considered are the empirical aspects of the project activity as it proceeds, which almost inevitably will arise; every project is a Learning Activity. Adapting the plan to suit contemporary circumstances, to take advantage of new knowledge, and avoid previously unseen risks seems both reasonable and necessary, if not essential. The three important words that are applied every day, everywhere, whenever ordinary people go about their daily business; adaptability, experiential learning, empirical behaviour. Agile behaviour is normal, usual and essential!

It does not take a very great leap of logic to be able to view almost any activity undertaken by people as being a project that is planned, but is naturally and inevitably modified by what is discovered along the way. Most people in their daily activities, intuitively understand this, and any purposeful activity is undertaken in this manner. This is the nub of this narrative.

6 Management Versus Leadership

In [26], the question was posed "*Does the quote from Peter Drucker have relevance here?"* which was *"Most of what we call management consists of making it difficult for people to get their work done."*. Quoting from [27], "*The primary 'resource' in software systems development is 'the people'. It is all about 'the people'*. McConnell [10] states [at page 12] *'we now know with certainty that peopleware issues have more impact on software productivity and software quality than any other factor.* And at page 13 *'it is now crystal clear that any organization that's serious about improving productivity should look first to the peopleware issues of motivation, teamwork, and staff selection and training*. Other significant authors in the information systems field have discussed the significance and relevance of 'people' in the development process [see Demarco and Lister [28], Constantine [29]]. Looking outside the information systems field to the business management literature, we can see suggestions and assertions about 'people' and 'people management' that are extremely apt to the software development situation. Champy [30] states *'... You cannot have a culture of obedience to chains of command and the job slot. It just won't work'* and *'... enabling [people]; redesigning work so that people can exercise their skills and capabilities to the fullest extent possible – then stepping back and letting it happen.'*

There are many case studies that support the idea of 'leadership' rather than 'management'. Referring to [31] *"... The efforts of the team were ultimately highly successful ... The role of the leader of the team was a major factor in this success. The success of the team's project was attributed to (1) dedication to a common purpose (2) acceptance of a performance challenge (3) a sense of mutual accountability (4) candour and mutual respect between team members (5) a shared affection for each other, arising from the shared experience, and the shared success"*. Another telling case study on the idea of leadership as a success factor, published in Maani and Cavana [32] is where, in 1995, a team from New Zealand won the famous and prestigious yacht trophy, called the Americas Cup. The amazing thing was that the NZ team's performance surpassed any previous campaign. How did they do it? The success has been attributed to the inspirational leadership of the syndicate Leader, the strong sense of community within the team, the openness of communication between team members.

We are uncovering better ways of developing software by doing it and helping others to do it. Through this work we have come to value:		
Individuals and interactions	Over	Processes and tools
Working software	Over	Comprehensive documentation
Customer collaboration	Over	Contract negotiation
Responding to change	Over	Following a plan
That is, while we value the items on the right, we value the items on the left more		

Following this line of thinking, this author has developed what he terms the Agile Project Management Manifesto, following the style of the original Agile Development Manifesto [1], included here.

We are seeking better ways of managing software development projects. Through this endeavour we have come to value:		
Leadership	Not	Management
Team empowerment	Not	Compliance to control
Collaboration and cooperation	Not	Organisation hierarchy
Adaptive behaviour	Not	Planning and estimating rigidity
Experiential learning and development	Not	BDUF and static requirements
Welcome change	Not	'Frozen specs' and fear of 'scope creep'
That is, while traditional thinking values the items on the right, we value the items on the left instead.		

In the same manner of thinking and publication, the Agile Project Management Manifesto, developed by this author, is stated here. Support for these concepts has been found in the literature, but, unfortunately, not in the Computer Science or Information Technology literature [apart from inference and hint in the published literature on Agile Development].

