An Idealet-Centric Scheme for Large Scale Open Innovation Systems

Matti Hämäläinen[1], Qing Li[2], Zhangxi Lin[3], Alin Tomoiaga[3], and Jia Wang[2]

[1] Aalto University
[2] Southwestern University of Finance and Economics
[3] Texas Tech University
matti.hamalainen@tkk.fi, liq_t@swufe.edu.cn,
{zhangxi.lin,alin.tomoiaga}@ttu.edu, xiaojiajia198796@sina.com

Abstract. This paper intends to demonstrate how open innovation systems could be developed by tackling the challenging knowledge management problems that are encountered when aiming at involving very large audiences. This is the case when generalizing open innovation approach beyond companies to a wider societal context like in the case of national innovation systems. The Open Innovation Banking System (OIBS) project, funded by the European Social Fund (ESF) and the participating higher education institutions in Finland, is used as a basis for our discussion. It specifically aims at bringing the largely underutilized creativity of students and senior citizens to play. Among several technologies to develop OIBS, mashups as hybrid web applications can play an important role in such constantly evolving system and contents. However, relying only on unstructured text inputs, the services of textual content sharing for OIBS would require intelligent text processing that far exceeds the capability of such applications. In this paper, we propose an "idealet"-centric solution for representing the data submitted by users, enabling concise description, refinement and linking of ideas as input for innovation processes. An idealet is defined as the core knowledge about an innovative idea. The relationships among idealets and essays can be represented in a semantic network in terms of their relationships. This scheme allows the mashup applications for OIBS to more effectively retrieve, process, extract, and deliver the most important knowledge from an ocean of information contributed by participating information composer, reviewers, and users. The paper also discusses how the idealet-centric approach can be employed for a functional open innovation system.

Keywords: Open Innovation, Mashup, Semantic Network, Knowledge Modeling, "Wicked" Problems.

1 Introduction

The idea of open innovation has gained exceptional momentum with web 2.0 enabled social media and collaborative services and has quickly gained popularity since it was promoted by Henry Chesbrough in 2003 [2]. The central idea behind open innovation is that in a world of widely distributed knowledge, companies cannot afford to

R. Sharman, T.S. Raghu, and H.R. Rao (Eds.): WEB 2009, LNBIP 52, pp. 19–29, 2010.

rely entirely on their own research, but should instead buy or license processes or inventions (e.g. patents) from other companies. In addition, internal inventions not being used in a firm's business should be taken outside the company, for example, through licensing, joint ventures, or spin-offs. While Chesbrough originally focused on firm's perspective the open innovation approach has potential to foster solution of many of today's challenging problems - or so called "wicked problems" - when generalized from organizations to wider societal context and supported by appropriate discourse mechanisms. Here a "wicked problem" is referred to as "to describe a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize. Moreover, because of complex interdependencies, the effort to solve one aspect of a wicked problem may reveal or create other problems." [12].

This paper is intended to demonstrate how such systems could be developed by tackling the knowledge management problems in Open Innovation Banking System (OIBS) project, an EU sponsored open innovation system project in Finland aiming at bringing the largely underutilized creativity of the students and senior citizens to play [13]. Specifically, OIBS is a kind of textual contents sharing system, which is similar to Wikipedia in its open content service [14] but with the focus on innovative ideas contributed by the users.

When applying the open content approach to the development of the idea bank system – not just to the process of producing its content - it becomes important to support approaches that enable very easy and lightweight linking of various value adding services as well as composition of new services. The mashup approach [10] can be applied in the ongoing development and extension of such a system by providing support for analyzing information and evaluating contributions, for identifying related items and supporting composition, and for ranking, visualizing, and various other forms of processing. There are a number of most well-known mashup implementations, such as Microsoft Popfly (discontinued on August 24, 2009), Google Mashup Editor (Deprecated, since January 14, 2009), Mozilla Ubiquity, Yahoo! Pipes, and Open Mashup Alliance. However, relying only on unstructured text inputs, these services for typical idea banking and brainstorming support systems would require intelligent text processing that far exceeds the capability of such applications. In the last couple of decades great efforts have been exerted to develop various techniques, methods, and algorithms for automatically processing and summarizing unstructured text documents, but there is no practically feasible resolution that is applicable to our specific application context. The challenging problems include:

- How to effectively and efficiently manage a large amount of textual contents, and
- How to support composition, refinement and more structured processing and discourse on the submitted contents.

