Information Processing and Management 50 (2014) 634–652

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

A recommender mechanism for social knowledge navigation in an online encyclopedia

Yung-Ming Li *, Lien-Fa Lin, Yu-Hui Lin

Institute of Information Management, National Chiao Tung University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Received 8 April 2013
Received in revised form 17 April 2014
Accepted 28 April 2014
Available online 29 May 2014

Keywords:
Social media
Online encyclopedia
Knowledge-sharing
Knowledge ontology
Recommender systems

ABSTRACT

In today's world, knowledge is important for constructing core competitive advantages for individuals and organizations. Recently, Web 2.0 applications and social media have provided a convenient medium for people to share knowledge over the Internet. However, the huge amount of created knowledge can also leads to the problem of information overload. This research proposes a social knowledge navigation mechanism that utilizes the techniques of relevant knowledge network construction, knowledge importance analysis, and knowledge concept ontology construction to generate a visualized recommendation of a knowledge map of sub-concept and knowledge of an article reading sequence for supporting learning activities related to a free online encyclopedia. The results of experiments conducted on Wikipedia show that the proposed mechanism can effectively recommend useful articles and improve a knowledge seeker's learning effectiveness.

1. Introduction

People exchange information and experiences to generate and enrich their personal knowledge. Knowledge is important for constructing core competitive advantages for individuals and organizations in today's world (Foss & Pedersen, 2002). In recent years, Web 2.0 applications and social media have provided a convenient environment for people to share their personal knowledge on the Internet. Many people and enterprises also enrich their own knowledge by means of this convenient environment (Cole, 2008).

Online collaborative encyclopedias are one of the most famous Web 2.0 applications for knowledge sharing (Leshed, Haber, Matthews, & Lau, 2008), and are well used by individuals and companies. For instance, Pfizer's pfizerpedia, which is a wiki used in Pfizer's research, provides an environment for the company's employees and partners to build knowledge database (Libert & Spector, 2007). The Social Intranet Study 2011 notes that social media such as blogs, wikis and other tools appear on most of the corporate intranets: 61% have at least one tool available to some or all employees. Among the reasons for the popularity of online encyclopedias is that they can be edited frequently so that they remain up to date (wiseGEEK, 2012), and that the articles are of good quality (Giacomo, 2008). The shared content is also a good resource for knowledge learning.

There are many online encyclopedias that provide a lot of knowledge, and people seek new knowledge from them every day. For instance, Baidu Baike is a Chinese collaborative online encyclopedia with more than 4.9 million articles (Baidu Baike, 2012), surpassing Chinese Wikipedia in this respect. Hudong is the largest Chinese encyclopedia, news, and neologism website in the world. It is the largest wiki site in China with over 6.4 million articles and more than 4.5 million volunteers (Hudong, 2012). It is ranked 55 in China, and most of its readers are college-educated or browse this site at school (Alexa,

* Corresponding author. Tel.: +886 3 5712121.
E-mail addresses: yml@mail.nctu.edu.tw (Y.-M. Li), lienfa0704@gmail.com (L.-F. Lin), clc0330@gmail.com (Y.-H. Lin).

http://dx.doi.org/10.1016/j.ipm.2014.04.006
0306-4573/

2012a). Wikipedia is a free encyclopedia with more than twenty two million articles. According to Alexa's survey in 2012, it is ranked sixth in the world and its readers are often users who browse at work and school (Alexa, 2012b). This indicates the importance of online encyclopedias for seeking and sharing knowledge.

Online collaborative encyclopedias still have several weaknesses, however, in terms of seeking and reusing knowledge resources, and many researchers have identified the disadvantages of online encyclopedias such as Wikipedia (Wu & Wu, 2011). For example, there are several questions which users ask, such as "Which article is important and worth reading?" "What is the concept in related articles?" "How can I get comprehensive understanding of new knowledge by reading just a few articles?" Lacking a practical way to reduce these disadvantages and improve performance, the results would be a waste of time and information overload. As a result, a recommender mechanism of social knowledge navigation which can support people to reuse the huge volume of available knowledge urgently needs to be established.

In order to solve the above problems, this research proposes a novel social knowledge navigation system (SKNS) to support users in knowledge navigation of online free encyclopedias. This study proposes an innovative recommender mechanism of knowledge visualization and navigation which would contains related knowledge findings, judges the importance of knowledge, and offers a visualized knowledge navigation interface. It would construct a more comprehensive view of the knowledge map and highlight the important knowledge flows so people could find knowledge information efficiently. In the proposed system, a knowledge seeker could key in a word and the system would help to find the specific knowledge information required. If the query results were not in accordance with the users' requirements, then the system would generate sub-concepts as suggestions which would direct the user to search again.

The remaining parts of this article are organized as follows. Section 2 of this paper is a review of the related literature on the research topic. Section 3 details the social knowledge navigation recommendation system, which is combined with relevant knowledge network construction, knowledge importance analysis, knowledge concept ontology construction, and a knowledge visualization and navigation interface. An empirical experiment is described in Section 4. Section 5 provides comprehensive experiment results and evaluations. Section 6 concludes and suggests future research directions.

2. Related literature

2.1. Social media and social knowledge

In recent years, in the light of emerging trends and the growing popularity of social media, researchers have attempted to analyze the characteristics and enhance the practical uses of social media. Social media gives researchers an unprecedented research opportunity (Kwak, Lee, Park, & Moon, 2010). Social media are "a group of Internet-based applications built on the ideological and technological foundations of Web 2.0, that allow the creation and exchange of user-generated content" (Kaplan & Haenlein, 2010). The information created on the social media platforms can be divided into two categories; user-generated content and non-content information (for example the links between articles and user feedback ratings) (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008). Both of these are widely used to extract knowledge and mine hidden information to solve the problem, decision-making, or predict the future (Agichtein et al., 2008; Choudhury, Sundaram, John, & Seligmann, 2008; Stein & Maier, 1995). This kind of content and information is termed 'social knowledge'. For instance, Wikipedia is a form of output of social knowledge; it is a product generated from the wisdom of the crowd. Cheong and Lee (2009) apply artificial intelligence-based data mining approaches to analyze messages on Twitter to find activity trends. Choudhury et al. (2008) analyze the communication dynamics of blogs and use it to determine correlations with stock market movements. Stein and Maier (1995) discuss how to use online forums as a powerful tool for problem solving.

The present research considers that social knowledge is not only a good resource for problem solving or decision-making, but that it also helps people to seek and learn new knowledge. Therefore, this study aims to improve learning effectiveness by utilizing the power of social knowledge in online encyclopedias.

2.2. Knowledge relevance and analysis of importance

In today's network environment, a huge amount of user-generated content and non-content information is created on social media platforms (Agichtein et al., 2008). It is important to alleviate information overload on a social media platform and to improve the seeking performance. In order to find the useful knowledge of user generated content, being able to identify high quality of information is an important. Traditionally, information retrieval (IR) techniques were used to improve search performance. Recently, network structure mining techniques, such as link mining, have been successfully used with web hyperlink data in order to evaluate the importance of web pages (Wu & Wu's, 2011; Chiang, Chen, & Yang, 2008). Wu and Wu (2011) utilize internal link analysis to measure the importance of Wikipedia's articles. He, Pei, Kifer, Mitra, and Giles (2010) propose a context-aware citation recommendation system for finding high-quality citation articles. Social network analysis (SNA) pays attention to the relationships between interactive social entities. Graph theory, theoretical concepts and matrix operations act as the foundations of social network analysis (Hage & Harary, 1983). Social rating mechanisms (Sarma, Sarma, Gollapudi, & Panigrahy, 2010), allow users to evaluate the quality of nearly anything (e.g. blog posts, books, movies, hotels, etc.). Agichtein et al. (2008) explores methods of using community feedback (ratings) to automatically recognize essential and high-quality content. Social rating systems are increasing in popularity among commercial sites (e.g. Digg.com, Amazon.com and Netfix.com).

In our work, social rating theory is used as a theoretical foundation for social interaction (SI), and we use social rating mechanisms to evaluate which articles are worthy to read and learn in the knowledge network. The importance of social interaction (feedback rating) is measured by counting the views of the crowd and the experts.

One of the most common approaches used to find the similarity between two Web contents is a text mining approach (Berendt, Hotho, & Stumme, 2002). There are three famous similarity measures which are usually used in the text mining process: cosine similarity, Jaccard similarity and Pearson correlation (Strehl, Ghosh, & Mooney, 2000). The technique of quantifying similarity between different objects and text documents can be used in many different modules and Web applications (Erk, 2007; Haveliwala, Gionis, Klein, & Indyk, 2002; Kim & Choi, 1999). For instance, Kim and Choi (1999) present a comparison of collocation-based similarity approaches to determine extra search terms support query expansion; Yuan and Sun (2005) present a structured cosine similarity (SCS) method for document summarization; Singh, Krishna, and Saxena (2009) propose an augmented Jaccard similarity function for the preservation of privacy.

