From Experience to Expertise: Digesting Cumulative Information for Informational Web Search*

WEI-GUANG TENG*, WEI-HSUN WEN AND YUE-CEN LIU
Department of Engineering Science
National Cheng Kung University
Tainan, 701 Taiwan

Web searching has become a necessary activity for most Internet users. People can reach a specific website or collect desired information by using search engines. Nevertheless, when a user is unfamiliar with his or her search target, e.g., learning a previously unknown topic, browsing and understanding the web pages in the corresponding search result requires extra effort. In such an informational web search, the user accumulates information cues by reading web pages to approach the search target. For example, a user can gradually understand “cloud computing” after obtaining some information about relevant terms, e.g., distributed computing, Google App Engine, and Amazon EC2. In this work, we propose an “immediacy of interest” indicator, measured in terms of information freshness, to evaluate the possible user interest in a web page contained in a search result. Moreover, to address relevance in search sessions, the topic map is also utilized in our proposed scheme. Empirical studies show that informational web searching tasks can be completed efficiently and effectively with our approach.

Keywords: web search, browsing history, information cue, relevance feedback, visualization

1. INTRODUCTION

Web searching has become one of the tasks most frequently performed by users, as the information gathered on the Internet increases rapidly [10]. One typical use of search engines is acquiring relevant information to learn or understand previously unknown topics. In such an informational web search, the user is usually unfamiliar with his or her search target. Thus, accomplishing this search task may require multiple search sessions. Nevertheless, it is a paradox that search engines help filter irrelevant information in a single search session, while delivering more information in successive search sessions. As suggested in prior works [13], more advanced search engine designs can help users approach their search targets in increasingly efficient ways.

User behavior during an informational search task can be explored in the two scenarios below. First, unknown targets attract the users. For example, when one does not understand the term “cloud computing” in a result page, the user may initiate a new search session for this unfamiliar term. Second, the user cognition changes as information cues are accumulated. For example, the user understands the concept of “cloud computing” after browsing web pages describing relevant terms such as “distributed computing” and “utility computing.” Thus, originally unfamiliar terms become familiar ones. These two scenar-
ios help resolve typical user behavior patterns in conducting informational web searches.

Furthermore, to provide further insight into users approaching their search targets, we address relevance in successive search sessions. During an informational web search, the phenomenon of implicit relevance also occurs. Specifically, the implicit relevance indicates that two topics are seemingly irrelevant but share something in common and exhibit previously unknown relevance. For example, when mentioning the U.S. president Barack Obama and the NBA star LeBron James, the most common impression is that they are both African American. There is also the usually unknown fact that they are acquainted with each other and once played basketball together at a birthday celebration. A user may be interested in the details of such stories. However, this implicit relevance cannot be noticed by using conventional search engines. Instead, users usually have to spend time filtering irrelevant or uninteresting information from numerous result pages to realize the relationship between topics within their search scope.

To reflect the user’s cognition towards a specific concept, we propose an “immediacy of interest” indicator, hereafter “information freshness”. The value of this indicator shows the possible interestingness of a web page for a user at that moment. Therefore, a user can easily distinguish interesting pages, i.e., those with higher information freshness values, from the entries in the search result. This is similar to the concept of novelty [22], i.e., the degree to which the content of a result page is new to the user, which is discussed in prior works. In addition, we adopt a topic map to organize what a user has already browsed (or learned) during the search process. Each search session is represented by keywords as topics in the topic map. Relevant sessions in a specific user’s browsing history are identified and presented in the topic map. Thus, a user can switch between topics easily to keep his or her thoughts clear. The information freshness value and topic map evolve as one browses web pages and learns concepts during a search.

The rest of this paper is organized as follows. Relevant studies and improving techniques are reviewed in section 2. Our proposed scheme and relevant issues are explored in section 3. Empirical results, including case studies, are presented in section 4. Finally, this paper concludes with section 5.

2. PRELIMINARIES

User behavior in conducting information searches is generally explored in section 2.1. Current improvements for presenting web search results are reviewed in section 2.2.

2.1 User Behavior in a Web Search

Using a search engine is the predominant way to help users obtain information. During the search process, the interaction between a user and the search engine can be separated into three main stages [8]. In the first stage, query formulation, the user translates thoughts into words and submits these to the search engine. Then, in the second stage, relevance prediction, the user screens the search results to predict whether the desired information can be found from an entry, according to its title, snippet, and URL. Finally, in the third stage, relevance evaluation, the user evaluates the availability of selected pages from the search results. Once the user reaches a destination page, he or she may navigate
more pages by following hyperlinks on the page [12]. The user can then decide whether the desired information can be found in this stage. If not, he or she may start over from the first stage by refining or redefining the query terms.