To address these problems, we suggest a solution, *idealet-centric* knowledge modeling, for structuring the information elements (i.e. inputs and possible outputs of the mashups) in a more meaningful way, so that the information can be better understood and processed. The structuring also supports discourse, idea refinement, and composition processes for providing input for the innovation process. The analogous problem of supporting of scientific collaboration has been earlier addressed by proposing a structured discourse model [5] based on the IBIS-model [12]. With the current pervasive

Internet connectivity and social network systems even very large scale of introduction of such approach has become viable. An *idealet* is defined as the core knowledge about a piece of innovative idea. Based on the idealet-centric model, we show how the essays contributed by the users of an open innovation system can be processed incorporating with the information retrieval techniques, such as the latent Dirichlet allocation (LDA) model [1]. Then a semantic mashup approach using the latest text snippet extraction technique can be employed to extract innovative ideas. This scheme allows the mashup applications for OIBS to more effectively process the most important knowledge from an ocean of information contributed by participating information composer, reviewers, and users. In this way, the suggested idealet concept and structure as a lightweight knowledge representation scheme that can be applied in Web environment both from technology and from user perspectives thus promoting scalability and adoption of open innovation systems. The idealet model and tools making use of it is suggested to provide value both to idea contributors and to participants looking to further processing of ideas in the innovation process. Thus, we can expect it to help boosting the initial adoption and scaling of use of OIBS and other such systems, which is typically a major challenge in their successful introduction.

2 Challenges in OIBS - Mashing Up the Innovative Ideas

The OIBS concept is based on the idea that Online Social Networks can be utilized as a critical part of National Innovation System i.e. National Open Innovation Systems (NOIS) as illustrated in Fig. 1 that presents the general Innovation Triangle framework which consolidates NOIS [13].

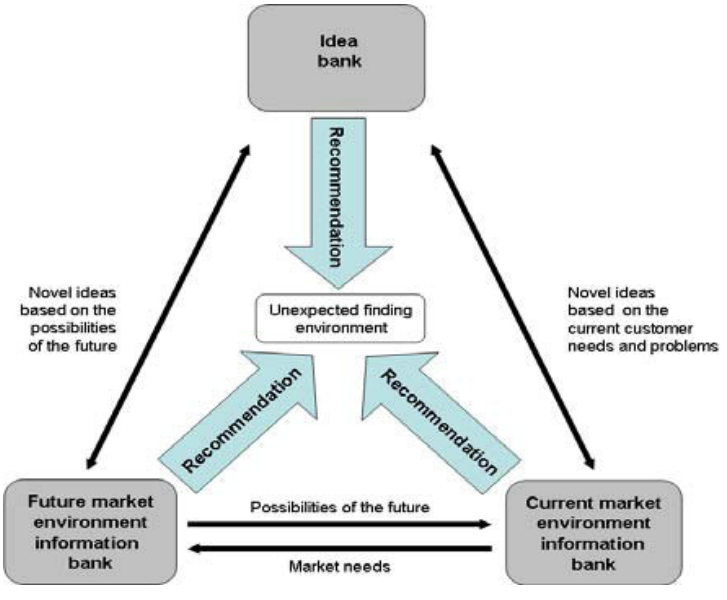

Fig. 1. General Innovation Triangle framework [13]

With the aim of generating new ideas (i.e. the top box) NOIS framework includes two complementary innovation sources: first, future market environment information, presenting visions of the future (i.e. the left box) and second, current market environment information, presenting today's challenges (i.e. the right box). By combining these, the approaches of collaborative content production and intelligent content recommendation will significantly boost the possibilities of unexpected findings, which have been identified as a major innovation source. It has also been recognized by [13] and others that a successful innovation system must include options for rewards for all key actors.