7 The System Developer as a Professional

For the various items in the Agile Project Management Manifesto to be relevant and applicable, systems developers must be considered to be professionals, as the term 'professional' is often applied. That is, practitioners well educated and knowledgeable in their field of endeavour, subject to certain ethical constraints, and dedicated to life-long learning to ensure their relevance in the contemporary environment. Such professionals, manifesting these professional attributes, can effectively work in a team without strict management control; that is, Team Empowerment. Well educated and confident practitioners will be willing and able to collaborate and cooperate with others, understanding their own weaknesses and strengths, and empathising with the often awkward needs of their clients; Welcome Change. The ability to adapt to circumstances and to overcome problems, Adaptive Behaviour, comes with professionalism which itself is an outcome of good education and training, as well as the ability to learn good lessons from personal experience. Thus we can state a project management manifesto when we have professional developers. Without this certainty of professionalism, old-fashioned Taylorisms will prevail, which are not appropriate to software systems development, which have been described as 'wicked problems'.

Where do budding software development professionals learn those soft, professional skills, of adaptable behaviour, self-motivation, creative thinking and all those beneficial attitudes that will allow them to overcome the inevitable adversities when seeking a solution to a 'wicked problem'? In the universities and colleges, surely. It is this aspect of professional development education that has been found wanting!

As an example of the Essential Professional, consider the airline pilot, the Captain of an A380 mega-airliner, responsible for the safety of a 430 million dollar aircraft and up to 500 passengers and crew. Such a person requires long training, a long 'apprenticeship' as a Second Officer, frequent re-testing to maintain competence.

Such a person also follows a plan that has been carefully developed. Partly, that plan is in the form of a checklist which must be assiduously followed before take-off, to ensure that none of the myriad of actions to be done are overlooked. There is also a checklist to be followed prior to landing, but that is not attended to until needed; that is planned, but comes into play in the future without any problematic effect in the present situation. During the flight again planning is attended by empirical observation, and the pilots' training comes into play in any emergency, such as flying through a cloud of volcanic ash. The point here is that, again, planning is a normal, usual and indeed essential activity in any purposeful endeavour, even with highly trained and professional actors, but must be modified as necessary by immediate circumstances.

8 The Risks Inherent in Change

In computer systems development many project managers are constrained by the Fear of Change. Forty years ago, in the 1970s and 1980s, this was a realistic fear in that the tools and software available at that time did not allow changes to be made simply. Indeed, at least in Australia, there was an informal notion of 'C&E' changes, meaning changes that could only be made at Christmas or Easter. Why? Because only at these times were there sufficient holidays, consecutive non-business days, for changes to be made to the database, tested and, if necessary, backed out and the original database schema reinstated, without disrupting business activities. Because the prevailing DBMS's of the time imposed such enormous constraints on change management, even the simplest change was viewed with horror and dismay! In modern times however, as discussed in [33] the widespread use of Relational Databases, and associated schema management tools has made creating and maintaining databases almost 'commoditised' and simple, and the plethora of software development tools have placed significant development power in the hands of developers which have enabled 'agile development' of software by code generators, report and graphics generators, use of readily available function libraries and GUI development environments. Change is now almost the 'fear of dragons' relevant to the past, not the present or future.

9 Agile Education

From almost the beginning of the 'agile age', education institutions failed to include Agile Development (in its various forms and names) in their curriculum. To this day official curriculum standards published by such as IEEE do not include specifics of Agile Development. In 1994 I published a paper on the challenge faced by tertiary education institutions in information systems education [22]. My conclusion then was that "*tertiary education of IST practitioners must be more broadly business based, emphasising ongoing professional development as a career necessity… tertiary educational institutions could implement both a technological stream, and a managerial stream in their MIS curriculum*". The clear implication here was the necessity to broaden such education beyond the technical. This was emphasised again in 2010 [23] and 2014 [34]. An example of the broader curriculum is indicated in [28] which

discusses Lean Manufacturing and uses the Toyota Motor Company as an exemplar. From that paper *"The principles of lean production enabled the company to deliver on demand, minimize inventory, maximize the use of multi-skilled employees, flatten the management structure and focus resources where they were needed"*.