This research considers and integrates the methods of the above studies in order to construct a comprehensive method for analyzing knowledge relevance and importance, and aims to provide important and relevant knowledge articles for knowledge seekers.

2.3. Knowledge maps and knowledge flow

Knowledge visualization investigates ways of using visual illustrations to enhance knowledge sharing between two people or a group (Burkhard, 2004). In the light of Burkhard's (2004) research, knowledge visualization can effectively reduce problems of overload, misinterpretation, and misuse.

A knowledge map is one of the methods used in knowledge visualization, and can be defined as "a tool or technique that enables visualizing knowledge and relationship in a clear form via a way so that the relevant features of the knowledge can be clearly highlighted" (Vail, 1999). It can be applied to a number of different systems or mechanisms. For instance, in Ong et al.'s (2005) research, text mining and SOM categorization are used to construct a knowledge map for reducing the information overload on online news websites. Wang, Su, and Hsieh (2011) propose an implicit knowledge extraction mechanism to build a knowledge map in order to carry out more effective assessments. Other research summarizes the advantages and disadvantages of the different display formats of knowledge maps (Lin & Yu, 2009). The three major display formats of a knowledge map are (1) hierarchical displays (Ong et al., 2005), (2) network displays (Liu, Ke, Lee, & Lee, 2008), and (3) map displays (Wu & Wu, 2011). According to Lin and Yu (2009), the map display format is the most suitable display format. It can help users find the major concepts and the relationships between different objects easily, but can still contain more information than hierarchical displays.

A Knowledge flow fulfills people's knowledge requirements and the referencing of the order of articles about the tasks conducted (Liu & Lin, 2012), and it can be defined as "a process of knowledge passing between people or knowledge processing mechanism" (Zhuge, 2002). It can offer task-related knowledge to help meet people's knowledge demands quickly and effectively (Zhuge, 2002) and can be built up by focusing on knowledge needs (Kim, Hwang, & Suh, 2003; Xiangfeng & Jie, 2010). Other researchers find that a knowledge flow can be applied to offer recommendations to workers (Lai & Liu, 2009).

Therefore, in order to avoid the information overload problem inherent in seeking new knowledge on online encyclopedia platforms, this study uses knowledge visualization to reduce it, and chooses a map display format for the visualization of knowledge articles. It also uses knowledge flow to help knowledge seekers view necessary knowledge effectively. If sufficiently supported by the information displayed on the knowledge map and a knowledge flow, people can seek new knowledge effectively.

3. The system framework

A framework of the proposed social knowledge navigation mechanism used in online encyclopedias is designed and presented in this section. Fig. 1 depicts the main components and procedures of the proposed social knowledge navigation system (SKNS) architecture, which includes four main modules.

The first component is a relevant knowledge network construction module; the aim of this component is to collect relevant articles from the online encyclopedia platform. The second component is a knowledge importance analysis module; the purpose of this component is to evaluate which articles are worth reading. The third component is a knowledge concept ontology construction module; it is used to find and construct sub-concepts contained inside special knowledge. The last component is a knowledge visualization and navigation interface. With this component, the study can help knowledge seekers to understand the information and knowledge in the knowledge network effectively.

3.1. Relevant knowledge network construction module

The aim of this module is to collect relevant articles from online encyclopedia platforms. In order to discover and learn new knowledge from online encyclopedia platforms, this system gathers relevant articles and filters out unrelated articles. The processes of the module can be divided into two steps.

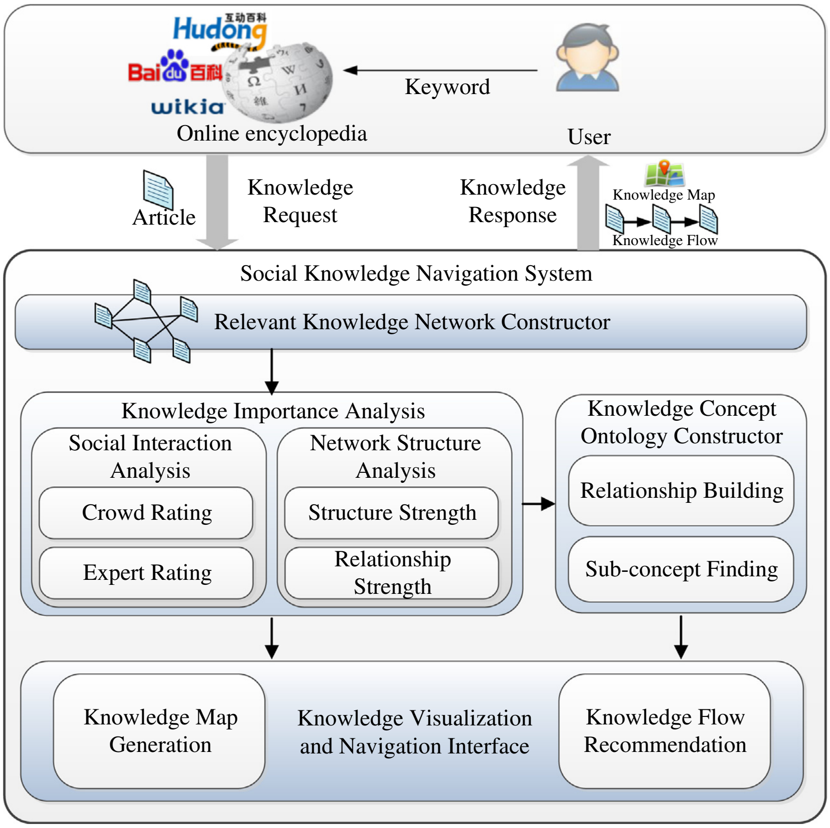

Fig. 1. System architecture.

3.1.1. Knowledge network construction

The first step is to construct the knowledge network, which includes articles relevant to the knowledge the users want to learn. This study uses the keyword as the input from a knowledge seeker, and regards it as the seeking target. The keyword search is one of the most common methods used to explore unknown knowledge, and the keyword search function is provided on almost all of the online encyclopedia platforms. Therefore, this study uses the keyword input as a knowledge seeker's seeking target, and uses it to find the start page of the knowledge article.

Almost all of the knowledge article pages contain internal hyperlinks, and a huge amount of reference and extension information can be obtained by accessing them. The internal hyperlinks may include sources about further readings related to this knowledge, so this research collects crucial internal hyperlinks (e.g. extended reading), which are contained in the knowledge articles, to find other relevant article pages. Those crucial internal hyperlinks represent the candidate articles which are closely related to the learning object of users. In this knowledge network, each article page is represented by a node and each internal hyperlink is represented as an edge. The knowledge network is constructed by expanding the hyperlinks and including the referenced articles recursively.

3.1.2. Knowledge relevance evaluation

The purpose of this step is to identify which articles included in the constructed knowledge network are highly relevant to the seeking target of a knowledge seeker. When a knowledge seeker wants to learn new knowledge on the encyclopedia platform, he/she is only interested in the articles matching his/her seeking target. To achieve this purpose, the technique of cosine similarity (Steinbach, Karypis, & Kumar, 2000) is applied to evaluate and remove less relevant nodes (articles) from the knowledge network. The formula of cosine similarity score is calculated as:

$$CSS_{SC} = \text{cosine}(\vec{D}_S, \vec{D}_C) = \frac{\vec{D}_S \cdot \vec{D}_C}{|\vec{D}_S||\vec{D}_C|}, \quad (1)$$

where $\vec{D}_S$ is the vector of the internal link term frequency (TF) of the start page (the keyword page on the encyclopedia platform), and $\vec{D}_C$ is the vector of the internal link TF of a candidate page, which is retrieved from one of the internal links. By calculating the similarity between start page and candidate page, we can obtain the articles which are related to the learning object of the knowledge seeker.

3.2. Knowledge importance analysis module

The purpose of the knowledge importance module is to find the articles which are worth reading and learn from the constructed in the knowledge network. This study combines the factors of social interaction and social structure to analyze the importance of an article. The importance of social interaction is measured by counting the views of crowds as well as experts. The importance of social structure is evaluated according to the node's importance in the network structure.

3.2.1. Social interaction analysis

The importance of social interaction (SI) includes two parts, crowd rating (CR) and expert rating (ER), which stand for the aggregated feedback score of an article evaluated by crowds and experts respectively. Each article in this knowledge network has its own CR and ER scores. CR_i is the crowd rating of article i aggregated from the feedback provided on the online encyclopedia platform. The formula of CR_i is written as:

$$CR_i = \frac{1}{n}\sum_{k=1}^{n}(cr_i^k/d), \tag{2}$$

where k is the evaluation metrics in the encyclopedia platform. d means the crowd feedback as evaluated by a d-point Likert scale. For instance, if the platform's user feedback uses a five-point Likert scale as the first evaluation metric, then d is equal to five. cr_i^k is the average score of metric k's user feedback for article i. n is the total number of evaluation metrics. A higher CR_i indicates the greater value of knowledge article i which the general users perceive.