The above process may happen within seconds or last for more than a few hours, depending on the user intent. Specifically, the taxonomy of a web search can be classified into three categories based on user intent, i.e., navigational, transactional and informational searches [4, 6, 11, 15]. The goal of a navigational search is reaching a specific website, such as Facebook, that the user already knows. The goal of a transactional search is performing some web-mediated activities, e.g., online shopping, downloading a file, or finding a map. Finally, the goal of an informational search is studying or discovering information in web pages to understand or learn some particular knowledge. In practice, the search process is generally very short for the first two categories, navigational and transactional, since their corresponding search targets are usually clear. Moreover, as reported in a previous study [23], the distribution of navigational and transactional searches is rarely affected by the time period of a day, but informational searches usually happen in the daytime.

In this work, we discuss informational web searches in which users start with a vague understanding of their goal. For example, a student who lacks background knowledge wants to know more about “the history of classical music.” After the first attempt, some famous musicians’ names appear in the search results. The next search session then starts with the query term “Beethoven.” After reading some pages, he wants to get an idea about “the musical style of Beethoven.” He realizes that “the Baroque period” (1600-1750) is a highly important period for classical music. Similarly, this student may issue successive queries until he is satisfied with the obtained information on the history of classical music. In summary, an informational web search may lead to several instances of query refinement or query re-formulation.

An informational model of berrypicking [2] best describes the actual behavior of information searchers, especially the exploratory search process. Search behavior is not merely a single movement through which users reach the best retrieved result, but it is, instead, a berrypicking or evolving pattern. As shown in Fig. 1, the path of a searcher is usually a curve instead of a straight line. Also, searchers may iteratively change the direction of their query after gaining new knowledge from retrieved documents and clarifying their own thoughts.

People tend to browse web pages according to information cues obtained during the search session. In the simplest form, these cues can be text or images around hyperlinks. For example, when a user sees the alternative text “ESPN.com provides comprehensive sports coverage” around a banner image, the user expects that this hyperlink can lead him to the official ESPN website, even though he has not yet clicked the hyperlink. Furthermore, a cue can also be knowledge perceived by the user after reading some relevant web pages. By accumulating sufficient cues, a searcher can move to the next step. This information foraging phenomenon is then analogous to how wild animals decide which resource to select, keep or give up [14]. However, such a search process requires some efforts, e.g., time consumption, uncertain search targets, frequent query changes, and irrelevant browsed pages.
2.2 Improvements for Presenting Search Results

When a user submits query terms to a search engine, a large number of search results may be returned. Thus, it is usually a time-consuming process to browse all these result pages. Prior works suggest several possible improvements [5, 7, 19-21], which are briefly categorized as follows.

(1) Re-organizing the Search Results
The conventional approach is using a long list to present search results. Alternatively, once document clustering techniques are applied, search results can be organized into groups. Within the same group, the topics of corresponding result pages are closely related. Consequently, users can browse only the search results contained in relevant groups.

Carrot2 [5] is representative of such a clustering search engine. Instead of performing the search task, Carrot2 redirects the query request and then gathers search results from various search engines, e.g., Google, Yahoo, and MSN. As shown in Fig. 2 (a), when an example query “Java” is submitted to Carrot2, several clusters are created, such as Sun Microsystems, Java Virtual Machine and so on.

Recent works have studied the process of re-ranking entries in the search result to achieve a personalized search. For example, because the well-known algorithm PageRank [3], adopted by the Google search engine, does not consider user intent, ClickRank [24] is proposed to estimate the importance of web pages based on an intentional surfer model. The main idea behind this personalized algorithm is that users search through targets according to their intents rather than searching based on the random surfer model. Alternatively, search result entries can be re-ranked according to the implicit relevance feedback [1]. Specifically, implicit relevance feedback indicates the interaction between the user and the search engine, such as accessing web pages, clicking ads, or reformulating queries. These activities are considered useful hints in improving search result ranking.
DIGESTING CUMULATIVE INFORMATION FOR INFORMATIONAL WEB SEARCH

(a) A clustering engine: Carrot2.

(b) Relevance mode in Pensieve.

Fig. 2. Examples of user interfaces.

