

# A semantic-expansion approach to personalized knowledge recommendation <sup>☆</sup>

Ting-Peng Liang <sup>a,\*</sup>, Yung-Fang Yang <sup>b</sup>, Deng-Neng Chen <sup>c</sup>, Yi-Cheng Ku <sup>d</sup>

<sup>a</sup> Department of Information Management and Electronic Commerce Research Center, National Sun Yat-sen University, Kaohsiung, Taiwan

<sup>b</sup> Electronic Commerce Research Center, Institute of Information Industries, Taipei, Taiwan

<sup>c</sup> Department of Management Information Systems, National Pingtung University of Science and Technology, Taiwan

<sup>d</sup> Department of Computer Science and Information Management, Providence University, Taiwan

Available online 13 May 2007

## Abstract

The rapid propagation of the Internet and information technologies has changed the nature of many industries. Fast response and personalized recommendations have become natural trends for all businesses. This is particularly important for content-related products and services, such as consulting, news, and knowledge management in an organization. The digital nature of their products allows for more customized delivery over the Internet. To provide personalized services, however, a complete understanding of user profile and accurate recommendation are essential.

In this paper, an Internet recommendation system that allows customized content to be suggested based on the user's browsing profile is developed. The method adopts a semantic-expansion approach to build the user profile by analyzing documents previously read by the person. Once the customer profile is constructed, personalized contents can be provided by the system. An empirical study using master theses in the National Central library in Taiwan shows that the semantic-expansion approach outperforms the traditional keyword approach in catching user interests. The proper usage of this technology can increase customer satisfaction.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Personalization; Customization knowledge management; Internet recommendation systems; Semantic-expansion method; Electronic commerce

## 1. Introduction

The rapid growth of the Internet has changed the nature of many businesses. The large amount of transactional data collected from the use of information

systems allows for better understanding of customer needs and the integration of knowledge for the customization of products and services. This is particularly important for content-based applications, such as consulting, news services, and knowledge management.

Due to the importance of product and service customization, *Internet recommendation systems* (also called the *Internet recommender systems*) have become an important research area in electronic commerce [46]. Its major purpose is to reduce irrelevant content and provide users with more pertinent information or product. A recent study indicates that the use of personalized recommendation can significantly increase user satisfaction due to its ability to offset information overload [27].

<sup>☆</sup> Note: The research was partially supported by the MOE Program for Promoting Academic Excellent of Universities under the grant number 91-H-FA08-1-4 and a research grant from National Science Council under the contract 92-2416-H-110-006-CC3. The paper was partially based on the thesis of the second author.

\* Corresponding author.

E-mail addresses: [tpliang@facmail.nsysu.edu.tw](mailto:tpliang@facmail.nsysu.edu.tw) (T.-P. Liang), [david@universalec.com](mailto:david@universalec.com) (Y.-F. Yang), [dnchen@mail.npust.edu.tw](mailto:dnchen@mail.npust.edu.tw) (D.-N. Chen), [ycku@pu.edu.tw](mailto:ycku@pu.edu.tw) (Y.-C. Ku).

Many information filtering and recommendation methods have been developed in literature, most existing techniques fall into three categories: rule-based filtering, content-based filtering, and collaborative filtering. *Rule-based filtering* uses pre-specified *if-then* rules to select relevant information for recommendation. *Content-based filtering* uses keywords or other product-related attributes to make recommendations. *Collaborative filtering* uses preferences of similar users in the same reference group as a basis for recommendation.

Content-based filtering and collaborative filtering are more popular in practical applications. However, they both have limitations. The content-based approach can classify services based on their nature, but often have difficulties in identifying related interests of the same user. Collaborative filtering can find similarities among different users but is unable to handle new items that do not have existing usage information.

Content-based filtering is better than collaborative filtering when it is applied to digital products such as customized news services and document recommendation in knowledge management because documents like reports have certain semantic linkages that cannot be captured by collaborative filtering.

Two general directions are popular for using these methods: profile generation and maintenance, and profile exploitation. *Profile generation and maintenance* include user profile representation, profile generation, and relevance feedback. *Profile exploration* includes information filtering, user profile-item matching, and profile adaptation [34]. Profile generation explores the interests of a particular user, while profile exploration finds information relevant to a particular user query for recommendation. Since most queries use keywords for document search, information retrieval techniques can facilitate content-based recommendations. For example, if a user reads a report about knowledge management and auctions, recommending reports of interest to the user is similar to retrieving documents using *knowledge management* and *auction* as two keywords.

Typical research in information retrieval uses key word weighting to find documents relevant to a particular query [5]. Since key words in a user query often have semantic meanings and certain semantic relationships may exist in certain documents, simple key word matching may result in an underweight or overweight of certain key words due to their semantic similarities. Advanced techniques that take semantics into consideration in building and exploitation of user profiles are useful.

In this paper, we propose a semantic-expansion approach to build user profile and content recommendations. This approach uses semantic networks and the

spreading activation model (SA) in cognitive psychology to build user profiles and then make recommendations accordingly [12,13]. The method includes three modules: (1) analyzing document structures, (2) building user interest profiles, and (3) making recommendations. An experiment was conducted to compare the performance between the semantic-expansion approach and a typical keyword-based approach. The result indicates that the semantic-expansion approach significantly outperformed the key-word-only approach.

The remainder of this paper is organized as follows. In the next section, literature in information filtering and content-based recommendation is reviewed. In Section 3, the semantic-expansion approach is described. Section 4 presents a prototype implementation and findings from our experimental study. Finally, conclusions and suggestions for future research are discussed.

## 2. Literature review

### 2.1. Personalization and recommendation systems

Personalization is defined as “the ability to provide content and services tailored to individuals based on knowledge about their preferences and behavior” or “the use of technology and customer information to tailor electronic commerce interactions between a business and each individual customer” [1]. A major vehicle that makes personalization possible is the recommendation system that matches potential products with customer preferences.