In [35], published in 2014, the proposition is that computer systems development education [using a catch-all term] can and should be itself agile, as well as the curriculum containing a significant element on Agile Development. The unfortunate fact, even today, in the face of a dynamic, fast moving and almost revolutionary change in computer systems development style and activities, computer systems development education is still fundamentally entrenched in last century thinking and practices, which may be categorised somewhat whimsically as 'giving long lectures on data flow diagrams'. As stated in that article *"A typical course curriculum plan for courses in Business Computing, Information Technology, Information Systems and the like usually includes Systems Analysis, Design, Programming and Database Design as distinct and separate subject areas. Overall, most curriculum is taught in discrete time limited subjects, by different academics with diverse experience, and often divergent intentions and philosophies. Many subjects are on theory, with a promise that the subject matter will prove to be useful and applicable in future subjects, and in the future generally. Theory subjects particularly are taught using the conventional 'chalk and talk' lecture approach, implying passive learning by the students."*

The proposition is made that computer systems development is essentially a practical hands-on activity, and computer systems development curriculum would be best taught based on a Project-Based Learning approach, with an 'industrial strength' project as the main activity over the period of the course. Theory would be learned and immediately applied in situ, concepts of 'just-in-time' learning applied, and waste in the form of forgotten or poorly learned curriculum eradicated. I can only wonder if such suggestions are likely to find adherents who also feel that 'agility' in most things, including education, is essential.

There are many examples of what is termed Project-Based Learning being successfully applied [36–43], and Action Research [44–46] being considered a very appropriate, practical, approach to Information Systems Research. It seems that educational and research practices based on practical learning and practical research are considered to be effective and efficient approaches to what I am terming 'Agile Education', meaning both including Agile Development in curriculum, and approach the Teaching and Learning activities in an agile manner. Personal experience in managing the industrial experience capstone project units for over 150 students has demonstrated that this approach is greatly approved of by students, who declared that their experience with Project Based Learning was highly successful and their learning was greatly enhanced [47, 48].

Bringing these propositions together under the heading of Agile Education, references to Lean Manufacturing and subsequently Lean Software Development have been drawn together into what I have termed the 7 Wastes of Education, and the Agile Education Manifesto was published in [49], and is reproduced here. Figure 1 is this author's attempt at correlating the 7 Wastes of Manufacturing with education practices and processes.

The 7 Wastes of Manufacturing	The 7 Wastes of Education
Inventory	Concepts and ideas that must be 'stored' until a future time to be applied which may be forgotten or become obsolete.
Extra Processing	Relearning, due to initial shallow learning and time lapse between related subjects.
Overproduction	Extra and Unnecessary Curriculum and assessment activities.
Transportation	Knowledge must be carried from one Subject to another which may not allow a proper continuation in the learning 'stream'
Waiting	The relevance of the topics in a subject must be kept 'on hold' until a later subject.
Motion	Moving from one subject to another without fully understanding the first subject.
Defects	Shallow Learning, forgotten curriculum, failure to comprehend subject matter relevance.

Fig. 1. The 7 Wastes of Education

In the Agile Education Manifesto, the Individuals referred to are, of course, the teachers and students with the emphasis being placed on the interactions between them. Team teaching, team learning, self-directed learning, mentoring of students by teachers are some of the concepts envisaged under this heading. Hands-on practice, experiential learning with reflective practices, theory taught in situ, immediate comprehension of relevance of subject matter are aspects of the Meaningful Learning. Under Stakeholder Collaboration I would group all of the teachers, the students, future employers and industry bodies working together to ensure quality and relevant curriculum, and deep learning of that curriculum. Especially in the computer industry, Responding to Change is especially relevant if somewhat vexing for the teaching academic. The rate of change in that industry almost makes a mockery of curriculum planning where institutions have an administratively imposed planning horizon of at least 5 years, during which time the industry has almost inevitably 'moved on' (Fig. 2).

We are uncovering better ways of educating children by doing it and helping others do it. Through this work we have come to value:		
Individuals and interactions	**over**	**processes and tools**
Meaningful learning	**over**	**the measurement of learning**
Stakeholder collaboration	**over**	**complex negotiation**
Responding to change	**over**	**following a plan**
That is, while there is value in the items on the right, we value the items on the left more.		