(2) Visualization of Browsing History

When users act on the web, logs are stored on different online services or in local browsers. These browsing logs can help users perform searches in more efficient and effective ways. For example, the Re:Search Engine [19] refers to the browsing history to re-rank the search results to emphasize new and possibly interesting entries for a user.

Pensieve [7] is another work that utilizes the browsing history of a user and then presents it in an organized structure. Specifically, a user can review his or her browsing history in either timeline mode or relevance mode. The timeline mode emphasizes recent search sessions with finer granularities; the relevance mode, as shown in Fig. 2 (b), is used to represent the relationships between search sessions.

3. REALIZING AND UTILIZING USER COGNITION IN A WEB SEARCH

Terminology used in this work is defined in section 3.1. As a user may browse more than a few pages to obtain desired information, we illustrate the process of accumulating information cues in section 3.2. Accordingly, we propose an indicator to reflect the user cognition as described in section 3.3. Furthermore, we adopt the topic map in section 3.4 to present the relevance between search sessions. Finally, the phenomenon of implicit relevance is illustrated in section 3.5.

3.1 Terminology

In an informational web search, users demand to learn a concept, e.g., recognizing a particular person or event or acquiring desired information. As discussed in previous sections, an informational search may contain a number of search sessions, i.e., times submitting a query and browsing returned information. Each session contains some result pages and each page is called an entry. For each entry, there are representative keywords or terms regarded as features of the entry. In summary, these elements in an informational web search have the hierarchical relationship illustrated in Fig. 3.

During a search session, a user may browse several entries to obtain information. Information cues are gradually accumulated as one reads some relevant contents. Evaluating the usefulness of such information cues to determine their adequacy for a user’s comprehension of the target concept then becomes a crucial issue.
3.2 Cumulative Cues during an Informational Web Search

As reported in a previous work [11], more than 80% of the queries submitted to search engines can be categorized as informational web searches. Users often attempt to acquire concepts for an uncertain or unfamiliar target when conducting the web search.

The following example illustrates the common process of an informational web search. Suppose a student is asked “Who is the most powerful emperor in world history?” as the presentation topic in his history class. He may submit the query terms “the most powerful emperor in world history” to a search engine, e.g., Google, and then obtain some entries, as shown in Fig. 4.

![Fig. 4. A list of search results returned by the Google search engine.](image-url)
He may then examine every entry on the first few pages in the search results. In the first entry, he may be attracted by the famous name “George Bush” and feel curious to know why this former U.S. president is regarded as a powerful emperor. Consequently, he clicks the web page of the first entry and reads the contents. Afterwards, he learns that, to answer the question completely, more cues should be obtained by browsing more web pages. In other words, he may want to know about more emperors to compare their achievements. He may then notice the Roman Empire and its emperors in another entry. He may read the unfamiliar name “Gustavus Adolphus of Sweden” in yet other entries. Once he locates a good candidate for his presentation, he starts a new search with the query terms “Gustavus Adolphus” to know more about this emperor. Similarly, he may also start another search session after reading some stories about the Holy Roman Emperor “Charles V” to make further comparisons between these two emperors. This example clearly shows that a process of informational web searching may consist of more than a few successive and highly relevant search sessions. In general, information cues are accumulated as one reads more during the process. Nevertheless, more effort is required when one is unfamiliar with his or her search target.

3.3 Reflecting the Change of User Cognition

As observed from many examples, the user may understand originally unfamiliar terms after browsing some pages. This is intuitive: user cognition changes as one learns or gets familiar with something new. In addition, one may try to filter redundant entries contained in the current search results because these entries are already known, not new, or irrelevant. We thus propose to both identify these behavior patterns as changes in user cognition and develop an indicator to concretely reflect this phenomenon. The change in user cognition is not addressed in conventional search engines, resulting in a time-consuming process for users to identify and read something helpful in a large amount of search results.