A recommendation system is a computer-based system that uses profiles built from past usage behavior to provide relevant recommendations. For instance, a video rental store may analyze historical rental data to recommend new movies to individual customers. An online newspaper may customize news services to different readers based on their reading interests. The objective of a recommendation system is to retrieve information of interest to users from a large repository. Such systems can reduce the search efforts of users and mitigate the encumbrance of information overload [28].

There are several research streams in personalized recommendation. One stream aims at improving the accuracy of algorithms used in recommendation systems [2,3,9,10,19,26,31,52]. The second stream is focused on the interaction between a recommendation system and its customers. For instance, some studies investigated the persuasive effect of recommendation messages [49,50], developed better data collection mechanisms [38], and enhanced awareness about privacy issues [4]. Furthermore, a few studies focused on the effect of

moderating factors such as user characteristics and product features on the performance of recommendation [20,21,39,48].

## 2.2. Approaches to personalized recommendation

A typical personalization process includes three steps: understanding customers through profile building, delivering personalized offering based on the knowledge about the product and the customer, and measuring personalization impact [1]. Montaner, et al. [35] simplified the process into two stages after analyzing 37 different systems: profile generation and maintenance, and profile exploitation.

One key to the performance of personalized recommendation is the nature of the mechanism it uses to build customer profiles. In previous research, a number of different algorithms have been proposed. These systems are classified based on various characteristics. For example, Beyah et al. [7] divided recommendation systems into four types: collaborative filtering (people to people correlations), social filtering, non-personalized recommendation systems, and attribute-based recommendations. Wei et al. [53] classified recommendation systems into six approaches based on the type of data and technique used to arrive at recommendation decisions. These approaches are recommendations based on: popularity, content, association, demographics, reputation and collaboration.

Schafer et al. [47] argues that recommendation systems may be categorized by the data they use, which includes original data search, expert ratings, statistical rankings, content-based recommendation, object associations, and user associations. A system using the original data search does not analyze user profiles. Rather, it provides a flexible user interface and allows users to query the database directly. Expert ratings use comments or ratings from experts in the domain (e.g., music or movies) and make recommendations accordingly. Statistical ranking is a simple but popular method, which uses descriptive statistical data such as order frequency to rank different objects for recommendation. Content-based recommendation uses attributes of the content to match user interest profiles. Object associations use found relationships among objects to make recommendations. A popular example is the market basket analysis that finds items often ordered simultaneously by a customer.

There exist three popular methods for extracting user preferences: direct, semi-direct, and indirect extraction [45]. The direct approach asks the user to tell the system explicitly what he prefers. For instance, a knowledge management system may list all document categories

and ask the user to check those of interest to him. The semi-direct approach asks the user to rate all documents he has read and gains knowledge of user preference through these ratings [25]. The indirect approach captures user preference from browsing behavior recorded by the computer, such as hyperlink clicks [40] or time spent on reading a document [27].

For making recommendations, two approaches are popular: collaborative filtering and content-based filtering [6,23]. *Collaborative filtering* is an approach to making recommendations by finding correlations among the shared likes and dislikes of the system users. It is capable of finding items of potential interest from ratings of previous users. *Content-based filtering* makes recommendation by analyzing the items rated by the user and the item to be recommended [43]. Generally speaking, content-based information filtering has proven to be effective in locating textual documents relevant to a topic [36]; whereas collaborative filtering is popular with e-tailors that sell physical products such as in Amazon.com [30]. The integration of these two types of filtering is reported to exhibit good performance in some domains (e.g., [2,31,42]).

## 2.3. Personalized document recommendation

Personalized recommendations are applied to both physical and digital/information products. Recent application domains include books [30], news [22,32], movies [10,37], advertisements [22], one-to-one marketing campaigns [52], and bundle purchases [17]. Although Amazon.com has been applauded for its success in using personalized recommendation, information goods such as news and documents in digital libraries are popular for personalization on the Internet due to its nature of content modularity.

The application of personalized recommendation to information goods also has a long history. For example, Mock and Vemuri [34] developed the Intelligent News Filtering Organization System (INFOS) that reorganizes the order of news based on revealed preferences. The results of a pilot test show that INFOS can effectively reduce the reader's search load. Sakagami and Kamba [45] developed the ANATAGONOMY system that learns reading preferences from the browsing behavior of a user; a learning engine and a scoring engine produce personalized Web news. Billsus and Pazzani [8] designed an intelligent news agent to automatically learn the user's preferences and provide personal news. Mooney and Roy [36] proposed a content-based book recommending system that utilizes information extraction for text categorization and produces accurate

recommendations. Lai et al. [24] designed a news recommendation system based on customer profiles to provide customized Internet news. Fan, et al. [15] presented a method for generating profiles of news readers. Liang, et al. [28] also reported experimental findings that personalized services produced significantly higher user satisfaction for online news readers.

#### 2.4. Information retrieval techniques

The technique used for identifying relevant items for recommendation is called *information retrieval (IR)*, which begins with a user query and searches for relevant documents from a large database [51,55]. For example, if a user is looking for documents related to *knowledge map*, these two words are entered into the system and then transferred into queries. The main function of the system is to match user queries with the features of documents stored in the database. Information retrieval is different from information filtering in that its major goal is to find relevant documents that meet the needs of different ad hoc queries [18]. However, both approaches share many techniques and applications in document categorization and extraction.

For processing information goods, information retrieval systems may use full-text search or keyword-based search [29]. Based on their query styles, document representation methods, and matching results, IR systems can further be classified into four basic models: Boolean model, probabilistic model, vector space model, and linguistic model [41]. Boolean models provide Boolean operators such as AND, OR, and NOT to increase search flexibility. Probabilistic models use keywords to estimate the probability that a document matches the query [16,46]. Vector space models (VSM) utilize arrays of keywords to represent queries and use their similarities to make recommendations [46]. Linguistic models use natural languages to process queries [11].