The encyclopedia platform can also recruit knowledgeable people as experts to rate the articles, so this study provides ER as the expert rating for a typical article i. The formula of ER_i is written as:

$$ER_i = \frac{1}{n}\sum_{k=1}^{n}(er_i^k/d), \tag{3}$$

where er_i^k is the article i's average score of experts' feedback on metric k.

Combining the crowd and expert ratings measured above, the knowledge importance for article i, evaluated from the social interactions SI_i, can be formulated as:

$$SI_i = w_r \cdot CR_i + (1 - w_r) \cdot ER_i, \tag{4}$$

where w_r is the weight of CR_i, and $(1 - w_r)$ is the weight of ER_i. This weight can be adjusted by the encyclopedia platforms.

3.2.2. Network structure analysis

We can further measure the importance of an article from the aspect of knowledge network structure. This research considers the in-links, out-links, bidirectional links, and degree centrality of each article to measure the connection structure strength and the relationship tie strength between two different articles. We improve the structure strength analysis method used in Wu and Wu's (2011) study. The formula of structure strength (SS) for article i is represented as:

$$SS_i = \left(\frac{|\Theta_B(n_i)|}{\sum_{n_i \in KN}|\Theta_B(n_i)|} + w_I \cdot \frac{|\Theta_I(n_i)|}{\sum_{n_i \in KN}|\Theta_I(n_i)|} + w_O \cdot \frac{|\Theta_O(n_i)|}{\sum_{n_i \in KN}|\Theta_O(n_i)|}\right) \cdot \Phi_D(n_i), \tag{5}$$

where KN is the set of the articles included in the knowledge network and $w_I + w_O = 1$. $\Theta_B(n_i)$ is the set of bidirectional links between article i and the other articles in this knowledge network. $\Theta_I(n_i)$ denotes in-links of article i. $\Theta_O(n_i)$ *denotes* out-links of article i. w_I and w_O is the weight for in-links and out-links respectively. Also, because all of the links represent relevant articles inside this knowledge network, the study considers the degree centrality of each node. Let Φ_D (n_i) denote the CDF (cumulative distribution function) of the degree of article i's internal links. Note that SS for article i will be normalized before further score aggregation.

Next, if articles i and j have a higher common in-links and out-links, the tie strength of relationship between them is stronger. A common in-link means two articles have the same source or reference information; conversely, a common out-link means they are used in the same application or method. The tie strength of the relationship between articles i and j is measured as:

$$RS_{ij} = \frac{1}{2}\left(\frac{|\Theta_I(n_i) \cap \Theta_I(n_j)|}{|\Theta_I(n_i) \cup \Theta_I(n_j)|} + \frac{|\Theta_O(n_i) \cap \Theta_O(n_j)|}{|\Theta_O(n_i) \cup \Theta_O(n_j)|}\right). \tag{6}$$

The self-loop link and the addition links will be detected and removed. Finally, the importance score for article i is formulated as:

$$KIS_i = \frac{1}{2}(SI_i + SS_i). \tag{7}$$

3.3. Knowledge concept ontology construction module

The knowledge concept ontology construction module is developed to extract the sub-concepts and their relationship from the constructed knowledge network according to the user's seeking target. Ontology is a structured knowledge base which describes the concepts and their relations in a domain (Cui, Lu, Li, & Chen, 2009; Liu & Lin, 2012). For the purpose of learning new knowledge effectively, this study builds seeking target-driven knowledge concept ontology. $O =< NC, R >$ denotes a ontology, where NC is a set of knowledge concepts and R is the parent-child relation between two different knowledge concepts. Because all of the article nodes contain category labels in the online encyclopedia environment, the concepts can be extracted by NLP. For instance, the category labeled "Methods in sociology" will extract two concepts, "Method" and "sociology." Also, if the title of the article is also a category, this category label should be considered as a concept, too. The set of concepts a typical article i involves can be expressed as:

$$\Omega_i = \{c_i^1, c_i^2, \ldots, c_i^{n_i}\}. \tag{8}$$

The set of aggregated concepts extracted from the knowledge network can be written as:

$$NC = \bigcup_{n_i \in KN} \Omega_i \tag{9}$$

To construct this parent-child relation from the extracted concepts, this study uses the platform's category and sub-category hierarchy relation of each article (Cui et al., 2009). The schema of knowledge concept ontology is illustrated in Fig. 2, in which a circle represents a unique concept, and the link represents their parent-child relationship. For example, c_2 has a parent relationship with c_1; on the contrary, c_1 has a child relationship with c_2.

The construction process of the knowledge concept is illustrated by the following example. Assume a user wants to seek and learn the new knowledge "social networking service" and search for it on an online encyclopedia platform. This keyword "social networking service" is his/her seeking target. The system first identifies several relevant and important articles, such as the "social network analysis" and "social software" articles. Then, the system extracts the category labels in those articles, such as "category: social network" in the article "social network analysis" and "category: social networking service" in the article "social software." Finally, according to the category tree established by the encyclopedia, the system finds that a parent-child relationship exists between the "category: social network" and "category: social networking service." The relationship is built between c_2 (category: social network) and c_1 (category: social networking service) in Fig. 2.

3.3.1. Sub-concept finding

The aim of sub-concept finding is to construct the knowledge map satisfying the seeking target. The main steps needed for the sub-concept finding are as follows. The first step is to map the concepts of each article to the knowledge concept ontology. The second step is to calculate each article's concept score by the concepts it has.

The knowledge concept ontology is the forest-like structure in the encyclopedia platform. Each root of a tree in the forest represents a main sub-concept. The score of a sub-concept equals the number of nodes included in the path from article node concepts to the sub-concept nodes. The final step is to compare each article node's score. The concept with the largest score represents this article node's sub-concept. If there are two concepts with an identical score, those two concepts also can be regarded as this article's sub-concepts. For instance, there are two main sub-concepts "social network" (node c_3) and "social networking service" (node c_4) in Fig. 2. Assume "category: social networking" and "category: social networking service" are the only category labels contained in the "social network analysis" article (node c_2) and in the "social software" article (node c_1). The score of the main sub-concepts (c_3,c_4) with respect to c_1 and c_2 is (3,3) and (4,0) respectively. Therefore, the sub-concept of the "social network analysis" article is node c_3 ("social network"); and the sub-concept of the "social software" article is node c_3 ("social network") and node c_4 ("social networking service").

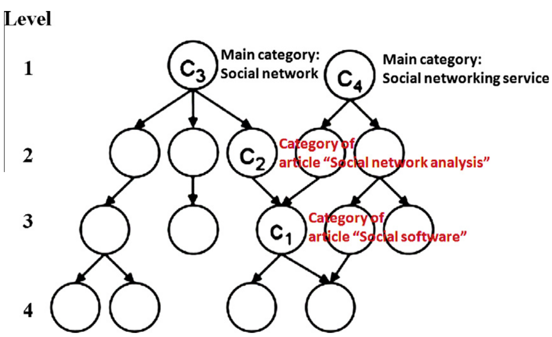

Fig. 2. Schema of knowledge concept ontology.

3.4. Knowledge visualization and navigation interface

In order to let knowledge seekers understand the information in the knowledge network effectively, this study builds a visualized social knowledge map and navigation interface with NodeXL, which provides a Microsoft Excel 2007 and 2010 template. It supports network analysis, popular SNA statistics, clustering, and the visualization features of the spreadsheet (Ahn, Taieb-Maimon, Sopan, Plaisant, & Shneiderman, 2011; Smith et al., 2009). The visualized social knowledge map can help knowledge seekers to view the article's relationship (e.g. article's sub-concept), and the navigation interface can suggest reading sequences for knowledge seekers. The following paragraphs describe how to build the knowledge map and reading sequence.

3.4.1. Knowledge map generation

This study uses a knowledge map to help users to understand the overall relationships and sub-concepts of target knowledge. The knowledge map is generated in the following steps. According to the knowledge concept ontology, we map the sub-concepts of each article in the knowledge network. Then, these sub-concepts corresponding to the article on the knowledge network are expressed by different symbols and colors. The knowledge map with the sub-concepts is shown by NodeXL. In this step, a knowledge seeker can see the sub-concept information related to the seeking target, such as "What are the main sub-concepts of this special knowledge?" With the support of this knowledge map, knowledge seekers can understand new knowledge effectively.

3.4.2. Knowledge flow recommendation

In order to help knowledge seekers to learn the new knowledge effectively, a knowledge flow recommendation is generated for the seeking target. The knowledge flow recommendation is a reading path highlighted by the knowledge map. An article on online encyclopedia platforms is usually very long. Therefore, in order to reduce the number of articles to be read, this proposed system only recommends a few articles. Then, the system uses the evaluated knowledge importance, the article node's link, and the knowledge concept ontology to build the sequence in the following steps.