Fig. 2. Agile Education Manifesto

10 Reference Disciplines for Agile Development

In [50], I first published my version of the Agile Project Management Manifesto, and also identified what I consider to be essential and appropriate Reference Disciplines. These included references to Peter Senge's "The Fifth Discipline", [51] in which Peter Drucker was referenced as defining 'the Learning Organization" considered necessary because *'The function of the society of post-capitalist organisations ... is to put*

knowledge to work … it must be organised for constant change', chaordic systems [52] which describes '*the behaviour of any self-governing organism, organization or system which harmoniously blends characteristics of order and chaos*', which I suggest is entirely applicable to the activity of software development, which is a system that indeed manifests those characteristics., Champy's 'Reengineering Management' [53] in which it is stated that '*You must have a culture that encourages qualities like relentless pursuit… bottomless resources of imagination … and both smooth teamwork and individual autonomy'* (and therefore) *'… You cannot have a culture of obedience to chains of command and the job slot. It just won't work.'* and *'the best approach to such a system is '… enabling* (people*); redesigning work so that people can exercise their skills and capabilities to the fullest extent possible – then stepping back and letting it happen.'* and Liker on The Toyota Way of management [54], 'The Minding Organisation' [55], in which The Model of Concurrent Perception was presented, which *'moves us from questions to answers, from divergent perceptions to convergent perceptions, from individual creativity to team implementation, from abstract thinking to concrete action, from quick experimentation to quality results, from deliberate chaos to emergent order'* and *'chaos should be deliberately created up front'*, Kanban Management [56] and team development [57–60], and including also Lean Thinking [61] from which the concept of just-in-time delivery, an important concept in Agile Development thinking, is derived. All of these references from the management literature have lessons for software development methodologists and developers. In general, these authors demonstrate the benefits of team development (collaboration), software development as a 'learning' activity, that software project management is best served by Leadership', not command-and-control management, problem analysis approaches that comprehend the 'chaordic' nature of team-based development and a method for converging on an acceptable and accepted solution.

11 Conclusion

An attempt has been made to construct a new narrative for Agile Development, perceiving this important approach as the normal manner in which humans behave whenever they undertake a purposive and purposeful activity, even the mundane activities of daily life. When discussing Agile Development, the greater variety of intellectual and academic areas of human endeavour and research need to be considered, and a Reference Disciplines Framework has been suggested and elaborated previously. The narrative is based on the acceptance of Uncertainty of the future, the essential empiricism and collaboration inherent in Agile Development, all of which are essential human behavioural characteristics. Agile as a learning approach, and an approach to learning is part of the narrative, issuing a challenge to colleges and universities vis-à-vis their curriculum and educational style.