Most IR systems developed recently use the VSM because its retrieval performance is better than that of the traditional Boolean model. However, the VSM method may encounter problems that lead to inefficient searches [56]. For example, a user query may contain only two or three terms, which is inadequate for locating a highly relevant document.

One possible solution for overcoming the problem of inadequate information is to expand the query by adding more semantic information to better describe the concepts. However, adding appropriate terms to expand the queries is not easy. Relevance feedbacks and knowledge structure are used to locate terms for expansion. Relevance feedbacks are information on the

items selected by the user from the output of previous queries. IR systems can modify original queries based on the relevance feedback to retrieve documents that truly satisfy users [44].

An alternative approach to query expansion is to use existing knowledge structure such as the *Longman Dictionary of Contemporary English* or *WordNet* [33]. Several methods have been proposed for knowledge-based query expansion. They include the use of co-occurrence data [54], document classification, syntactic context [33], and spreading activation [13,14].

#### 2.5. The Spreading Activation Model

The Spreading Activation (SA) Model is an extension of content-based filtering that is popular in knowledge recommendation. Concepts are expanded based on the semantics in the process of identifying customer profile and matching items and the model has been applied to expand queries in information retrieval [13]. Since a large amount of semantic information may exist in most documents, it is useful to include this information in making recommendations.

The SA model was developed in cognitive psychology as a mechanism for interpreting how semantic networks function in human brains [12]. It includes two major components: a spreading activation network and an activation spreading mechanism. Spreading activation networks are similar to semantic networks. A network consists of nodes and links that are gradually accumulated from the life experiences. A node represents a concept and a link represents the semantic relationship between two concepts.

A spreading mechanism includes several major steps: (1) adjusting inputs, (2) concept spreading, (3) calculating outputs, and (4) spreading termination. Spreading is a diffusion process among nodes. Two actions control the spreading handling procedures, namely pulse spreading and termination check. A pulse continuously spreads data to surrounding nodes and includes three handling actions, namely input value, message spreading, and output value adjustment. Input and output value adjustments control the range and weight of spreading. A termination check is used to determine whether the termination condition is met. This method allows semantic information to be used to expand a query.

### 3. A semantic-expansion approach to document recommendation

As described before, content-based filtering is proven useful in recommending information goods such as

organizational documents. If only keywords in the content are used, some semantic information will be missing and some important cues may not be captured. Therefore, we propose the *semantic-expansion approach* that integrates semantic information for spreading expansion and content-based filtering for document recommendation. The proposed system includes three main modules: user preference extraction, semantic-expansion, and document recommendation, as shown in Fig. 1. Each module is described in detail below.

### 3.1. User preference extraction

The first step for making recommendations is to find user interests from the documents he or she has read previously. A typical approach for representing the content of a document in information retrieval and filtering is to use a keyword vector. For instance, this paper may be represented as a vector, [knowledge management, recommendation system, semantic-expansion method, electronic commerce].

The profile of user interest is then extracted from the relevance feedback provided by the user through rating the documents they have read. The rating scores are indicators of their relative interests in the concepts represented as key words in the documents. Documents scored higher by users imply that they had higher interests [25]. Since each document has only one score, all keywords in the document are assigned the same interest score.

As keywords may contain synonyms, similar keywords are grouped together to become a concept. For example, Internet marketing and e-marketing are quite similar and may be considered to be the same concept.

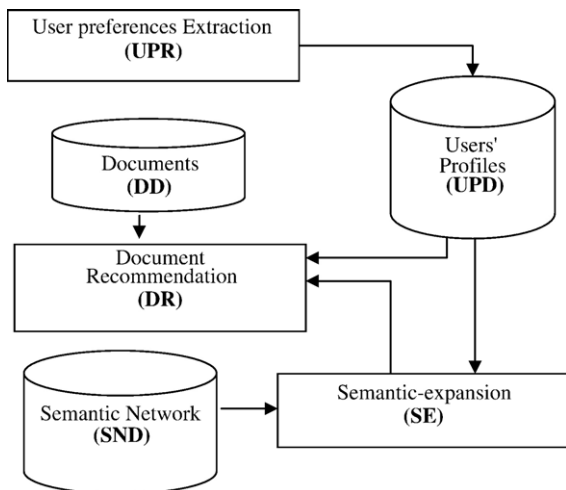


Fig. 1. Structure of the recommendation mechanism.

Table 1  
User ratings of two sample documents

Document	Keyword	Score
A	Consumer purchasing behavior, electronic store, adaptive website	4
B	Marketing research, consumer behavior, personalized website	6

When two documents containing the same concept but different scores, we use temporal information such as the rating time to modify the score. A sample mechanism is that the old interest score will increase by one point if the score of a newer document containing the same concept is higher than the previous one. Oppositely, the interest score will be reduced by one point if the score of the new document containing the same concept is lower than the old one.

Table 1 shows the user ratings of two documents read by a user sequentially and their interest scores are 4 and 6, respectively. Since *consumer behavior* (document B) and *consumer purchasing behavior* (document A) belong to the same concept and document B has a higher interest score, the score of *consumer purchasing behavior* is rated 4+1, rather than replacing it with the score of document B (which is 6). Based on the rule, the adjusted score of the concepts in the two documents are showed in Table 2. This allows the basic element, i.e., the interest level of individual concept, to be established.

### 3.2. Semantic-expansion network

After the concepts of interest to a user are extracted from the reading history, these concepts can be further expanded with the semantic-expansion module that consists of a library of semantic trees and a set of spreading rules.