First, we use the keyword's article node as the start node and locate the position of the keyword node's concepts in the knowledge concept ontology. In the second step, we use in-links of the start node to find the cause of the learning object, and out-links of the start node to find the effect of the learning object. So we can determine the first cause node and effect node, the following two conditions must be met: the candidate nodes should be among the top five in terms of knowledge importance. The concept of the candidate nodes should have a parent relationship with the start node's concept. The in-linked candidate node with the maximum number of related concepts is selected as the first cause node. The out-linked candidate node with the maximum number of related concepts is selected as the first effect node. By repeating the above rules, we can further define the cause of the cause node and the effect of the effect node until a few article nodes are produced to construct the reading sequence.

4. Experiments

Our proposed framework is implemented and validated on a famous online encyclopedia platform, Wikipedia, which is a knowledge-sharing website. Wikipedia provides a huge amount of knowledge article resources supplied by countless volunteers around the world. According to Alexa's survey in 2012, Wikipedia is ranked sixth in the world. Alexa estimates that 20% of this site's visitors are from the USA, and Wikipedia's audiences are often users who browse at work and school (Alexa, 2012b). We use Java and NodeXL to develop a Wikipedia application called SKNS (social knowledge navigation system). The detailed data analysis and experimental procedures are described as follows.

4.1. Data description

This research selects three keywords: "social networks," "electronic commerce," and "marketing" as the seeking targets. These three keywords represent relative but different academic fields. This research invited people who have used the English version of Wikipedia to participate. All of them have a university degree or above, and they are interested in or familiar with these seeking targets. In May 2012, 94 users (35% male, 65% female) participated in our experiment on the seeking target "social network", 92 users (42% male, 58% female) on the seeking target "electronic commerce," and 90 users (58% male, 42% female) on the seeking target "marketing." For simplicity, in the following section, we use the seeking target "social network" as an example to describe how to implement this system on Wikipedia. The participants are asked to join quiz and self-evaluation with two different systems (basic Wikipedia vs. SKNS) for assessing learning performance and use the five-point Likert scale to verify the recommendation of knowledge flow on learning performance.

Human experts are invited to evaluate the relevance between the recommendation articles and the recommendation results of the knowledge map. In this study, a group of experts, including 1 professor, 3 assistant professors, 3 PhD students, and 30 graduated students major in MIS or Marketing is formed to judge whether an recommender article is related to the seeking target or not and evaluate the quality of knowledge map.

4.2. Data analysis

4.2.1. Relevant knowledge network construction

To identify the relevant articles, the internal hyperlinks of the start page (e.g. "social network" article in Wikipedia) are used to expand and find the related pages recursively for three layers. Then, we analyze the similarity between the learning object's article and other articles in the established knowledge network. In this research, we use 0.3 as an appropriate threshold value for cosine similarity. The initial knowledge network is shown in Fig. 3. The detailed information on the initial knowledge network is shown in Table 1.

4.2.2. Knowledge importance computing

We measure the importance of an article based on social interaction analysis and network structure analysis. For the social interaction analysis, we collect each article's crowd rating (CR) from user feedbacks on the article page. A feedback in Wikipedia contains four metrics: "Trustworthy," "Objective," "Complete," and "Well-written." Then we obtain the expert rating (ER) from each article's talk page. Almost all of the articles have an expert rating on their quality and importance in Wikipedia. Those articles which do not have expert ratings are set at the lowest levels. "Wisdom of the crowd" is an important belief in Wikipedia, so the opinions of crowd could be relatively important to this study. We compare the precision rates with respect to different weighting sets. From Fig. 4, we find that the weight setting $(CR,ER) = (0.6,0.4)$ generates the best performance.

In the network structure analysis, both factors of structure strength (SS) and relationship strength (RS) are calculated by the in-links, out-links, and bidirectional links of each article. According to ()()(5)–(7), the system can obtain SS for each article and RS for each link between two articles. Fig. 5 compares the precision rates with respect to different weighting sets. As we can observe, the weight set $(SS,RS) = (0.4,0.6)$ generates the best performance. Finally, this research can obtain KIS from (8). The value of KIS is interval $[0,1]$, and the threshold value of KIS is 0.5.

The characteristics of the knowledge network for "social network" in Wikipedia are shown in Fig. 6. The detailed information on the final knowledge network for "social network" is shown in Table 2.

4.2.3. Knowledge concept ontology construction in Wikipedia

After knowledge importance analysis, we obtain a knowledge network in which all of the nodes are important and relevant to the seeking target. To provide the knowledge map and recommendation reading sequence, we construct the knowledge concept ontology related to the seeking target. The final social network knowledge concept ontology is shown in Fig. 7. Note that to present a more concise and clear knowledge map, some category labels are reorganized. For example,

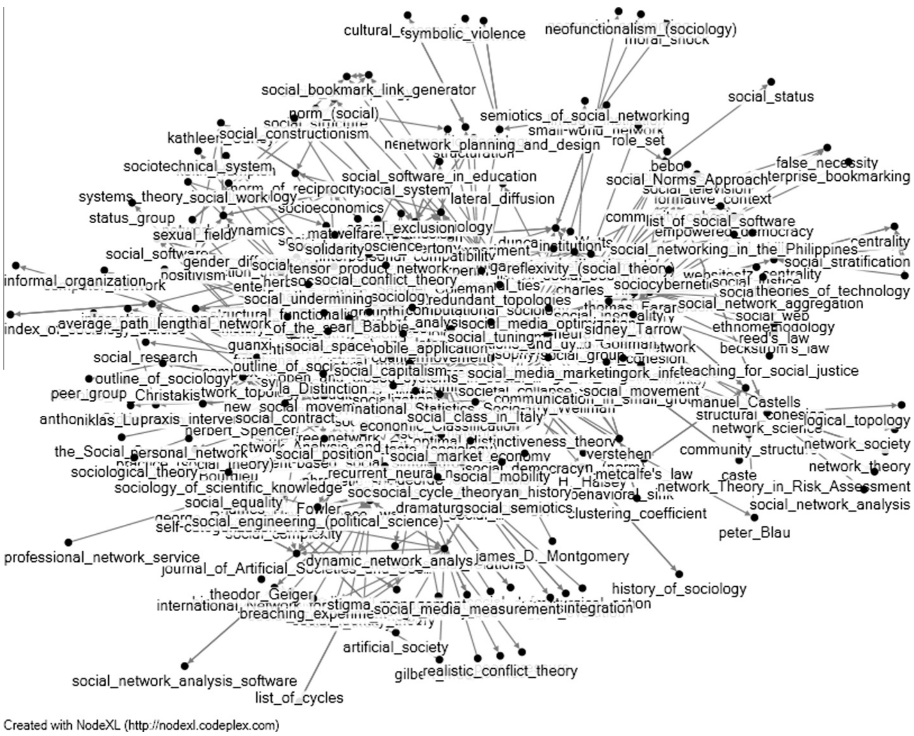

Fig. 3. Initial knowledge network about the seeking target "social network".

Table 1
Information on the initial knowledge network "social network".

Detailed information on the initial knowledge network "social network"	
Vertices	260
Edges	701
Maximum geodesic distance (Diameter)	7
Average geodesic distance	3.618077
Graph density	0.01040986

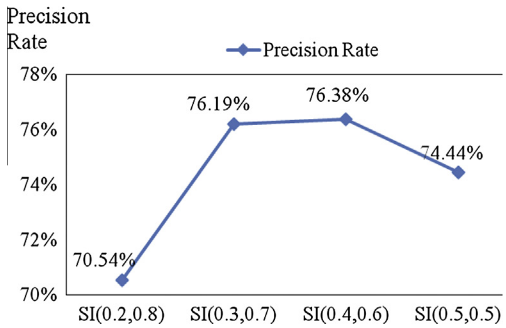

Fig. 4. Precision of different SI weighting sets.

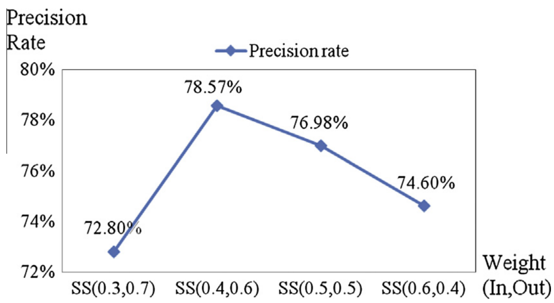

Fig. 5. Precision of different SS weighting sets.

one category label in article "social network analysis" is named "Methods in sociology"; this study uses the category label by simplifying it as "Methods" and "Sociology". After the label simplification process, some of category levels can be consolidated.

4.2.4. Knowledge visualization and navigation strategies

The final step is to recommend articles of high importance and good quality to knowledge seekers. In order to reduce the information overflow problem, this study provides a knowledge map and reading sequence. A knowledge map can help a knowledge seeker to view the knowledge overview of the seeking target. A knowledge flow offers a reading path so they can learn the seeking target conveniently and quickly. We term it "reading sequence" in the interface. The navigation interface is shown in Figs. 8, 9 and 10 are respectively the recommended knowledge map and knowledge flow for seeking the target "social network."