The OIBS project launched in 2008 aims at creating an environment that supports the OIBS approach and takes it into large scale operational use starting from a network of higher education institutions. The challenges in the development of such a system are described in [6]. The intended widespread use of OIBS requires that the system should be easy to learn and use. Still it should have very powerful features for the management, processing and presentation of huge amounts of fairly unstructured data. Once the user community gets large, the OIBS system will contain huge amounts of valuable pieces of information just waiting to be combined intelligently to form innovative solutions to the needs of companies and the society. The technology development subproject adopts the open approach that is basis for the content creation. It is based on enabling development of the web based applications collaboratively and has made the technology entirely open source. Specifically, the potential of mash-up type of approach has been recognized for supporting community innovation. As a relatively new concept, the project has a challenge to attract developers and content contributors.

Among several technologies to develop OIBS, mashups as hybrid web applications play an important role in the system as they can combine contents from multiple web sites into a single application. They have become popular because of the emphasis on interactive user participation in which they aggregate and stitch together the data from multiple sources. However, before mashups can make the transition from cool toys to sophisticated applications, much work will have to go into distilling robust standards, protocols, models, and toolkits. Due to complexity of the input information, most of current mashup applications have limited functions and majority of them are based on Google map applications [11]. Therefore, they are incapable of aggregating a large amount of textual information.

3 The Idealet-Centric Knowledge Modeling Scheme

3.1 The General Idea

We suggest a knowledge structuring approach - the "idealet" that can help boost the open development approach of both the environment itself as well as the content of OIBS. Idealets can represent the knowledge embedded in the previously unstructured text inputs at least semi-structured with the focus on the innovative components. In this way, they allow new functionalities and services, like intelligent text analysis, incentive mechanisms and reward systems to be developed and plugged on the core OIBS services. In addition, by incorporating the text mining techniques into mashup

applications based on the semantic representation of the input information, the idealet-centric modeling scheme will enrich the knowledge being processed. The basic mechanism for semantics are provided using the "idealet" attributes and can be enriched with user defined tags.

Our model is expected to increase the dynamics of the individual's creativity by creating an online environment where a conventional habit is easily exceeded. The collaborative content production and intelligent content recommendation approach together will significantly boost the possibilities of unexpected findings which are a major innovation source. We conceive a new people-to-people interaction based approach to support the innovation system. This way, people with fresh ideas and people with practical knowledge can effectively combine their forces in an open innovation based social networking community.

To better illustrate the knowledge modeling scheme we narrow down our discussion to open textual content systems, where the pieces of textual contents produced by users are named as "essays". These essays are short articles that may contain limited multimedia information that are associated to some part of text information. As the basic inputs, essays are contributed by content composers – a subcategory of users. The idea behind this phenomenon is that the content users may take the role of content composers if they want to contribute. This kind of user-driven services has been studied and advocated recently [4].

The essays in the open innovation system are short articles consisting of one or multiple idealets. Fig. 2 shows the structure of such a system for managing shared textual contents. The essays in the database are converted into knowledge via a semantic network, which appear as the innovative ideas at the user end. There are three user interfaces. The Innovation Services interface is the main one to content users, which allows content users to create, edit, retrieve, summarize, aggregate, and market essays. The System Admin interface is used to manage the system and services by the