Table 2  
Scores of derived concepts

Keyword	Concept	Score after reading A	Score after reading B
Consumer purchasing behavior	Consumer behavior	4	5
Electronic store	Electronic store	4	4
Adaptive website	Personalized website	4	5
Marketing research	Marketing strategy		6
Consumer behavior	(Consumer behavior)		
Personalized website	(Personalized website)		

Note: Concepts in parentheses mean they are a duplication of existing concepts.

A semantic-expansion network is composed of a set of networked semantic trees. A semantic tree is a group of related *concepts* and *relationships*. Each node represents a concept, whereas a link between two nodes represents their relationship. Two kinds of relationships exist. The solid arrow shows an *is-a* relationship that implies certain property inheritance. The dashed line shows a *non-is-a* relationship, i.e., all other possible associations between two semantic trees with no property inheritance. A semantic-expansion network is a set of interconnected semantic trees. Fig. 2 shows a sample semantic-expansion network.

Given the semantic-expansion network, there are three major strategies for spreading: generalization, specialization, and relevance expansion. *Generalization* is the process of activating a concept above the current concept in the semantic network (e.g., activating the concept *personalized website* to the concept *WWW* in Fig. 2). *Specialization* is the process of activating a concept below the current concept (e.g., activating the concept of *brand strategy* from *marketing strategy* in Fig. 2). *Relevance expansion* is the process of activating a concept through a non-is-a relationship in the network (e.g., activating the concept of *consumer behavior* from *electronic store* in Fig. 2). Each strategy is given a weight between 0 and 1 to show a decrease in interests during the spreading process. The exact weight may be adjusted by the user or through a specific regression process. If two concepts are exactly the same, then the weight for their link is 1.0.

Using these three strategies, the basic user profile can be expanded to become more complete. Three para-

eters are essential in the semantic spreading process: activation value, spreading distance, and threshold. The *activation value* is the interest score of a concept during the spreading activation process. Its initial value is often set to zero. *Spreading distance* is the maximum number of levels that the system plans to spread. The longer the activation distance, the more likely irrelevant concepts may be activated. The distance is generally set to 2 (e.g., if we start from *marketing strategy* in Fig. 2, the system will activate concepts down to *Internet advertising* while performing specialization). *Threshold* is the minimum activation value necessary for a concept to activate its next-level concepts. That is, the concept spreading will stop when the activation value of a concept is below the threshold or the maximum activation distance is reached. Given the above information, activation spreading is an iterative process that calculates activation values for all relevant concepts and puts together the final interest profile.

The spreading process is quite straightforward. If the activation value of a node exceeds the threshold and the maximum spreading distance has not been reached, then the system will spread to find its upper-level nodes (generalization), lower-level nodes (specialization), and association nodes (relevance). The contribution of spreading from an original node to the destination node is the activation value of the original node multiplied by the weight of the spreading strategy. That is, if the original node has a value of 5 and it reaches a node through generalization with an assigned weight of 0.6, then the activation value of the activated destination

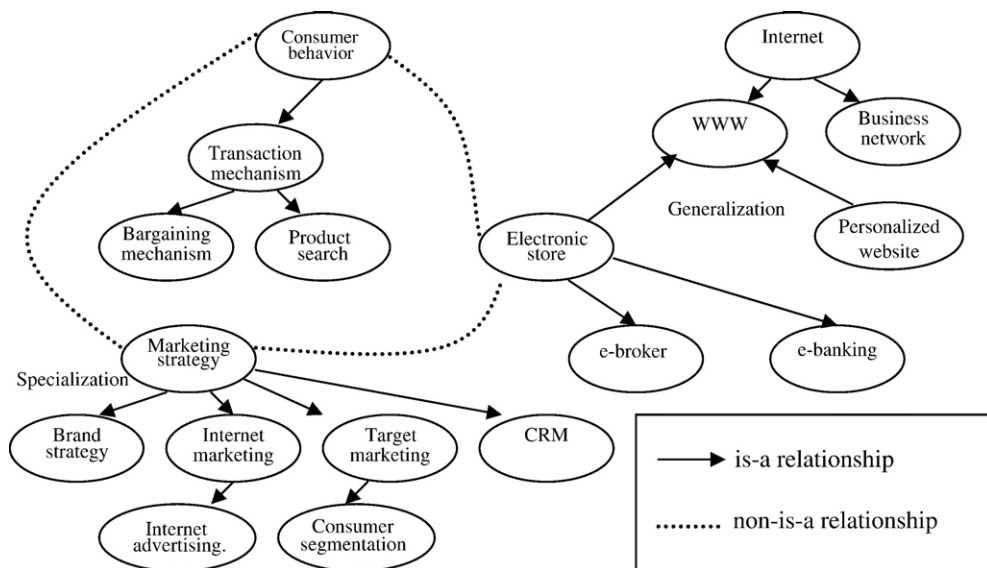


Fig. 2. A sample semantic-expansion network.

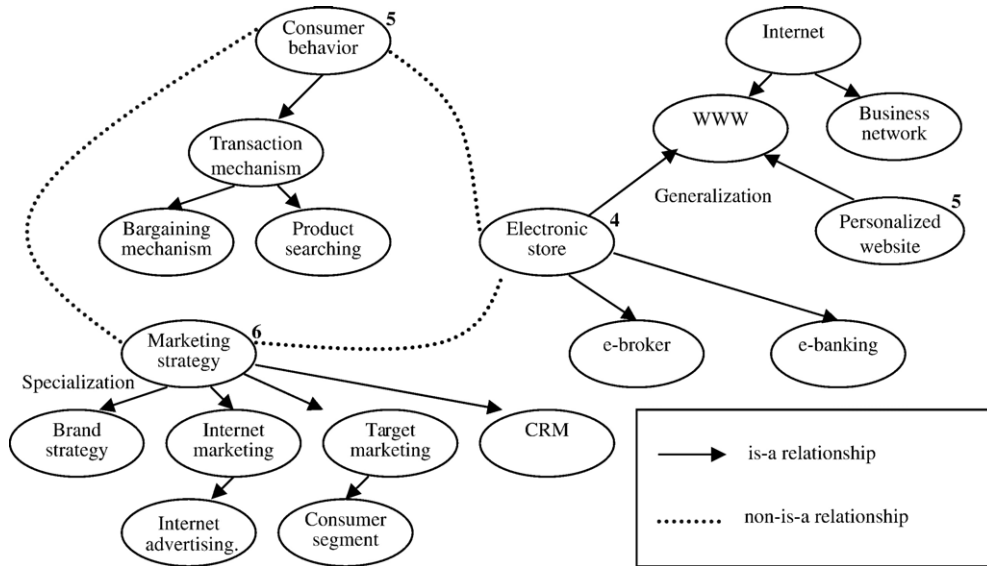


Fig. 3. Initial basic user profile.