In this work, we seek to develop an indicator to represent the interestingness of each entry in the search results. With this indicator, one can simply filter the entries which are already known, i.e., not so fresh. This indicator can also be smoothly integrated with conventional search engines. Because the user cognition evolves with time, the value of this indicator for a single page may also change with time. Then, the crucial issue is representing the change of user cognition through the accumulation of information cues. This indicator, referred to as immediacy of interest or information freshness, can be best understood as the interest a user has in a result page at the present time. Specifically, an entry with a higher value indicates that the corresponding page contains more undiscovered information within this session and the user is more likely to read it. This helps the user to master the concept he or she reads. Consequently, the value of this indicator for a specific page is based on the quantity of unfamiliar content included in the page. In other words, if a page contains many identical terms and similar paragraphs as what the user regularly reads, then this page is unlikely to have high information freshness. Not that a user typically needs to carefully examine each entry in search results when there is no such indicator. Nevertheless, as information freshness is introduced and visually represented, a user can glance over the whole list and read entries with higher values.
3.4 Identifying the Relevance between Search Sessions

A user may require many sessions to complete an informational web search. Various topics are included in a single search session and the same topics may be shown in relevant sessions. Moreover, a user may still need to filter undesired information from search results. It is thus crucial to identify the relevance between search sessions. Consequently, the topic map is adopted to organize all sessions and emphasize the influence of the same topics in relevant sessions. The corresponding details are illustrated as follows.

In the previous example of discovering the most powerful emperor in world history, the user browses various topics including “George Bush,” “Gustavus Adolphus,” “the history of Sweden,” and “Charles V.” These topics are successively explored and can be represented on a timeline as shown in Fig. 5. Furthermore, one can easily understand that the two topics “Gustavus Adolphus” and “the history of Sweden” have higher relevance after learning the corresponding stories. On the other hand, the topic “Charles V” seems to have little relevance to either “Gustavus Adolphus” or “the history of Sweden.” However, it is difficult to appropriately represent the relevance on a temporal diagram.

![Figure 5](image1.png)  Fig. 5. An example of successively explored topics.

![Figure 6](image2.png)  Fig. 6. A topic map showing the relationship and the relative significance of browsed sessions.

Because of this difficulty, we propose the adoption of the topic map to represent the relevance between pairs of all topics. The topics in Fig. 5 can then be organized to form a topic map, as shown in Fig. 6. A topic is represented as a rectangle (or a node) and the relevance between two topics (if any) is represented as a link. Consequently, one can easily recognize relationships between topics that he or she has ever browsed.

The size (and the color shade) of a session in our topic map is proportional to its freshness. A larger rectangle in a more vivid color indicates a topic of higher freshness. Specifically, the freshness of a session is calculated by averaging the information freshness of its corresponding entries. The thickness of a link is proportional to the relevance strength. A thicker link indicates that two corresponding sessions are more closely relevant.

Moreover, the topic map in our proposed scheme evolves as the user reads more pages or starts more search sessions. In the example, as shown in Fig. 6, the relevance between Gustavus Adolphus and Charles V is established as the user tries to compare their achievements by reading pages containing both names. The relevance between informational search sessions can generally be properly addressed with the proposed topic map.

3.5 Phenomenon of Implicit Relevance

In this work, a phenomenon called implicit relevance in informational web searches is extensively discussed. Specifically, implicit relevance indicates that two seemingly irrelevant topics share something and exhibit previously unknown relevance. This implicit
relevance may greatly interest a user who wants to know more details to connect learned information. The following example illustrates the usefulness of identifying implicit relevance.

Suppose a student wants to learn about the history of aircraft. He then submits “history of aircraft” to a search engine and reads some returned entries. Soon he is interested in an entry discussing the Wright brothers and realizes that they invented airplanes. At the same time, he notices the familiar name “da Vinci” in these entries. He then starts the next search with the query “da Vinci” to confirm the relevance of the famous Italian artist Leonardo da Vinci to the history of aircraft. As one may expect, the ensuing search results contain all art-oriented entries. This student may thus become confused as to whether this relevance exists. In fact, Leonardo da Vinci was also an outstanding engineer, regarded as a predecessor of the Wright brothers in designing a flying machine. The volume of this information, however, is far less than Leonardo da Vinci’s artistic achievements.

Note that the implicit relevance tends be personal. For a person who is not interested in the history of aircraft at all but is a religionist, Leonardo da Vinci is mainly a painter who created *The Last Supper*, the most reproduced religious painting. In view of this, to help a user utilize information cues well, we propose providing suggested web pages once this implicit relevance occurs. If confirmed by the user, i.e., by reading several suggested web pages, the relevance between topics becomes explicit and is established automatically in the proposed topic map.

4. EMPIRICAL STUDIES

To realize our concepts of information freshness and a topic map for supporting informational web search, we propose and implement a prototype of an appropriate scheme in section 4.1. Further experimental studies based on our prototype system compare the feasibility of using our proposed scheme to that of a conventional search engine. The corresponding results are presented in section 4.2.