To show whether our approach has better performance to help user to learn new knowledge than today's Wikipedia platform, the participants are asked to join the three-stage experiments with two different systems (basic Wikipedia vs. SKNS). These three stages are: seeking stage, exam stage, and evaluation stages when using different user interfaces (basic Wikipedia vs. SKNS) in the experiment. In the seeking stage, the participants are allowed to spend a maximum 30 min searching for and learning the knowledge about the learning objects (e.g. social network, electronic commerce, and marketing). After the seeking stage, this research automatically redirects to the exam stage and request the participants to answer three multiple choice questions, which are randomly selected from the question database. Finally, this study requests the participants to evaluate the learning effect by previous experience (evaluation stage). Here this work uses a questionnaire of five-point Likert scale. This study repeats those three stages on two different systems. This study starts the experiment in the basic Wikipedia interface first. The day after the completion of the experiment in the basic Wikipedia interface, this study conducts the experiment in the SKNS interface. This design is to reduce the learning experience on the first interface to help

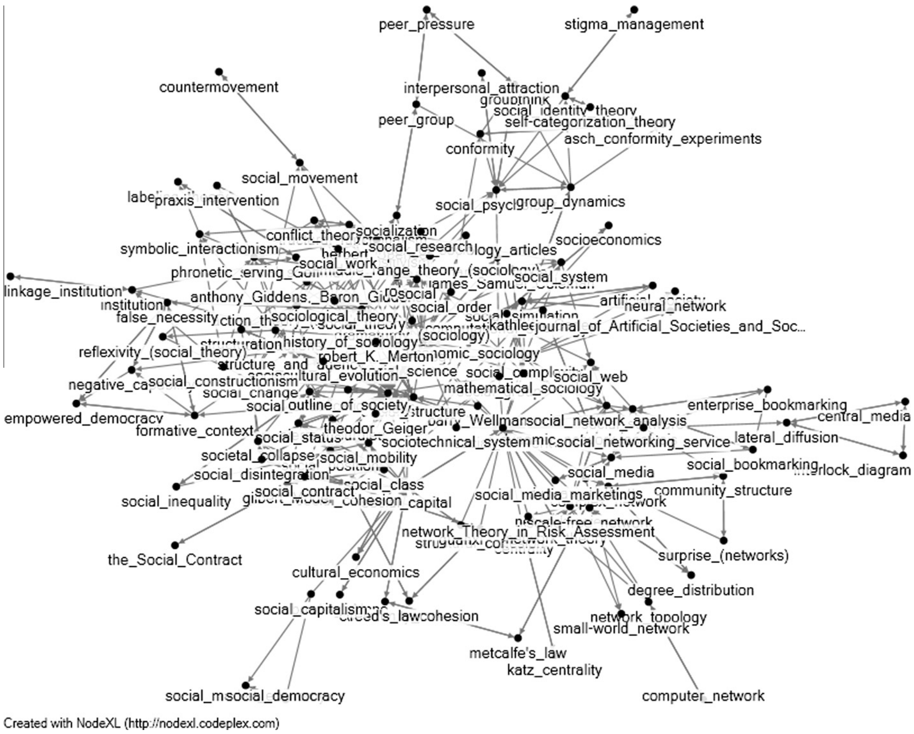

Fig. 6. The final knowledge network for "social network".

Table 2
Detailed information on final knowledge network.

Detailed information on final knowledge network	
Vertices	123
Edges	416
Maximum geodesic distance (Diameter)	7
Average geodesic distance	3.164651
Graph density	0.0277222

the learning experience on the first interface. By the examination and self-assessment with these two different interfaces, the learning effectiveness can be evaluated and compared.

4.3. Experiment processes

In order to verify the performance of our approach, this study uses and compares three different approaches to generate a knowledge map and knowledge flow for each seeking target. Four types of recommendation approach were used:

1. Random approach: the easiest approach to determining the recommendation article, knowledge map, and knowledge flow for a seeking target. It simply follows the article's internal hyperlinks and randomly selects some of those articles as the recommended reading articles.
2. Important degree approach: this approach decides the knowledge map, knowledge flow, and recommended reading articles based on the article's feedback rating. It selects the articles with high feedback rating as the recommended reading articles.
3. Similarity-based approach: this mechanism uses the similarity of other articles and the seeking targets to determine the recommended articles. Articles which are less similar will be filtered out in this mechanism.
4. SKNS approach: the approach we propose in this study. This approach uses similarity analysis and knowledge importance analysis to determine the recommended articles.

As we know that precision and recall are widely used as relevancy measures for information retrieval (Salton & Buckley, 1988). Lin and Hsueh (2006) and Su and Wang (2010) adopted precision to evaluate the quality of recommended knowledge

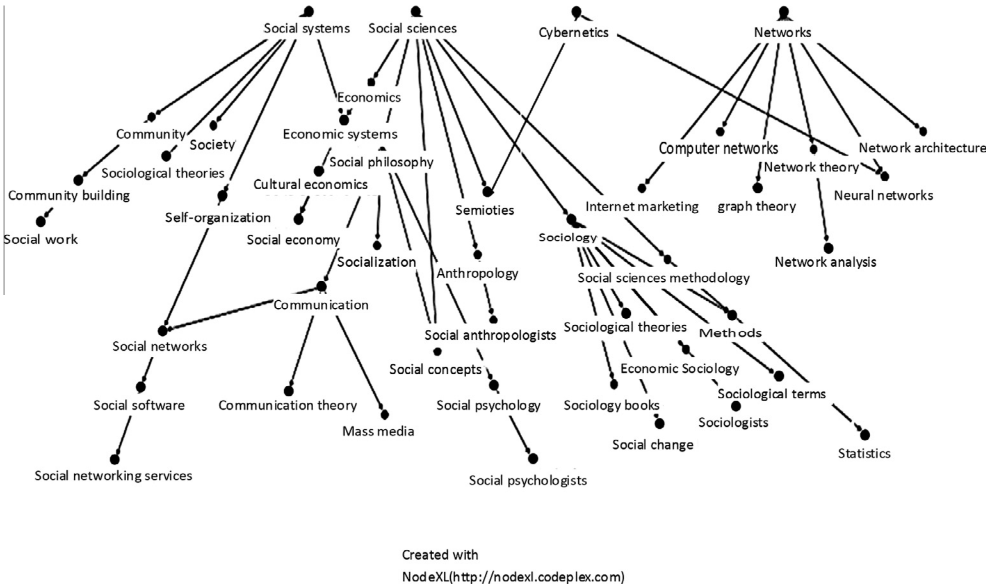

Fig. 7. Knowledge concept ontology generated for the keyword "social network".

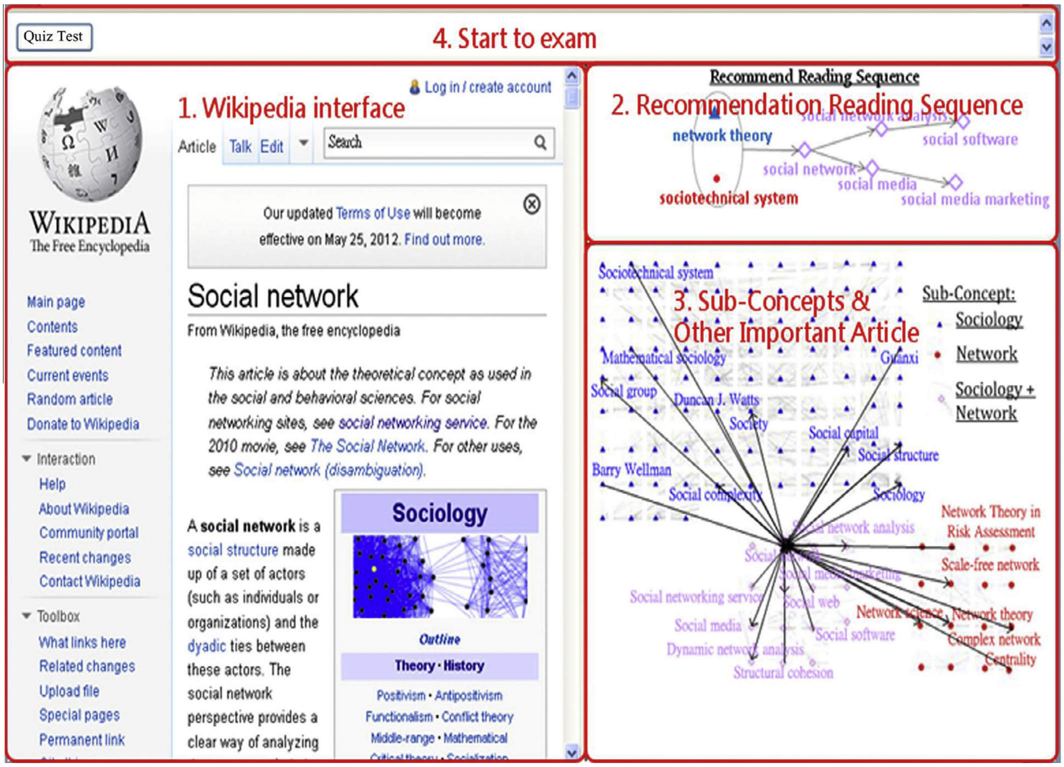

Fig. 8. SKNS interface.