node is 3.0 (5 \* 0.6). If a node received activations from multiple sources, then its final activation value is the sum of the values from all incoming sources.

We use the same example shown in Table 2 and Fig. 2. The initial basic interests can be represented as a vector [consumer behavior (5), electronic store (4), personalized website (5), marketing strategy (6)], as shown in Fig. 3.

We set the following parameters:

- (1) initial activation value of other nodes = 0;
- (2) spreading distance = 1;

(3) threshold value = 2.4;

(4) generalization weight: 0.7; specialization weight: 0.4; relevance weight: 0.5.

After the spreading activation, the final user profile is shown in Fig. 4. Since the threshold value is 2.4, those concepts with activation values below the hurdle will not be included in the final expanded profile. The concept expanded from *electronic store* and *marketing strategy* includes *WWW*, *target marketing*, *Internet marketing*, *brand strategy*, and *CRM*. Table 3 shows

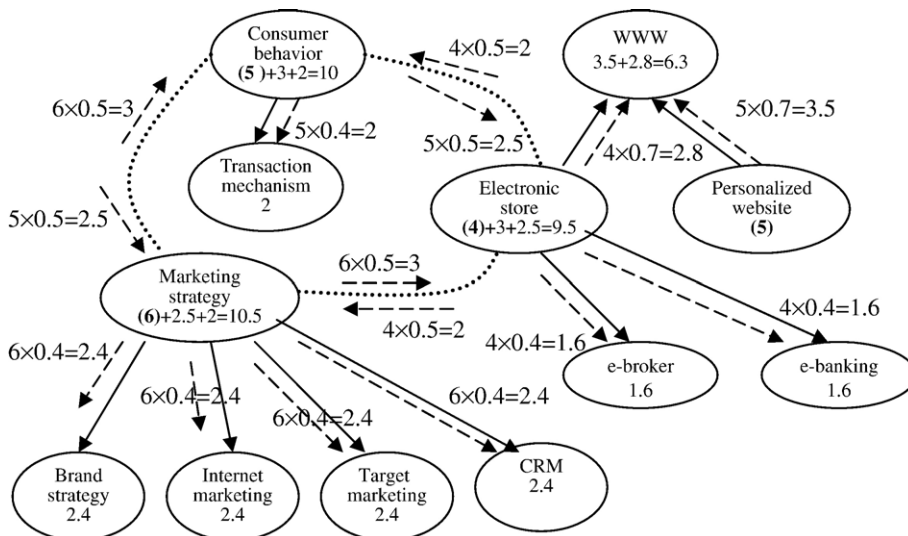


Fig. 4. User profile after spreading.

Table 3  
Result of the semantic-expansion example

Concepts	Activation values	Expanded concept	Concepts	Activation values	Expanded concept
Consumer behavior*	10		Internet marketing	2.4	√
Electronic store*	9.5		Brand strategy	2.4	√
Personalized website*	5		CRM	2.4	√
Marketing strategy*	10.5		Transaction mechanism	2	NA
WWW	6.3	√	E-banking	1.6	NA
Target marketing	2.4	√	E-broker	1.6	NA

Note: \* Initial concepts; NA indicates not included due to values below the threshold.

the resulting activation values of the initial concepts and expanded concepts.

### 3.3. Document rating and recommendation

After obtaining the expanded user profile, we use this profile to assess the potential interest of the user in various documents. This involves matching keywords of the documents and the concepts in the expanded user profile. Since each concept may match with multiple keywords in a document, the concepts in a user profile are converted into keywords before matching with document keywords. Table 4 shows a sample conversion between concepts in Table 3 and possible keywords. All keywords derived from a concept are assigned the same interest value initially.

Once the interest values of keywords are determined, the rating of a document is the aggregation of all interest values of its keywords. For example, if a document has

three keywords [Electronic CRM, Data Mining, and marketing strategy], then we can find that the interest value of these keywords to the user are 2.4, 0 and 10.5, respectively. Its total interest value is 12.9. The system can then make recommendations according to their interest values of the analyzed documents.

## 4. An experimental study

In order to evaluate whether the semantic-expansion approach can improve the performance of recommendation systems, a prototype system was developed and an experiment was conducted in the computer lab. The prototype system was implemented in the Microsoft Windows environment and development in ASP, VB Script, and SQL Server. Fig. 5 illustrates its architecture.

The documents database contained 200 master theses or doctoral dissertations in information systems that were sampled from the National Central Library in Taiwan (<http://www.ncl.edu.tw>). These documents included abstracts and keywords. Four graduate students constructed a semantic-expansion network that included 265 concepts, 470 keywords, and 56 semantic trees.