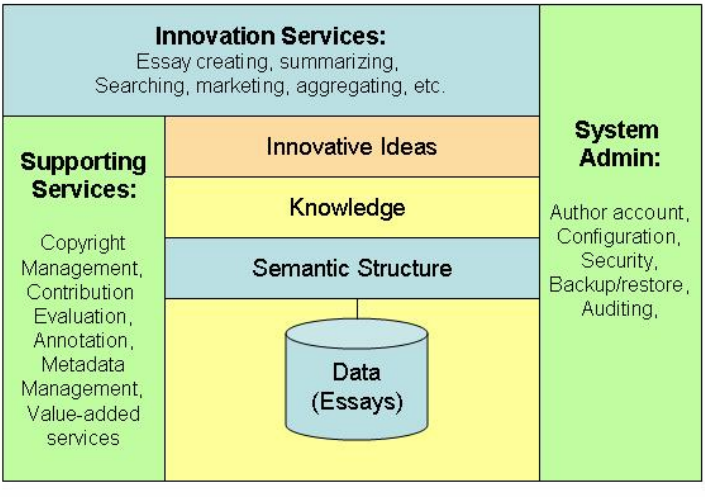

Fig. 2. The logic structure of an open innovation system with textual content services

system administrators. The Supporting Services interface provides advanced services, such as copyright management, essay evaluation, annotation, user-configurable services, etc., mainly for those advanced or power users. The idealet-centric approach can help composers to structure their essays in the way that will ease the access of them. Most importantly, the essays organized in idealets with relationships among them are more informative for mashups to aggregate innovative contents in better quality to users.

3.2 Knowledge Modeling with Idealets

A semantic network based knowledge modeling approach for idealets can be applied to the open innovation system with a specific focus on the identified innovative contents. Idealets are then associated to each other with regard to their categories and attribute values in conjunction with the relationship among essays. The attributes also enable implementation of relevant models of structured discourse for argumentation-based approaches [5].

When users input their essays they will be advised to input key information of the essay as its attributes, such as the theme of the essay, keywords, references, category, type of the essay, etc. Some of the attributes are generic and some are specific to the context and user group (e.g. in the context of OIBS this ranges from course assignment with students to open ideation for thematic area, or locally or situationally relevant issue, with elderly people). Importantly, inputting the attributes happens as natural part or side effect of the process minimizing the user efforts. These attributes obviously enrich the input information and will serve as a good source for the applications to mash multiple inputs up into a piece of new essay. Since an essay may contain more than one idea, idealets then become the fundamental components in a semantic network formed by essays. An example of semantic network for the relationships among idealets is shown in Fig. 3. There are two major types of relationships among idealets – 1:1 Relation and 1:M Relation. Examples of 1:1 Relations are "extended from", "referred to", and "differentiated from"; examples of 1:M Relations include "multi-referred to", "integrated from", and so on. In addition to the relationships among idealets that represent the major knowledge structure of the essays, the relationships among essays also form a semantic network at a higher level. The relationships among essays can be inherited by the idealets within the essays. Also, they help discovery of relations among seemingly unrelated essays (e.g. as described above on in the OIBS use cases).

It is worth noting that this idealet-centric semantic network is a customized application of the semantic network method to meet the specific requirements of open innovation systems. There has been abundant research in extending the classical semantic network knowledge representation scheme that was very popular in the artificial intelligence research area. Recent relevant research efforts include those in knowledge discovery and semantic web, for example, personalized knowledge recommendation [9], the application in discovering knowledge from computer-mediated discussions [7], and automatic web content archiving [3].

The main purpose of an open innovation system is to allow users to exchange their ideas and generate input for innovation process. This involves knowledge creation, dissemination, integration, and refinement. Naturally, idealets become the elemental