map. Precision measures the percentage of relevant documents in relation to the number of documents retrieved. In this study, we view a seeking target as a category designation, and adopt precision and measures to evaluate how accurate the knowledge map creation method complies with experts in assigning documents to specific categories and use the five-point Likert scale for users to verify the recommendation of knowledge flow on learning performance. Φ_I denotes the set of documents allocated to a category by an expert, and Φ_R denotes the set of documents assigned to a category by the knowledge map creation method. The formula of recommendation's precision rate (PR) is written as:

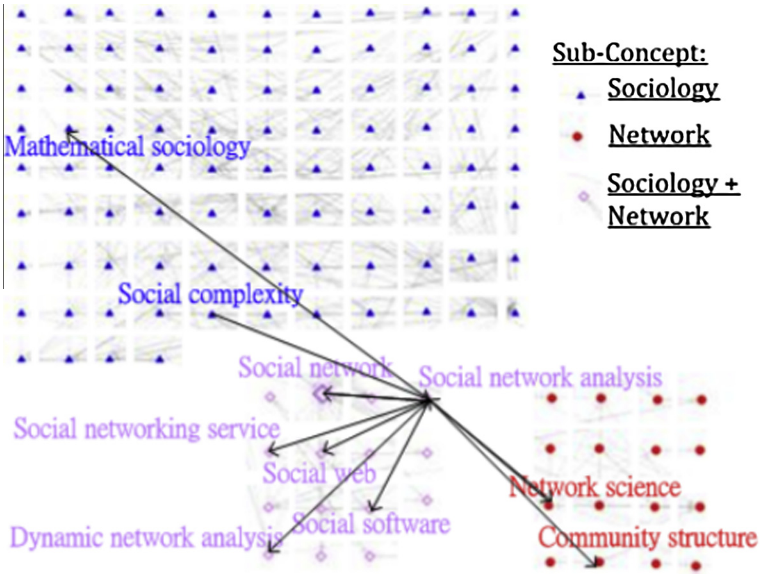

Fig. 9. Knowledge map.

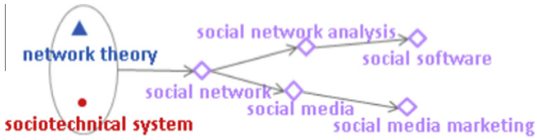

Fig. 10. Recommended knowledge flow.

$$PR = \frac{|\Phi_R \cap \Phi_I|}{|\Phi_R|}. \tag{10}$$

The whole knowledge map's precision rate is calculated by averaging the values of precision and recall respectively from all categories of the modified knowledge map.

5. Results and evaluations

In this section, the experimental results and inspect insights which obtained from the observation and analysis are presented. First, the recommendations' precision rate is used to verify whether the system effectively provides correct articles for knowledge seekers. Then, we compare the knowledge learning performance between the basic Wikipedia interface and the SKNS interface. A quiz and self-evaluation questionnaire is used to verify the navigating performance of knowledge seekers (Shee & Wang, 2008).

5.1. Article recommendation accuracy

In this paper, our recommendation approach is compared with three other benchmark approaches via evaluating the precision of article recommendation. All of the recommendation approaches collect three layer's articles from the seeking target, but the method used to filter out irrelevant articles is different. The results of different recommendation approaches are shown in Fig. 11. As we can see, the proposed SKNS approach having the highest precision rate of all approaches with respect to all three seeking targets.

5.2. Knowledge map quality

We compare the quality of the knowledge map generated by the proposed SKNS approach and the three other approaches. To examine the quality of knowledge maps, we invited experts to evaluate them with a score ranging from zero to five. Fig. 12 show the results. Tables 3–5 outline the results of two-paired sample *t*-test. The results show that at the 95%

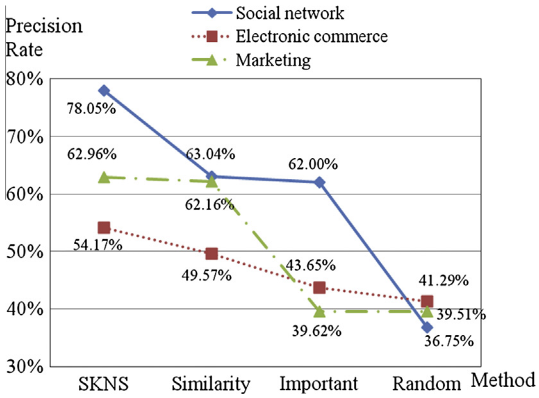

Fig. 11. Precision of recommendation approach.

significant level the knowledge map generated by the SKNS approach significantly outperforms those of the other three approaches.

5.3. Knowledge flow quality

We also compare the quality of the knowledge flow generated by the proposed SKNS approach and the other three approaches (Fig. 13). The results of two-paired sample t-tests are shown in Tables 6–8. The results show that at the 95% significant level the knowledge map generated by the SKNS approach significantly outperforms those of the other three approaches.

5.4. SKNS learning performance

Pretest-posttest, which measures the change resulting from experimental treatments, is widely used in behavioral research. In this study, we conduct an experiment based on pretest-posttest (Linn & Slindle, 1977) to evaluate and compare the knowledge-seeking performance of proposed SKNS mechanism with that of Wikipedia. Quiz is one of the most common methods to evaluate people's learning performance. This research invited the participants to answer three questions after a

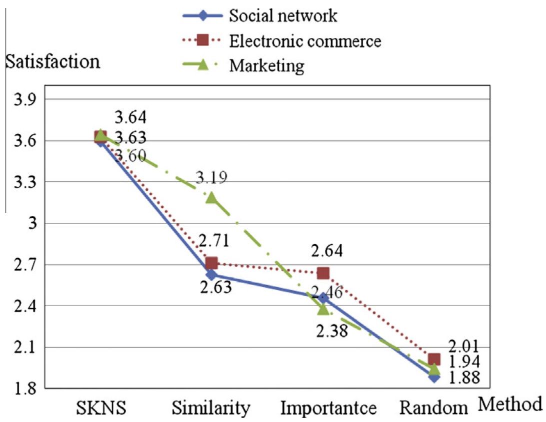

Fig. 12. Average satisfaction of knowledge map.

Table 3
Statistical verification of quality of knowledge map with different approach – social network.

Paired group		Mean	Std deviation	Std error mean	T value	Sig.
Random	V.S. SKNS	1.7692	1.0860	.1138	15.540	0.000
Similarity		1.0000	.9189	.0963	10.381	0.000
Important		1.1758	.9955	.1044	11.268	0.000

Table 4
Statistical verification of quality of knowledge map with different approach – electronic commerce.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
Random	V.S. SKNS	1.6170	.8564	.0883	18.306	0.000
Similarity		.9149	.7133	.0734	12.435	0.000
Important		.9894	.9332	.0963	10.279	0.000

Table 5
Statistical verification of quality of knowledge map with different approach – marketing.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
Random	V.S. SKNS	1.7000	.9651	.1017	16.710	0.000
Similarity		.4556	.8632	.0910	5.006	0.000
Important		1.2667	.9807	.1034	12.253	0.000

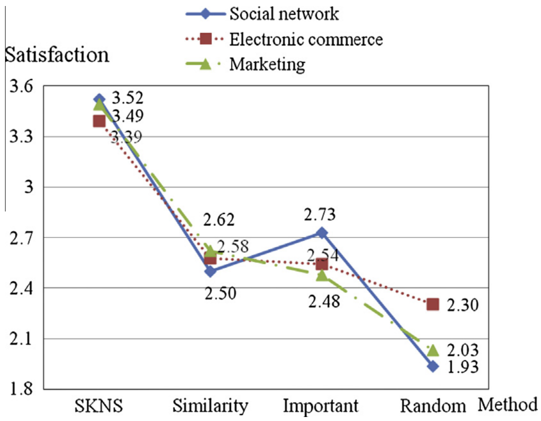

Fig. 13. Average satisfaction of knowledge flow.

