Cloud and Crowd Based Learning

Chun-Hsiung Tseng, Ching-Lien Huang, Yung-Hui Chen, Chu-Chun Chuang, Han-Ci Syu, Yan-Ru Jiang, Fang-Chi Tsai, Pin-Yu Su and Jun-Yan Chen

Abstract Speaking of new teaching methodology, "Flipped Classroom" is undoubtedly a very popular one. The basic concept of flipped classroom is to have students learn by themselves before attending a "real" class at school. Once the background learning stage is performed outside of the class time, tutors have free time to lead students to participate in higher-order thinking. However, as shown in the report of Katie Ash, the performance of the flipped classroom method is in fact still arguable. Our survey shows that the contents offered by most modern e-learning systems are relatively static. Consider how fast new information appeared on the Web! Of course, teachers, or material providers, can upload new contents to e-learning systems. However, creating contents requires efforts. Today, work load of our teachers is already heavy, so expecting teachers to update contents very frequently is not practical. The researchers believe that one of the major challenges faced by e-learning systems today is the richness of contents.

Keywords e-learning · Crowd-sourcing

C.-H. Tseng · C.-C. Chuang · H.-C. Syu · Y.-R. Jiang · F.-C. Tsai · P.-Y. Su · J.-Y. Chen
Department of Information Management,
Nanhua University, Dalin Township, Taiwan, ROC
e-mail: lendle_tseng@seed.net.tw

C.-L. Huang
Department of Industrial Management,
Lunghwa University of Science and Technology, Taoyuan 33306, Taiwan, ROC
e-mail: lynne@mail.lhu.edu.tw

Y.-H. Chen (✉)
Department of Computer Information and Network Engineering,
Lunghwa University of Science and Technology, Taoyuan 33306, Taiwan, ROC
e-mail: cyh@mail.lhu.edu.tw

J.C. Hung et al. (eds.), *Frontier Computing*, Lecture Notes
in Electrical Engineering 375, DOI 10.1007/978-981-10-0539-8_1

1 Introduction

Speaking of new teaching methodology, "Flipped Classroom" is undoubtedly a very popular one. The basic concept of flipped classroom is to have students learn by themselves before attending a "real" class at school. Once the background learning stage is performed outside of the class time, tutors have free time to lead students to participate in higher-order thinking. Typically, online materials such as videos and slides will be used in the background learning stage. To make the system effective, in most cases, the materials should be accessible from the Web. However, as shown in the report of Ash [1], the performance of the flipped classroom method is in fact still arguable. How is the quality of the materials? How active students are in the stage? How good is the performance evaluation method? Is there a reasonably-designed feedback system? Do practical tools provided for couches? These are all important factors affecting whether a flipped classroom system is successful or not. Our survey shows that the contents offered by most modern e-learning systems are relatively static. Consider how fast new information appeared on the Web! Of course, teachers, or material providers, can upload new contents to e-learning systems. However, creating contents requires efforts. Today, work load of our teachers is already heavy, so expecting teachers to update contents very frequently is not practical.

The challenges pointed out above are not freshly new in the Web 2.0 age. We have an explosive amount of information. It is beyond the capabilities of traditional material providers to always keep their material up-to-date. At the very beginning of the Web 2.0 age, the issue is solved with crowd sourcing. That is, instead of relying on few material providers to update their Web sites or their blogs, we simply allow everyone to become material providers. However, crowd sourcing is only a partial cure. There are already a few on-line knowledge-sharing Web sites utilizing crowd sourcing technologies. Among them, "Yahoo Answers" is a well-known example. To the best of the researchers' memory, there is no strict study for this, however, generally speaking, "Yahoo Answers" is not treated as an e-learning Web site. A possible reason is, materials on "Yahoo Answers" are too diverse and not strictly-organized.

The situation motivates this research. Looking at the enormous amount of information on the Web, the researchers wonder how to leverage the information in e-learning systems. Simply automatically collecting "similar" contents together will make little contribution to learning due to the diversity of the Web. In this research, the goal is to propose a module that can facilitate the following functionalities:

1. be automatic
2. generate structured information
3. take advantage of the crowd sourcing technology
4. adopt the cloud technologies

2 Related Works

Although Web search engines today are usually considered efficient, in some circumstances, they are not, especially when semantics and human intelligence are of concern. Some queries simply cannot be answered by machines alone. In such cases, human input is required [2]. The research field is typically named as crowd search or crowd searching. It is not an easy task to mediate between responses from human beings and search engines, and thus the research field is very challenging.