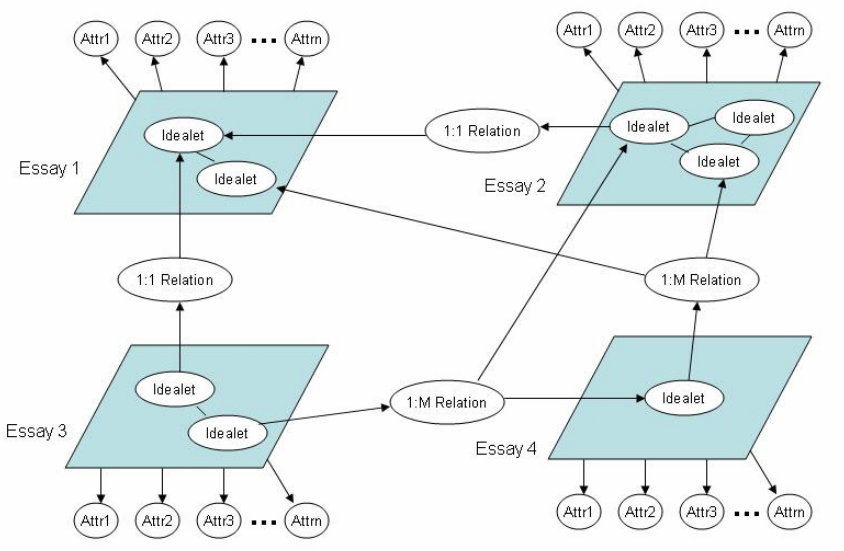

Fig. 3. The semantic network at the idealet level

unit of the knowledge in the semantic network. The essay is the carrier of idealets and predetermines its idealets' attributes. In this way, the two-level semantic network becomes the basic form of the idealet-centric knowledge model.

Idealets in an essay are normally tied to each other closely as they are tightly integrated together by a stream of textual contents. Theoretically, they can be identified in terms of latent Dirichlet allocation (LDA) [1], which stems from the Dirichlet distribution, a family of continuous multivariate probability distributions characterized by a vector of positive real values. LDA allows sets of observations explained by unobserved groups which explain why some parts of the data are similar. Once the vector of terms with their frequencies is generated from an essay, the terms can be partitioned into subgroups according to the number predefined. In this way, the outcome of an essay processed by the LDA model will be a set of idealets, each being associated with a term vector that comes with the probabilities for each term. To ensure that the automatically created idealets are meaningful, the results of this process may be verified by the one who submitted the essay – or by other members of the community. Thus, this combined process would provide substantial boost in idea processing, refinement and linking.

Our application of semantic network approach differs from others in that we focus on the level of idealet and the formation of the semantic network based on combining it with the web 2.0 approach. The better quality the description of the idealet is, the more comprehensive the semantic network is, and the more likely the innovative idea will be adopted. Therefore, essay composers have incentive to conceive their essay in a way making use of the idealet-centric knowledge representation scheme. In the best case, an essay composer can readily outline each idea with detailed information. By default, there is only one idealet in the essay but the composer may specify number of idealets. In summary, the idealets are identified by two approaches, either defined by

essay composers or generated by the LDA model. Each idealet is characterized by a vector of terms, each of the terms having a probabilistic value.

3.3 Idealet-Based Textual Information Processing

As the ideas are described by text (and optional media elements) it is important to support textual information processing in addition to processing of the structured data elements. Below we describe the basic textual information processing functions for an open innovation system idea bank:

- Idealet search
  Since every idealet is characterized in a weighted vector of terms, the search for the most matched idealets is a calculation process of semantic distance or cosine-similarity between the query and the candidate idealets. This can be seen an extension from the previous semantic network based search, which relies on the semantic links among idealets.
- Idealet comparison
  The "innovativeness" is an important criterion to justify the contribution of an essay. The weighted vector of terms for idealets allows the calculation of the distance between two idealets to determine the extent of innovativeness - in case we choose to use such a criterion as a measure of "innovativeness". Again, this automatic processing is meant as input for final judgment to be done by human users, and it remains to be studied what would be the appropriate criteria in various cases.
- Idealet aggregation
  Based on the enriched inputs of textual information in the idealet semantic network, we will apply text snippet extraction technique to reconstruct an essay based on many idealets [8]. Text snippet extraction is a special text processing technique based on language statistical model to identify certain personalized patterns in a textual document for the applications such as web search engine construction, customized text summarization and automated question answering. Li and Chen proposed that text snippet extraction can be generalized if a user's intention is utilized. This is achieved by constructing and using statistical language models which effectively capture the commonalities between a document and the user intention. In particular, this approach first employs a document filter to retrieve a set of documents from the corpus that are relevant to a profile representing a user's intention. Since this step is orthogonal to the subsequent text snippet extraction, any good filters can be applied to retrieve documents. Typically, a document filter conducts a certain similarity comparison between the profile and a subset of documents in the corpus which contain at least part of the profile. Next, a pair of statistical language models is constructed. One pertinent to the user's intention and the other is independent of this intention. Using these two models, two probabilities of each word in the relevant document being generated by either model are calculated. In this approach, a relevant document is treated as a sequence of words, each associated with these two probabilities. Thus, from every relevant document, one or more snippets that are deemed to best satisfy the user's intention can be extracted. In this way, the system can overcome the rigidness of existing approaches by dynamically returning more flexible start-end positions of text snippets, and is semantically coherent.