Table 6
Statistical verification of quality of knowledge flow with different approach – social network.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
Random	V.S. SKNS	1.5870	1.0393	.1084	14.645	0.000
Similarity		.8478	.9256	.0965	10.588	0.000
Important		.7935	.8456	.0882	9.000	0.000

Table 7
Statistical verification of quality of knowledge flow with different approach – electronic commerce.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
Random	V.S. SKNS	1.0870	1.0234	.1067	10.188	0.000
Similarity		.8152	.9713	.1013	8.050	0.000
Important		.8478	.8508	.0887	9.558	0.000

period of learning time with respect to two different interfaces (pretest/basic Wikipedia vs. posttest/SKNS). The questions are randomly selected from the question database as shown in Appendix A. To assess the validity of the questionnaire, it was examined by 37 MIS or Marketing researchers. The internal consistency and reliability were tested by means of the Cronbach's alpha coefficient, and the result was 0.94, indicating the questionnaire was acceptable with good internal consistency and reliability. In the experiment, we invited the participations to answer three questions after a period of learning time with interfaces of basic Wikipedia (pretest). After the day when the experiment in the basic Wikipedia interface is completed, we conducted the experiment in the SKNS interface (posttest). Fig. 14 shows the comparison of the quiz results of two different interfaces. The average number of correct answers of quiz is 2.6129 and 2.0323 in the posttest and the pretest on seeking target "social network". Statistical verification of quiz with pretest and posttest is evaluated by two-paired sample *t*-test

Table 8
Statistical verification of quality of knowledge flow with different approach – marketing.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
Random	V.S. SKNS	1.4565	1.0630	.1108	13.143	0.000
Similarity		.8696	1.01990	.1062	8.187	0.000
Important		1.0109	.9550	.0996	10.153	0.000

and shown in Table 9. At the 95% significant level, the test results show that the seeking knowledge performance with SKNS interface (posttest) significantly outperforms than basic Wikipedia interface (pretest) in all three seeking targets.

5.5. Learning interface satisfaction performance

Because the learning effective is a kind of subjective feeling, it cannot be measured by simply using quiz. So, this study also uses questionnaire to verify every participant's feeling of learning effectiveness (Self-Evaluation), and summarize it as the learning satisfaction (Shee & Wang, 2008). In this study, ease of use, user-friendliness, ease of understanding, up-to-date content, sufficient content, and useful content are considered to evaluate the learner-interface (basic Wikipedia vs. SKNS). The detailed content of questionnaire is shown in Table 10. Then, according to this questionnaire, this study can find the learning satisfaction with these two different interfaces as shown in Fig. 15. Besides, at the 95% significant level, the test results show that the overall learning satisfaction in the SKNS interface significantly outperforms than basic Wikipedia interface (see Table 11).

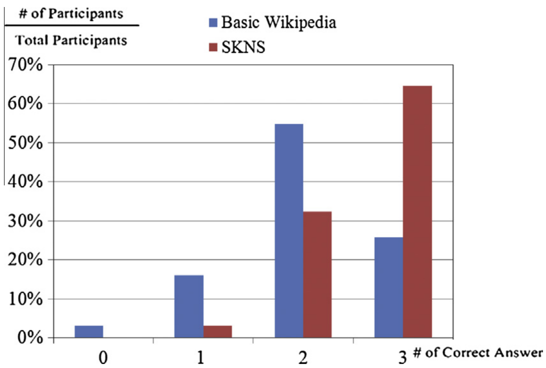

Fig. 14. Distribution of quiz's result on two pretest and posttest.

Table 9
Statistical verification of quiz with pretest and posttest.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
SKNS (posttest) (Social Network)	V.S. basic Wikipedia (pretest)	.5807	1.0796	.1371	4.235	0.000
SKNS (posttest) (Electronic Commerce)		.8681	.8591	.0901	9.640	0.000
SKNS (posttest) (Marketing)		.8901	.8360	.0876	10.157	0.000

Table 10
Content of questionnaire.

Content of questionnaire
1. In the learning process, I think the functionality provided by the system is easy to use
2. In the learning process, I think the information provided by the system is useful
3. In the learning process, I think this system can easily help me to find the desired information
4. In the learning process, I feel this system provides rich information
5. In the learning process, I think that the information provided by this system is not outdated
6. On the whole, I think this system is able to help me to learn new knowledge in the learning process

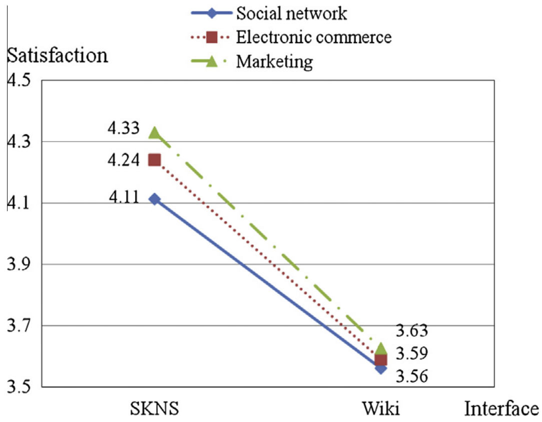

Fig. 15. Average satisfaction with different interfaces.

Table 11
Statistical verification of seeking knowledge satisfaction with two different interfaces.

Paired group		Mean	Std deviation	Std error mean	*T* value	Sig.
SKNS (Social Network)	V.S. Basic Wikipedia	.5430	1.0144	.1288	4.215	0.000
SKNS (Electronic Commerce)		.6520	.5745	.0602	10.827	0.000
SKNS (Marketing)		.6960	.5809	.0609	11.430	0.000

6. Discussions and conclusion

In online encyclopedia platforms which have a large amount of data, it is essential to find those articles which are worth reading. For example, Wikipedia, a free encyclopedia with more than twenty two million articles contributed by countless volunteers around the world. A vital issue is how to alleviate information overload and thereby facilitate effective information search activities. In order to find the useful knowledge of user generated content, being able to identify high quality of information is an important.

To obtain information need from a collection of information resources, information retrieval (IR) techniques (Cheong & Lee, 2009; He et al., 2010) were used to improve search performance. However, it cannot answer the question, such as "Which article is important and worth reading?". To solve the above issue, link analysis or social rating mechanisms (Wu & Wu's, 2011; Agichtein et al., 2008) are used to retrieve essential and high-quality content. Nonetheless, "What is the concept in the related articles?" is another issue. Knowledge visualization can reduce problems of overload, misinterpretation, and misuse. A knowledge map is one of the methods used in knowledge visualization to represent the relationships between relevant articles. A knowledge flow can provide relevant knowledge to people that helps them seeking their knowledge needs quickly and effectively.

This paper proposes a social knowledge navigation recommendation system for online encyclopedias. The proposed mechanism utilizes the techniques of relevant article identification, social interaction and network structure analysis, and knowledge concept ontology building, to recommend a knowledge map and knowledge flow (article reading sequence) for a requested learning object. A series of experiments conducted on Wikipedia verify that the proposed visualized interface of knowledge map and knowledge flow can effectively assist the knowledge seeker to acquire the relevant knowledge.

6.1. Research contributions

This study makes several significant contributions as follows. First, from the methodological perspective, it provides a novel and useful approach utilizing the techniques of text mining analysis, social network analysis, and knowledge ontology building to visually recommend articles on an online encyclopedia platform. Moreover, this proposed framework could be applied to develop recommendation applications in other text-based knowledge sharing platforms. Second, from the practical perspective, according to the experimental results, this study provides a visualized interface that can effectively improve the learning activities of knowledge seekers in online encyclopedias. The proposed visualized knowledge map and knowledge flow can reduce the information overflow problem in online encyclopedias. Because the system investigates and organizes the whole structure of an article, seeking and learning new knowledge in online free encyclopedias is easier and more effective. In addition, according to the experimental results, we found the proposed mechanism can increase not only the participants' leaning performance but also the overall learning satisfaction, besides system precision commonly used in information retrieval.

6.2. Limitations and future work

With regard to the study's limitations, knowledge seekers have to express explicitly what they want to learn without ambiguity and the keywords have to exist in the encyclopedia platform. In this context, semantic analysis is an important procedure that can be used to further improve the system and solve the ambiguity problem. Second, although our system can reduce the number of articles that needs to be read, the length of each article is still too long in the online encyclopedia platform and there is still an information overload problem for some knowledge seekers. Therefore, summarizing the articles would be another way to improve article quality and reduce the information overflow problem. Third, our mechanism only considers the value of articles from the reader side and does not consider the editor side. To evaluate the value of articles properly, the views of both article readers and editors should be taken into account in future work. Lastly, the proposed mechanism does not provide personalized recommendations, even though the requirements of different knowledge seekers must vary. If future research included the inputs of personal preference and experience, the recommendation results would be closer to the knowledge needs of individuals.

Acknowledgement

This research was supported by the National Science Council of Taiwan (Republic of China) under the Grant NSC 101-2410-H-009-009-MY3.

Appendix A

The content of multiple-choice questions – social network.

Content of multiple choice questions			
Question 1	Please select a "methods" or "technique" which is commonly used in the social media marketing		
	1. Direct marketing	2. Electronic word of mouth	3.Email marketing
Question 2	Please select a "methods" or "theory" can be used to analyze the social network.		
	1. Link analysis	2. Discourse analysis	3. Task analysis
Question 3	Please select an online media which is provided social network services.		
	1. UrMap	2. Google search	3. Epinions.com
Question 4	Please select a famous experiment which is related with social network		
	1. Schrödinger's cat	2. Michelson–Morley experiment	3. Small world experiment
Question 5	Please select a movie which is not related social network		
	1. The social network	2. Six degrees of separation	3. Roman holiday

The content of multiple choice questions – electronic commerce.