Table 4  
Interest values of keywords

Concepts	Activation value	Keywords	Interest score
Consumer behavior	10	Consumer behavior, Shopping on the Internet, Consumer purchase behavior	10
Electronic store	9.5	Electronic store	9.5
Personalized website	5	Adaptive website, Personalized website	5
Marketing strategy	10.5	Marketing strategy, Marketing research	10.5
WWW	6.3	WWW, Internet	6.3
Target marketing	2.4	Target marketing	2.4
Internet marketing	2.4	Internet marketing, Network marketing	2.4
Brand strategy	2.4	Brand strategy, Online brand	2.4
CRM	2.4	CRM, Electronic CRM, Customer maintenance	2.4

### 4.1. Experimental design

The benchmark for comparison was the traditional keyword-based approach as shown in Eq. (1) below. The weight of a keyword was measured by the frequency in which it occurred in the same document divided by the frequency of all keywords. Since the sample document didn't have full texts, the weight of keywords was assigned 1 and the equation was revised to Eq. (2) as our benchmark.

$$\begin{aligned} &\text{Interest value of a document} \\ &= \sum \text{interest value of a keyword} \times \text{weight} \quad (1) \end{aligned}$$

$$\begin{aligned} &\text{Interest value of a document} \\ &= \sum \text{interest value of a common keyword} \quad (2) \end{aligned}$$



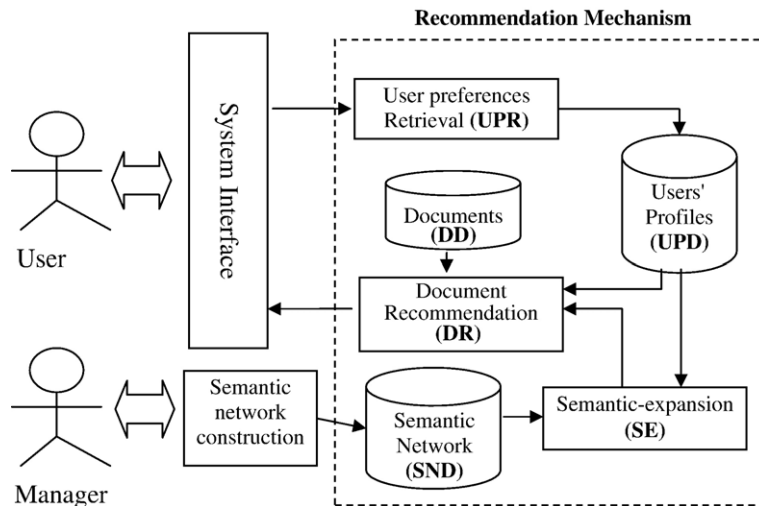


Fig. 5. Architecture of the experimental system.

One hundred twenty-nine students who majored in information systems were recruited to participate in the experiment. The experimental procedures were:

First, the system randomly selected 30 articles from the database. The subject read the articles, chose ten articles of interest to them, and gave an interest score to each of them on a scale between 1 and 7. Then, the system analyzed user interests and recommended 20 articles for further reading, 10 by the semantic-expansion approach and 10 by the traditional keyword approach. Articles recommended by different approaches were mixed to show in alphabetical order and could not be identified by the subject. Finally, the subjects were asked to read all recommended articles and give them their interest ratings.

#### 4.2. Experimental results

After removing incomplete data and outliers, the total effective sample size for analysis was 103. The mean and standard deviation of their interests on articles recommended by different methods are shown in Table 5. We can see that the articles recommended by the semantic-expansion approach better caught user interests,

Table 5  
Mean and standard deviation of two recommendation methods

Recommendation method	Mean	SD	t-test
Semantic-expansion	5.18	1.30	$P=0.008 (<0.01)$
Keywords	5.03	1.38	

compared with the keyword approach (5.18 vs. 5.03 on a 7-point scale). Since all subjects tried both mechanisms, the paired *t*-test can be used to examine the significance of the difference. The results show that the difference is significant statistically at 0.01 level.

In addition to comparing user interests in articles recommended by different methods, the correlation between concept activation values and user interest levels was also measured. The results from the Pearson correlation analysis indicate a positive correlation between them (coefficient=0.092,  $p<0.01$ ). Even though the coefficient is not very large, it does show that concepts with higher activation values are more likely to meet user preference and the semantic-expansion approach is capable of improving user satisfaction on personalized content recommendation.

#### 5. Concluding remarks

Personalization has been a major trend for e-commerce. Using recommendation systems to provide customized information services will be the mainstream in the future. In this paper, we have presented a semantic-expansion approach to document recommendation. It adopts the spreading activation model to broaden the scope for user profile analysis. A major feature of this method is the construction of a semantic-expansion network that includes “is-a” and “non-is-a” relationships to connect concepts. Results from an experimental study show that the proposed approach performs significantly better than the traditional keyword-based approach in capturing user interests.

Since the proposed method relies heavily on the semantic-expansion network for concept spreading, a major concern is how to build comprehensive and useful semantic-expansion networks to cover major concepts and their relationships. This requires a great deal of work from professionals in the application domain. Another limitation of the approach is that it requires the user to provide relevance feedbacks in order to build the user profile. This is important at least at the early application stage of the system. If the user is reluctant to provide feedback, then it is difficult for the system to build user interest profiles accurately. Unfortunately, this is a common drawback for all content-based recommendation methods. One way to alleviate the problem is to use implicit cues such as reading time as a substitute. Nonetheless, the findings in this study still provide much valuable insight into integrating semantic information in information filtering and document recommendation.