When a user enters a string of keywords to summarize relevant ideas proposed by others, the system will do the following steps based on the idealet-based knowledge representation:

- Step 1: retrieve a number of relevant essays and rank them in accordance with the semantic distance from the query
- Step 2: use a configurable cut-off level to keep those most relevant essays for further processing
- Step 3: these top-N relevant essays are treated as one document and fed into the text snippet extraction engine [8] to construct a new condensed essay.

This is in fact the main idea of text mining based mashup applications.

4 Implementing an Open Innovation System

Based on the idealet-centric approach, an open innovation system for text information sharing is depicted in Fig. 4. The idealet-centric data model is built upon a network infrastructure, normally the Internet or an intranet. These idealets organized by the model are constructive blocks for open innovation services delivered by semantic mashups, and technically implemented in the advanced methods, such as text summarization. A series of services may be configured to form certain business processes upon needs.

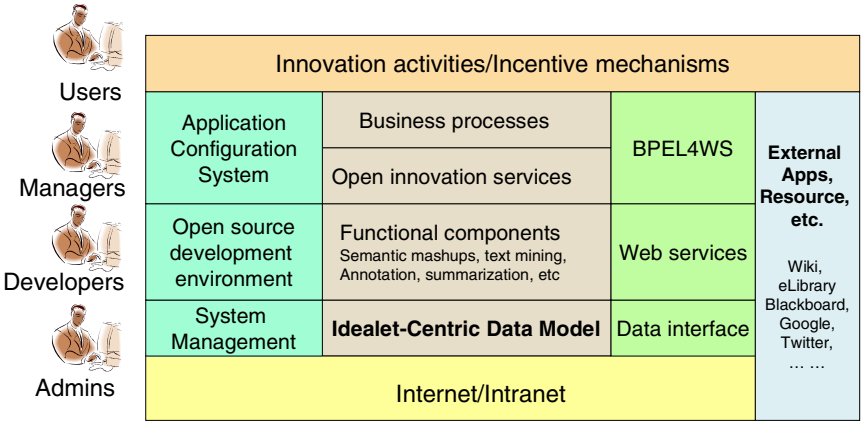

Fig. 4. The anatomy of idealet-centric open innovation system

There are four types of users: system administrators (Admins), who support the operation of the system, service developers (Developers), who provide the evolving functions of the system to users, project managers (Managers), who are leaders of specific projects which are configurable upon needs, and normal users (Users), who contribute the knowledge to the open innovation system. Users of the system could play different roles, including idealet contributor, idealet reviewer, and idealet follower.

To facilitate different needs of the open innovation system, a number of utilities must be available. In the left column of the system diagram in Fig. 4, the Application

Configuration System is for project managers to define their projects, the Open Source Development Environment is for service developers to enhance the functions of the system, and the System Management module is dedicated for system administration purposes.

There must be several incentive-compatible mechanisms to guarantee the effective operation of the system. For example, the following are the most important incentive-compatible mechanisms for an open innovation system:

- The incentive for users to contribute idealets. The idealet contributors should not be bothered by the copyright problem. Their contributions to the system must be clearly identified and properly evaluated.
- The incentive for idealet contributors to input the contents in the way that can better organize the text information for further processing.
- The incentive for users to follow the innovative ideas available in the system. Obviously, the number of followers can well indicate the value of an idealet or a set of idealets.
- The mechanism to promote high quality idealets that will attract better followers and keep the thread moving forward. For this purpose, some users of the system will be playing the role of reviewers of the idealets. Their contributions must be accounted and encouraged.