References

1. Manifesto for Agile Software Development. http://www.agilemanifesto.org/ (2001). Accessed Aug 2015
2. Royce, W.W.: Managing the development of large software systems. In: Proceedings of the IEEE WESCON, August, 1970. https://www.cs.umd.edu/class/spring2003/cmsc838p/Process/waterfall.pdf
3. Gilb, T.: Principles Of Software Engineering Management, 1st edn (1988)
4. Naumann, J.D., Jenkins, A.M.: Prototyping: The New Paradigm for Systems Development. MIS Quarterly, Sep 1982
5. Jenkins, A.M.: Prototyping: A Methodology for the Design and Development of Application Systems. Indiana University (1983)
6. Boar, B.H.: Application Prototyping: A Requirements Definition Strategy for the 80's. Wiley, New York, NY, USA©1984 ISBN:0-471-89317-X
7. Budde R., Kuhlenkamp K., et al.: Approaches to Prototyping. Springer (1984)
8. Hawgood, J. (ed.): Evolutionary Information Systems. North-Holland (1981)
9. Martin, J.: Rapid Application Development. Macmillan (1991). ISBN 0-02-376775-8
10. McConnell, S.: Rapid Development: Taming Wild Software Schedules. Microsoft Press (1996)
11. Schwaber, K., Beedle, M.: Agile Software Development with Scrum. Prentice Hall (2001). ISBN 0130676349
12. Beck, K.: Extreme Programming Explained: Embrace Change, 2nd edn. Addison-Wesley (2004). ISBN 0321278658
13. Stapleton, J.: DSDM: Business Focused Development, DSDM Consortium, 2nd edn (2003)
14. Palmer, S.R., Felsing, J.M.: A Practical Guide to Feature-Driven Development. The Coad Series. Prentice Hall (2002). ISBN 0130676152
15. Cockburn, A.: Crystal Clear: A Human-Powered Methodology for Small Teams. The Agile Software Development Series. Addison-Wesley (2004). ISBN 0201699478
16. Poppendieck, M.: Principles of Lean Thinking. http://www.poppendieck.com/papers/LeanThinking.pdf. Accessed 19 May 2005
17. Poppendieck, M., Poppendieck, T.: Lean Software Development: An Agile Toolkit. Addison-Wesley (2003). ISBN 0321150783
18. Poppendieck, M., Poppendieck, T.: Implementing Lean Software Development: From Concept to Cash. Addison-Wesley (2006). ISBN 0321437381
19. Agile User Groups Listing. http://www.agilealliance.org/resources/agile-usergroups/. Accessed Aug 2015
20. James, H., Morien, R.: Controlling rapid information systems development environments. In: EDP Auditors Conference, Canberra, May 1991
21. Morien, R.: Prototyping large online systems: a focal entity approach. In: 3rd Australasian Conference in Information Systems, Wollongong University, Oct 1994
22. Morien, R.I., Schmidenberg, O.: Educating information systems professionals: the tertiary educational challenge. In: Proceedings of Asia Pacific Information Technology in Training and Education, APITITE'94. Brisbane, June 1994
23. Morien, R.: Seeking a new paradigm for software project management. In: The 9th International Conference on e-Business, iNCEB2010, Kasetsart University, Bangkok Thailand, 18–19 Nov 2010
24. Horn, A.: IBM tries to block legal action over Queensland Health payroll disaster. http://www.abc.net.au/news/2015-08-25/ibm-tries-to-block-legal-actionover-qld-health-payroll-disaster/6723598?section=qld. Accessed 26 Aug 2015

25. Morien, R.: Agile development: it's only natural. In: Agile Symposium, Adjunct Symposium to the 8th Malaysian Software Engineering Conference, Langkawi, Malaysia, Sept 2014
26. Morien, R.: Back to basics—in support of agile development. In: 7th Malaysian Software Engineering Conference, Melaka, Nov 2013
27. Plauger, P.J.: Programming on Purpose II: Essays on Software People. PTR Prentice Hall, Englewood Cliffs, NJ
28. Demarco, T., Lister, T.: Peopleware: Productive Projects and Teams. Dorset House, New York (1987)
29. Constantine, L.L.: Constantine on Peopleware. Yourdon Press, Englewood Cliffs, NJ (1995)
30. Champy, J.: Reengineering Management: The Mandate For New Leadership. HarperCollins Publishers (1995)
31. Katzenbach, J.R., Smith, D.K.: The Wisdom of Teams: Creating the High-performance Organization. Harvard Business School (2008)
32. Maani, K.E., Cavana, R.Y.: Systems Thinking, Systems Dynamics—Managing Change and Complexity. Pearson Education, NZ (2007)
33. Morien, R.: Back to basics: in support of agile development. In: Ghani, I. et al. (eds.) Emerging Advancements and Technologies in Software Engineering, Engineering Science Reference, IGI Global. 7th Software Engineering Conference, Melaka, Malaysia, pp. 279–292. 20–21 Nov 2014
34. Morien, R.: An agile project management manifesto—a reference disciplines framework for agile development. In: KSII The 6th International Conference on Internet ICONI, Taiwan (2014)
35. Morien, R.: Streamlining business computing education. In: KSII The 6th International Conference on Internet ICONI, Taipei, Taiwan (2014)
36. Takeuchi, H., Nonaka, I.: New Product Development Game. Harvard Business Review, 86116, pp. 137–146 (1986). January 1, 1986. Accessed 12 March 2013, Cited in Wikipedia
37. Blumenfeld, P.C., Soloway, E., Marx, R.W., Krajcik, J.S., Guzdial, M., Palincsar, A.: Motivating project-based learning: sustaining the doing, supporting the learning. Educ. Psychol. **26**(3–4), 369–398 (2014)
38. Patton, A.: Work that Matters: The Teachers Guide to Project Based Learning, http://www.innovationunit.org/sites/default/files/Teacher's%20Guide%20to%20Project-based%20Learning.pdf. Accessed 1 Oct 2014
39. Felder, R.: Resources in Student Centered Teaching and Learning, Resources in Science and Education, blog. http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Student-Centered.html. Accessed 1 Oct 2014
40. Froyd, J., Simpson, N.: Student-centered learning addressing faculty questions about student-centered learning. In: Transforming Undergraduate Education in Science, Technology, Engineering and Mathematics. Washington DC (2013)
41. Nielsen, L.: Student Driven Learning = Passion-Based Classrooms, blog. http://theinnovativeeducator.blogspot.com/2011/04/student-driven-learning-passion-based.html. Accessed 1 Oct 2014
42. National University of Singapore: Learning to Teach, Teaching to Learn, A Handbook for NUS Teachers. http://www.cdtl.nus.edu.sg/handbook/home/foreword.htm. Accessed 1 Oct 2014
43. Alausa Y.A.: Continuous assessment in our schools: advantages and problems. http://www.nied.edu.na/publications/journals/journal9/Journal%209%20Article%202.pdf. Accessed 1 Oct 2014
44. Baskerville, R.: Investigating information systems with action research. Commun. Assoc. Inf. Syst. **2** (1999). Article 19