## References

- [1] G. Adomavicius, A. Tuzhilin, Personalization techniques: a process-oriented perspective, *Communications of the ACM* 48 (10) (2005) 83–90.
- [2] H.J. Ahn, Utilizing popularity characteristics for product recommendation, *International Journal of Electronic Commerce* 11 (2) (2006) 59–80.
- [3] A. Ansari, S. Essegai, R. Kohli, Internet recommendation systems, *Journal of Marketing Research* 37 (3) (2000) 363–375.
- [4] N.F. Awad, M.S. Krishnan, The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization, *MIS Quarterly* 30 (1) (2006) 13–28.
- [5] R. Baeza-Yates, B. Ribeiro-Neto, *Modern Information Retrieval*, ACM Press, Addison-Wesley, 1999.
- [6] M. Balabanović, Y. Shoham, Fab: content-based, collaborative recommendation, *Communications of the ACM* 40 (3) (March 1997) 66–72.
- [7] G. Beyah, P. Xu, H.G. Woo, K. Mohan, D. Straub, Development of an instrument to study the use of recommendation systems, *Proceedings of the Ninth Americas Conference on Information Systems*, Tampa, FL, USA, 2003, pp. 269–279.
- [8] D. Billsus, M.J. Pazzani, A personal news agent that talks, learns and explains, *Proceedings of the Third Annual Conference on Autonomous Agents*, Seattle, Washington, USA, 1999, pp. 268–275.
- [9] K.W. Cheung, J.T. Kwok, M.H. Law, K.C. Tsui, Mining customer product ratings for personalized marketing, *Decision Support Systems* 35 (2) (2003) 231–243.
- [10] K.W. Cheung, K.C. Tsui, J.M. Liu, Extended latent class models for collaborative recommendation, *IEEE Transactions on Systems, Man, and Cybernetics (Part A)* 34 (1) (2004) 143–148.
- [11] Y. Chieramella, B. Defude, A prototype of an intelligent system for information retrieval: IOTA, *Information Processing & Management* 23 (4) (1987) 285–303.
- [12] A.M. Collins, E.F. Loftus, A spreading-activation theory of semantic processing, *Psychological Review* 82 (6) (1975) 407–428.
- [13] F. Crestani, Application of spreading activation techniques in information retrieval, *Artificial Intelligence Review* 11 (6) (December 1997) 453–482.
- [14] F. Crestani, P.L. Lee, Searching the web by constrained spreading activation, *Information Processing & Management* 36 (4) (July 2000) 585–605.
- [15] W. Fan, M.D. Gordon, P. Pathak, Effective profiling of consumer information retrieval needs: a unified framework and empirical comparison, *Decision Support Systems* 40 (2005) 213–233.
- [16] N. Fuhr, H. Hüther, Optimum probability estimation from empirical distributions, *Information Processing & Management* 25 (5) (1989) 493–507.
- [17] R. Garfinkel, R. Gopal, A. Tripathi, F. Yin, Design of a shopbot and recommender system for bundle purchases, *Decision Support Systems* 42 (2006) 1974–1986.
- [18] U. Hanani, B. Shapira, P. Shoval, Information filtering: overview of issues, research and systems, *User Modeling and User-Adapted Interaction* 11 (August 2001) 203–259.
- [19] J.L. Herlocker, J.A. Konstan, A. Borchers, J. Riedl, An algorithmic framework for performing collaborative filtering, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, California, USA, 1999, pp. 230–237.
- [20] L.P. Hung, A personalized recommendation system based on product taxonomy for one-to-one marketing online, *Expert Systems with Applications* 29 (2) (2005) 383–392.
- [21] I. Im, A. Hars, Finding information just for you: knowledge reuse using collaborative filtering systems, *Proceedings of the Twenty-Second International Conference on Information Systems*, New Orleans, Louisiana, USA, 2001, pp. 349–360.
- [22] J.W. Kim, K.K. Lee, M.J. Shaw, H.L. Chang, M. Nelson, R.F. Easley, A preference scoring technique for personalized advertisements on Internet storefronts, *Mathematical and Computer Modeling* 44 (2006) 3–15.
- [23] J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl, GroupLens: applying collaborative filtering to Usenet news, *Communications of the ACM* 40 (3) (March 1997) 77–87.
- [24] H.J. Lai, T.P. Liang, Y.C. Ku, Customized Internet News Services Based on Customer Profiles, *Proceedings of the Fifth International Conference on Electronic Commerce*, Pittsburgh, PA, USA, 2003, pp. 225–229.
- [25] K. Lang, et al., NewsWeeder: learning to filter netnews, *Proceedings of the 12th International Conference on Machine Learning*, Tahoe City, California, 1995, pp. 331–339.
- [26] W.P. Lee, Applying domain knowledge and social information to product analysis and recommendations: an agent-based decision support systems, *Expert Systems* 21 (3) (2004) 138–148.
- [27] T.P. Liang, H.J. Lai, Discovering user interests from web browsing behavior: an application to Internet news services, *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, Big Island, Hawaii, USA, 2002, pp. 203–212.
- [28] T.P. Liang, H.J. Lai, Y.C. Ku, Personalized content recommendation and user satisfaction: theoretical synthesis and empirical findings, *Journal of Management Information Systems* 23 (3) (2006) 45–70.
- [29] S.H. Lin, C.S. Shih, M.C. Chen, J.M. Ho, M.T. Ko, Y.M. Huang (Eds.), *Extracting classification knowledge of Internet documents with mining term associations: a semantic approach*, *ACM SIGIR'98*, 1998, pp. 241–249.
- [30] G. Linden, B. Smith, J. York, Amazon.com recommendations: item-to-item collaborative filtering, *IEEE Internet Computing* 7 (1) (2003) 76–80.