Crowd search is highly related with social networking [3]. The opinions collected within friends and expert/local communities can be ultimately helpful for the search task. For example, the question "find all images that satisfy a given set of properties" can be difficult for machines to proceed, but with the help of human intelligence, answering the question becomes simpler [4]. A special query interface that let users pose questions and explore results spanning over multiple sources was proposed in [5]. Another type of crowd search and crowd sourcing is social bookmarking. As shown in Heymann's research work [6], social bookmarking is a recent phenomenon which has the potential to give us a great deal of data about pages on the web.

Various crowd search and crowd sourcing systems have been proposed. For example, the research of Parameswaran proposed a human intelligence-based methodology for solving the human-assisted graph search problem [7]. Amazon's Amazon Mechanical Turk is a commercial product that enables computer programmers (known as Requesters) to co-ordinate the use of human intelligence to perform tasks that computers are currently unable to do [8].

3 Cloud and Crowd Based Learning Module

The proposed system is separated into several sub modules: the data processing sub module, the data provider sub module, and the feedback processing sub module. Figure 1 depicts the overview of the system.

The data processing sub module is responsible for collecting data from the Web and extracting information from them. Data collected from the Web is full of noises and is unstructured. Interpreting and reusing unstructured data is difficult. The sub module will extract structured data according to some pre-defined rules from it. The sub module is based on the researchers' previous work: the Object-Oriented Schema Model (OOSM). The main issue to be addressed by this sub module is the unstructured nature of the Web. OOSM is in fact a grammar model. Here, we emphasize a database-centric design. Why is the concept of database important? The content of the Web is simply too diverse and ambiguous. To make information extracted from the Web usable for learning, we have to at first make it structured and thus it can be read and processed by applications easily. Databases are certainly

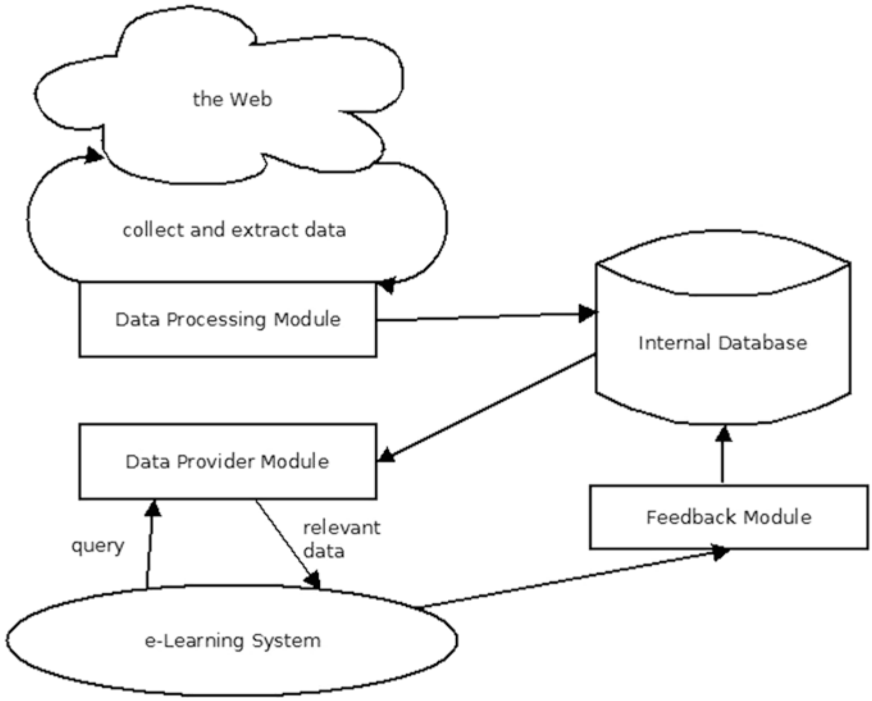

Fig. 1 System overview

structured information sources. OOSM contains three components: the schema model, the mapper tool, and the database.