45. Kock, N. (ed.): Information Systems Action Research: An Applied View of Emerging Concepts and Methods. Springer (2007)
46. Fu, C.C.: Information systems development as action research: changing a school. Doctoral thesis, Universidad ESAN, Peru (2006)
47. Morien, R.: Insights into using agile development methods in student final year projects. In: Proceedings InSITE2004, Issues in Informing Science and Information Technology. Central Queensland University, Australia. https://www.informingscience.org
48. Morien, R.: Student experience of using agile development methods in industrial experience projects. Proc. ISECON 2005, v22, Columbus, Ohio, and Inf. Syst. Educ. J. **4**(103) 2006. http://isedj.org/4/103/
49. Morien, R.: Agile and agility in computer systems education. In: KSII The 7th International Conference on Internet ICONI'2015, Kuala Lumpur, Malaysia, Dec 2015
50. Morien, R.: An agile project management manifesto—A reference disciplines framework for agile development. In: KSII The 5th International Conference on Internet (ICONI) 2013, Pattaya, Thailand, Dec 2015
51. Senge, P.: The Fifth Discipline—The Art & Practice of the Learning Organization. Currency Doubleday (1990)
52. Hock, D.: Birth of the Chaordic Age. Visa International (1999)
53. Champy, J.: Reengineering management: The Mandate for New Leadership. Harper-Collins Publishers (2005)
54. Liker, J.K.: The Toyota Way. McGraw-Hill (2004)
55. Rubinstein, M.F., Firstenberg, I.R.: The Minding Organization. Wiley (1999)
56. Gross, J.M., McInnis, K.R.: Kanban Made Simple. American Management Association (2003)
57. Katzenbach, J.R., Smith, D.K.: The Wisdom of Teams: Creating the High-Performance Organization. Harvard Business School (2008)
58. Surowiecki, J.: The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. Random House (2005)
59. The wisdom of the herd. http://en.wikipedia.org/wiki/Collective_wisdom. Accessed 31 May 2010
60. Wide-Band Delphi. http://en.wikipedia.org/wiki/Wideband_delphi. Accessed 31 May 2010
61. Womack, J.P., Jones, D.T.: Lean Thinking: Banish Waste and Create Wealth in Your Corporation. BE—Business & Economics, vol. 2553. Simon and Schuster (2003)