- [31] D.R. Liu, Y.Y. Shih, Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences, *Journal of Systems and Software* 77 (2) (2005) 181–191.
- [32] M. Maybury, W. Greiff, S. Boykin, J. Ponte, C. McHenry, L. Ferro, Personalcasting: tailored broadcast news, user modeling and user-adapted interaction, *User Modeling and User-Adapted Interaction* 14 (2004) 119–144.
- [33] G.A. Miller, WordNet: a lexical database for English, *Communications of the ACM* 38 (11) (1995) 39–41.
- [34] K.J. Mock, V.R. Vemuri, Information filtering via hill climbing, WordNet, and index patterns, *Information Processing & Management* 33 (5) (1997) 633–644.
- [35] M. Montaner, B. Lopez, L.D.A. Mosa, A taxonomy of recommender agents and the Internet, *Artificial Intelligence Review* 19 (2003) 285–330.
- [36] R.J. Mooney, L. Roy, Content-based book recommending using learning for text categorization, *Proceedings of the fifth ACM conference on Digital libraries*, San Antonio, Texas, United States, 2000, pp. 195–204.
- [37] R. Mukherjee, N. Sajja, S. Sen, A movie recommendation system — an application of voting theory in user modeling, *User Modeling and User-Adapted Interaction* 13 (2003) 5–33.
- [38] B.P.S. Murthi, S. Sarkar, The role of the management sciences in research on personalization, *Management Science* 49 (10) (2003) 1344–1362.
- [39] R. Nikolaeva, S. Sriram, The moderating role of consumer and product characteristics on the value of customized on-line recommendation, *International Journal of Electronic Commerce* 11 (2) (2006) 101–123.
- [40] D.W. Oard, J. Kim, Implicit feedback for recommender systems, *Proceedings of AAAI Workshops on Recommender Systems*, Madison, WI, 1998, pp. 81–83.
- [41] C.D. Paice, A thesaural model of information retrieval, *Information Processing & Management* 27 (5) (1991) 433–447.
- [42] H. Paik, B. Benatallah, Building adaptive e-catalog communities based on user interaction patterns, *IEEE Intelligent Systems* 17 (6) (2002) 44–52.
- [43] M.J. Pazzani, A framework for collaborative, content-based and demographic filtering, *Artificial Intelligence Review* 13 (5–6) (1999) 393–408.
- [44] J. Rocchio (Ed.), *Relevance feedback in information retrieval, The SMART Retrieval System-Experiments in Automatic Document Processing*, Prentice Hall Inc, 1971, pp. 313–323.
- [45] H. Sakagami, T. Kamba, Learning personal preferences on online newspaper articles from user behaviors, *Computer Networks and ISDN Systems* 29 (1997) 1447–1455.
- [46] G. Salton, M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw–Hill, New York, 1983.
- [47] J.B. Schafer, J.A. Konstan, J. Riedl, E-commerce recommendation applications, *Data Mining and Knowledge Discovery* 5 (1) (2001) 115–153.
- [48] S. Senecal, J. Nantel, The influence of online product recommendations on consumers' online choices, *Journal of Retailing* 80 (2004) 159–169.
- [49] S.N. Singh, N.P. Dalal, Web home pages as advertisements, *Communications of the ACM* 42 (8) (1999) 91–98.
- [50] K.Y. Tam, S.Y. Ho, Web personalization as a persuasion strategy: an elaboration likelihood model perspective, *Information Systems Research* 16 (3) (2005) 271–291.
- [51] H.C. Tu, J. Hsiang, An architecture and category knowledge for intelligent information retrieval agents, *Decision Support Systems* 28 (3) (2000) 255–268.
- [52] S.S. Weng, M.J. Liu, Feature-based recommendation for one-to-one marketing, *Expert Systems with Applications* 26 (4) (2004) 493–508.
- [53] C.P. Wei, M.J. Shaw, R.F. Easley, A survey of recommendation systems in electronic commerce, in: R.T. Rust, P.K. Kannan (Eds.), *e-Service: New Directions in Theory and Practice*, M.E. Sharpe Publisher, 2002.
- [54] J. Wei, S. Bressan, B.C. Ooi, Mining term association rules for automatic global query expansion: methodology and preliminary results, *Proceedings of the First International Conference on Web Information Systems Engineering*, Hong Kong, China, 2000, pp. 366–373.
- [55] J. Xu, Solving the Word Mismatch Problem through Automatic Text Analysis, PhD Thesis, Department of Computer Science, University of Massachusetts at Amherst, MA, USA (1997).
- [56] J. Xu, W.B. Croft, Query expansion using local and global document analysis, *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zurich, Switzerland, 1996, pp. 4–11.



**Ting-Peng Liang** (Fellow, AIS) is the Director of Electronic Commerce Research Center and National Chair Professor of Information Management at the National Sun Yat-sen University in Taiwan. Prior to the current position, he had been Dean of Academic Affairs and Dean of the College of Management, Director of the Graduate Institute of Information Management, and Director of the Software Incubator of the same University. He received his doctoral degree in Information Systems

from the Wharton School of the University of Pennsylvania and had taught at the University of Illinois, Purdue University, and Chinese University of Hong Kong. His primary research interests include electronic commerce, intelligent systems, decision support systems, knowledge management, and strategic applications of information systems. His papers have appeared in journals such as *Management Science*, *MIS Quarterly*, *Journal of MIS*, *Operations Research*, *Decision Support Systems*, and *Decision Sciences*. He also serves on the editorial board of *Decision Support Systems* and several academic journals.



**Deng-Neng Chen** is the Assistant Professor of Management Information Systems at National Pingtung University of Science and Technology in Taiwan. His research interests include knowledge management, electronic commerce and artificial intelligence applications. Dr. Chen received his PhD in Information Management from National Sun Yat-Sen University in Taiwan. His researches have been published in *Expert Systems With Applications*, *Industrial Management & Data Systems*, *Journal of Information Management* (in Chinese) and several conference proceedings.



**Yi-Cheng Ku** is an Assistant Professor in the Department of Computer Science and Information Management, Providence University, Taiwan. He received his PhD in Information Management from National Sun Yat-sen University. His research interests include recommendation systems, information system adoption and diffusion, and knowledge management. His research has been published in *Journal of Management Information Systems*, *Decision Support Systems*, *International Journal of Business*, *Electronic Commerce Studies*, and various conference proceedings.

*ness, Electronic Commerce Studies*, and various conference proceedings.



**Yong-Fang Yang** is a researcher at the Institute of Electronic Commerce of the Institute for Information Industry in Taiwan. He received his master degree in information management from National Sun Yat-Sen University in Taiwan.