5 Concluding Remarks

This paper introduces the main ideas on how to implement an idealet-centric knowledge representation scheme and how this scheme can be used for semantic mashups. In particular, we demonstrate how the scheme would help in the development of very large scale open innovation systems by tackling the challenging knowledge management problems encountered when involving wide and diverse audiences. The combination of the idealet model and semantic mashup tools making use of both structured and textual information processing are expected to provide value to idea contributors as well as to those looking to their further processing as inputs for the innovation processes. This should help boosting the initial adoption and scaling of use of OIBS and other such systems, which is typically a major challenge in their successful introduction. We may expect that successful development of very large scale open innovation support systems would require the described approach of combining automatic structured and textual content processing with the individual and community judgment and efforts. To take this approach further, the research tasks, among others, include the development of content aggregation algorithm, and experiments using text corpus.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. Journal of Machine Learning Research (3), 993–1022 (2003)
2. Chesbrough, H.W.: Open Innovation: The new imperative for creating and profiting from technology. Harvard Business School Press, Boston (2003)

3. Du, T.C.: Building an automatic e-tendering system on the Semantic Web. Decision Suppor Systems 47(1), 13–21 (2009)
4. Hamalainen, M.: Crowdsourcing and a Living Lab on Campus: enabling research in user driven service innovation. Presented in IBM Academic Days, Zurich, Switzerland (May 13-14, 2008)
5. Hamalainen, M., Hashim, S., Holsapple, C., Whinston, Y.: Structured Discourse for Scientific Collaboration: a Framework for Scientific Collaboration Based on Structured Discourse Analysis. Journal of Organizational Computing and Electronic Commerce 2, 1–26 (1992)
6. Heino, P., Pippola, T., Peltonen, J., Piispanen, P.: Innovative development of an environment for innovations. In: Proceedings of the Beyond the Dawn of Innovation 2009 Conference, Laurea University of Applied Sciences, Finland, June 15 – 17, pp. 256–260 (2009)
7. Khalifa, M., Liu, V.: Semantic network representation of computer-mediated discussions: Conceptual facilitation form and knowledge acquisition. Omega 36(2), 252–266 (2008)
8. Li, Q., Chen, Y.P.: Personalized Text Snippet Extraction Using Statistical Language Models. Pattern Recognition (2009) (in press)
9. Liang, T.P., Yang, Y.F., Chen, D.N., Ku, Y.C.: A semantic-expansion approach to personalized knowledge recommendation. Decision Support Systems 45(3), 401–412 (2008)
10. Palfrey, J., Gasser, U.: Case Study - Mashups Interoperability and eInnovation. Berkman Publication Series (November 2007), http://cyber.law.harvard.edu/interop/pdfs/interop-mashups.pdf (retrieved on September 10, 2009)
11. Reamy, T.: Mashup Mindset Moving Mashups to Next Level. Knowledge Architecture Professional Services, http://www.kapsgroup.com/presentations/Mashup%20Mindset.ppt (retrieved on June 1, 2009)
12. Rittel, H.W., Webber, M.M.: Dilemmas in a General Theory of Planning. Policy Sciences (4), 155–169 (1973); Reprinted in Cross, N. (ed.): Developments in Design Methodology, pp. 135–144. J. Wiley & Sons, Chichester (1984)
13. Santonen, T., Kaivo-Oja, J., Antikainen, M.: National Open, Innovation System (NOIS): Defining a Solid Reward Model for NOIS. In: Huizingh, K.R.E., Torkkeli, M., Conn, S., Bitran, I. (eds.) Proc. of the 1st ISPIM Innovation Symposium - Managing Innovation in a Connected World, Singapore, December 14-17 (2008)
14. Viégas, F.B., Wattenberg, M., McKeon, M.M.: The Hidden Order of Wikipedia. IBM Watson Research Center, Technical Report #07-12 (2007)