The data provider sub module provides two sets of utilities: query utilities and transformation utilities. In most cases, e-learning systems interact with the data provider sub module to acquire materials related to their contents. For the purpose, the data provider sub module defines the following function:

$$query(db\_namespace, db\_localname, criteria, sortby)$$

Here, db_namespace and db_localname are used for pointing to a specific database defined in the data processing sub module. The data provider sub module allows the following types of criterias:

1. static id: each entry extracted by the data processing sub module will be automatically assigned an unique id; e-learning systems can utilize this id to make a static link
2. null: by specifying null in this field, the data provider sub module will not perform any filtering
3. attribute filter: a JSON-style query language will be developed for detailed querying

Furthermore, without a solid feedback system, it will be impossible to evaluate the effectiveness of the proposed system. The feedback sub module offers two types of feedbacks:

1. direct feedbacks: collected from users by requiring users to submit rating information directly
2. indirect feedbacks: collected by analyzing user behaviors such as the click sequences

For direct feedbacks, simple functions are defined:

$$rate(domain, db\_namespace, db\_localname, contentId, score)$$
$$tag(domain, db\_namespace, db\_localname, contentId, tags)$$

As the names suggest, the functions are used for providing ratings and tags (labels) to content provided by a database. Scores give an overview of qualities of contents. By calculating average scores of contained contents, we can also have scores for databases. Tags reveal the characteristics of contents.

Indirect feedbacks are acquired through the analysis of click logs and tags. Click logs record user behavior when they are reading contents. How long did they stay on a specific content? How many contents are read before they leave? Which content is most frequently read after a user reads a specific content? Analyzing these tells us the popularity and potential (implicit) relationships of contents. On the other hand, tags represent users' subjective feeling about content. Roughly speaking, contents with the tags have something in common. Certainly, ambiguities can cause problems. An example is, without contextual information, we can never make sure of the correct meaning of the word "apple". Though completely getting rid of ambiguities is almost impossible, technologies such as word stemming can be used to alleviate the issue. Additionally, after collecting a proper amount of tags, grouping tags into clusters can help determine the similarities between tags and thus in turn reduce the level of ambiguities. Feedbacks are used for calculating scores of Web resources according to the following functions:

$$score\ of\ content(C) : \alpha \times direct\ scores + \beta \times indirect\ scores, \alpha + \beta = 1$$
$$score\ of\ database(D) : \frac{\sum_{i=1}^{n} \rho_i C_i}{number\_of\_contents\_in\_the\_database}, \rho = relative\ popularity$$

4 Conclusions and Future Work

In this manuscript, the concept of a cloud and crowd based learning module is proposed. The proposed mechanism is aimed at solving the lack of content of existing e-learning systems. The proposed method utilizes crowd intelligence to collect related materials from the Web. Besides, to lower the complexities of adopting the proposed method, we separate the concept into several sub modules and propose a cloud-based deployment structure. For now, the implementation is still in its prototyping stage. In the future, the following goals are set:

1. complete the implementation
2. integrate the implementation with an e-learning system for evaluation
3. develop a sound evaluation matrix

References

1. Ash K (2012) Educators evaluate 'flipped classrooms'. http://www.edweek.org/ew/articles/2012/08/29/02el-flipped.h32.html
2. Franklin J, Kossmann D, Kraska T, Ramesh S, Xin R (2011) CrowdDB: answering queries with crowdsourcing. In: Proceedings of the 2011 ACM SIGMOD international conference on management of data, pp 61–72
3. Bozzon A, Brambilla M, Ceri S (2012) Answering search queries with CrowdSearcher. In: Proceedings of the 21st international conference on world wide web, pp 1009–1018
4. Parameswaran A, Garcia-Molina H, Park H, Polyzotis N, Ramesh A, Widom J (2012) CrowdScreen: algorithms for filtering data with humans. In: Proceedings of the 2012 ACM SIGMOD international conference on management of data, pp 361–372
5. Bozzon A, Brambilla M, Ceri S, Fraternali P (2010) Liquid query: multi-domain exploratory search on the web. In: Proceedings of the 19th international conference on world wide web, pp 161–170
6. Heymann P, Koutrika G, Garcia H (2008) Can social bookmarking improve web search? In: Proceedings of the 2008 international conference on web search and data mining, pp 195–206
7. Parameswaran A, Sarma A, Garcia-Molina H, Polyzotis N, Widom J (2011) Human-assisted graph search: it's okay to ask questions. In: The proceedings of the VLDB endowment vol 4, pp 267–278
8. Amazon, https://www.mturk.com/mturk/welcome