Content of multiple choice questions			
Question 1	Kalakota and Whinston (1997) believes that e-commerce infrastructure; e-mail belongs to that kind of infrastructure?		
	1. General business services infrastructure	2. Message and send the information infrastructure	3. Multimedia content and network publishing infrastructure
Question 2	Which type of E-commerce transactions are from consumer to consumer transactions?		
	1. B2B	2. C2C	3. P2P
Question 3	Which of the following belong to e-Tailer?		
	1. Yahoo!	2. AOL	3. Amazon
Question 4	Which of the following belong to content provider?		
	1. CNN.com	2. War-Mart	3. Google
Question 5	Please select a model which is not use in E-commerce.		
	1. C2B	2. B2B2C	3. B2P

The content of multiple choice questions – marketing.

Content of multiple choice questions			
Question 1	Which of the following is non-store type of business?		
	1. Vending	2. Department stores	3. Shopping center
Question 2	The most simple pricing way is		
	1. The intuition price method	2. Cost plus pricing method	3. Target price method
Question 3	4P in marketing is price, place, promotion and		
	1. Packaging	2. trademark	3. Products
Question 4	In the market, some customers have the same preferences, known as a natural market segmentation, also known as		
	1. Homogeneous preferences	2. Dispersed preferences	3. Cluster preferences
Question 5	In the following items, which is the content of the product policy "extended product"?		
	1. Packaging	2. Brand	3. After-sales service

References

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. In *Proceedings of the international conference on web search and web data mining, February 11–12*. Palo Alto, CA, USA.

Ahn, J., Taieb-Maimon, M., Sopan, A., Plaisant, C., & Shneiderman, B. (2011). Temporal visualization of social network dynamics: Prototypes for nation of neighbors. *Social computing, behavioral-cultural modeling and prediction* (pp. 309–316).

Alexa (a). hudong.com. <http://www.alexa.com/siteinfo/hudong.com> Accessed 05.30.12.

Alexa (b). Wikipedia.org. <http://www.alexa.com/siteinfo/wikipedia.org> Accessed 04.03.12.

Baidu Baike. baidu.com. <http://baike.baidu.com/> Accessed 05.30.12.

Berendt, B., Hotho, A., & Stumme, G. (2002). Towards semantic web mining. In *Proceedings of the first international semantic web conference on the semantic web* (pp. 264–278).

Burkhard, R. A. (2004). Learning from architects: The difference between knowledge visualization and information visualization. In *Proceedings of IV'04, 8*.

Cheong, M., & Lee, V. (2009). Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In *Proceeding of the 2nd ACM workshop on social web search and mining* (pp. 1–8). Hong Kong: ACM.

Chiang, H. K., Chen, H. W., & Yang, J. R. (2008). The development and application of an automatic link analysis algorithm for social networks. *Journal of Information Management, 15*(3), 157–180.

Choudhury, De., Sundaram, M., John, H., & Seligmann, A. (2008). Can blog communication dynamics be correlated with stock market activity? *Hypertext 2008* (pp. 19–21). Pittsburg, PA, USA.

Cole, M. (2008). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers & Education, 52*(1), 141–146.

Cui, G., Lu, Q., Li, W., & Chen, Y. (2009). Mining concepts from Wikipedia for ontology construction. In *The 2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology*. *Web Intelligence & Intelligent Agent* (Vol. 03, pp. 287–290). IEEE Computer Society.

Erk, K. (2007). A simple, similarity-based model for selectional preferences. In *Proceedings of ACL-07* (pp. 216–223).

Foss, N. J., & Pedersen, T. (2002). Transferring knowledge in MNC's: The role of sources of subsidiary knowledge and organizational context. *Journal of International Management, 8*(1), 49–67.

Giacomo, P. (2008). *Wikipedia and the featured articles: How a technological system can produce best quality articles*, (Master thesis), University of Maastricht.

Hage, P., & Harary, F. (1983). *Structural models in anthropology*. New York: Cambridge University Press.

Haveliwala, T., Gionis, A., Klein, D., & Indyk, P. (2002). Evaluating strategies for similarity search on the web. In *WWW2002*.

He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th international conference on world wide web, April 26–30*. Raleigh, NC, USA.

Hudong. hudong.com. <http://www.hudong.com/> Accessed 05.30.12.

Kalakota, R., & Whinston, A. B. (1997). *Electronic commerce: A manager guide*. London: Addison-Wesley.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68.

Kim, M.-C., & Choi, K.-S. (1999). A comparison of collocation-based similarity measures in query expansion. *Information Processing and Management: An International Journal, 35*(1), 19–30.

Kim, S., Hwang, H., & Suh, E. (2003). A process-based approach to knowledge-flow analysis: A case study of a manufacturing firm. *Knowledge and Process Management, 10*(4), 260–276.

Kwak, H., Lee, C., Park, H., Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international World Wide Web (WWW) conference* (pp. 591–600). Raleigh, NC, USA.

Lai, C.-H., & Liu, D.-R. (2009). Integrating knowledge flow mining and collaborative filtering to support document recommendation. *Journal of Systems and Software, 82*(12), 2023–2037.

Leshed, G., Haber, E. M., Matthews, T., & Lau, T. (2008). CoScripter: Automating & sharing how-to knowledge in the enterprise. In *Proceedings of the twenty-sixth annual SIGCHI conference on human factors in computing systems, April 05–10*.

Libert, B., & Spector, J. (2007). *We are smarter than me: How to unleash the power of crowds in your business* (1st ed.). Upper Saddle River, NJ: Wharton School Publishing.

Lin, F. R., & Hsueh, C. M. (2006). Knowledge map creation and maintenance for virtual communities of practice. *Information Processing and Management, 42*, 551–568.

Lin, F. R., & Yu, J. H. (2009). Visualized cognitive knowledge map integration for P2P networks. *Decision Support System, 46*(4), 774–785.

Linn, L., & Slindle, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47*, 121–150.

Liu, D. R., & Lin, C. W. (2012). Modeling the knowledge-flow view for collaborative knowledge support. *Knowledge-based systems*.

Liu, D. R., Ke, C. K., Lee, J. Y., & Lee, C. F. (2008). Knowledge maps for composite E-services: A mining-based system platform coupling with recommendations. *Expert Systems with Applications, 34*(1), 700–716.
Ong, T. H., Chen, H., Sung, W. K., & Zhu, B. (2005). Newsmap: A knowledge map for online news. *Decision Support System, 39*, 583.
Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management, 24*(5), 513–523.
Sarma, D., Sarma, A. D., Gollapudi, S., & Panigrahy, R. (2010). Ranking mechanisms in twitter-like forums. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 21–30). New York, NY, USA: ACM.
Shee, D., & Wang, Y. (2008). Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Computer & Education, 50*, 894–905.
Singh, M. D., Krishna, P. R., & Saxena, A. (2009). A privacy-preserving jaccard similarity function for mining encrypted data. In *IEEE TENCON 2009 proc.* (pp. 1–4).
Smith, M., Shneiderman, B., Milic-Frayling, N., Rodrigues, E. M., Barash, V., Dunne, C., et al. (2009). Analyzing (social media) networks with NodeXL. In *Proceedings of the fourth international conference on communities and technologies*.
Stein, A., & Maier, E. (1995). Structuring collaborative information-seeking dialogues. *Knowledge-Based Systems, 8*(2–3), 82–93.
Steinbach, M., Karypis, G., Kumar, V. (2000). A comparison of document clustering techniques. In *Proc. TextMining workshop, KDD 2000*.
Strehl, A., Ghosh, J., & Mooney, R. (2000). Impact of similarity measures on web-page clustering. In *AAAI 2000 workshop on artificial intelligence for web search* (pp. 58–64).
Su, C. Y., & Wang, T. I. (2010). Construction and analysis of educational assessments using knowledge maps with weight appraisal of concepts. *Computers & Education, 55*, 1300–1311.
Vail, E. F. (1999). Mapping organizational knowledge. *Knowledge management review 8 (May/June)* (pp. 10–15).
Wang, T. I., Su, C. Y., & Hsieh, T. C. (2011). Accumulating and visualizing tacit knowledge of teachers on educational assessments. *Computers & Education*.
wiseGEEK. What is User Generated Content? <http://www.wisegeek.com/what-is-user-generated-content.htm> Accessed 06.02.12.
Wu, I. C., & Wu, C. Y. (2011). Using internal link and social network analysis to support searches in Wikipedia: A model and its evaluation. *Journal of Information Science, 37*(2), 189–207.
Xiangfeng, L., & Jie, Y. (2010). Building web knowledge flow based on interactive computing with semantics. *New Generation Computing, 28*, 113–121.
Yuan, S. T., & Sun, J. (2005). Ontology-based structured cosine similarity in document summarization: With applications to mobile audio-based knowledge management. *IEEE Transactions on Systems, Man, and Cybernetics – Part B Cybernetics, 35*(5).
Zhuge, H. (2002). A knowledge flow model for peer-to-peer team knowledge sharing and management. *Expert Systems with Applications, 23*(1), 23–30.