4.1 Proposed Scheme and the Prototyping

To transform our proposed concepts into practical applications, we devised a system scheme consisting of several key components: an interactive interface, data collection, part-of-speech tagging (POS tagging), session characterization, relevance evaluation, suggestion generation, and result visualization, as shown in Fig. 7. Moreover, a prototype of our scheme is implemented using the PHP script language and an ActionScript library. Specifically, major system components are briefly described as follows.

- Data collection: This step forwards the query terms provided by a user to the Google AJAX Search API [9] to obtain corresponding search results. The results can be from web pages, blogs, or news.
- POS tagging (part-of-speech tagging): This step receives the search results and extracts the nouns from each page to filter the redundant information, such as stop words and punctuation marks. Without loss of generality, we adopt the Stanford Part-Of-Speech Tagger [18] for tagging English articles and the SCWS (simple Chinese words segmentation) [17] for segmenting and tagging Chinese articles.
• Session characterization: This step produces the terms which are features of the session from a set of entries already processed by the POS tagging. We adapt the TF-IDF (term frequency-inverse document frequency) technique [16] to evaluate the significance of all terms. Specifically, the original TF-IDF equation is shown in Eq. (1).

$$W_{ij} = tf_{ij} \times \log \frac{N}{df_i},$$  \hspace{1cm} (1)

where $W_{ij}$ is the weight of term $i$ in document $j$, $tf_{ij}$ is the occurrence frequency of term $i$ in document $j$, $N$ is the number of documents in the corpus, and $df_i$ is the number of documents where term $i$ occurs at least once. In our environment, every session is considered as a different document. For fast processing, snippet texts instead of full texts of all the entries in a session are gathered. Furthermore, to emphasize the familiar terms that the user has read before, we slightly revise Eq. (1) to be as shown in Eq. (2).

$$W_{ij} = tf_{ij} \times \log \frac{N}{df_i} \times (C_i + 1),$$  \hspace{1cm} (2)

where $C_i$ is the occurrence frequency of term $i$ in the user’s browsing history. By taking $(C_i + 1)$ as a weight, the significance $W_{ij}$ of term $i$ in session $j$ can be obtained. For each session, the top $k$ terms are selected for the term list, regarded as representative of the session, and saved in the database for further usage.

• Relevance evaluation: After the term list of each session is generated, the relevance of all session pairs can be evaluated. Specifically, the relevance strength of two sessions is determined according to Eq. (3).

$$R(i, j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|},$$  \hspace{1cm} (3)
where $R(i, j)$ indicates the relevance strength between sessions $i$ and $j$, and $S_i$ and $S_j$ are the term lists in sessions $i$ and $j$, respectively. $R(i, j)$ ranges from zero to one. In our prototype system, two sessions are considered as relevant when the value of $R(i, j)$ is no less than a predefined threshold, e.g., 0.3. Furthermore, to address the phenomenon of implicit relevance as described in section 3.5, a simple implementation is to define another lower threshold, e.g., 0.2. Therefore, two sessions $i$ and $j$ are implicitly relevant if the value of $R(i, j)$ falls between these two thresholds, e.g., $R(i, j) \in [0.2, 0.3)$.

- Suggestion generation: Should there be implicit relevance between the current session and a previous one, this step generates suggested entries by exploiting information in the current session and the user’s browsing history. Intuitively, suggested entries can be obtained by submitting a new user query (which combines query terms in both sessions) to the Google AJAX Search API. If more than a few suggested entries are read by the user, the term lists generated according to Eq. (2) could change and thus the relevance strength may increase.

- Result visualization: This step visualizes the two proposed items, information freshness and a topic map, in an interactive web interface. Specifically, the information freshness (or immediacy of interest) of an entry in the search result is evaluated by Eq. (4).

$$F_i = \frac{|T_i - B|}{\sum_{t=1}^{n} \log_{a+c_i} (a + c_t)}$$

where $F_i$ is the information freshness of entry $i$ ranging from zero to one, $T_i$ is the set of terms in entry $i$, $B$ is the set of terms in the user’s browsing history, $a$ is a constant, $C_i$ is the occurrence frequency of term $t$ in user’s browsing history, and there are $n$ terms in entry $i$. Thus, the numerator denotes the number of terms that the user has never seen before; the denominator expresses the set size of all terms in entry $i$ weighted by the term familiarity, $C_t$. The denominator reduces to the number of terms $n$ in entry $i$ if the user has not read any of the terms.

Nevertheless, one may understand that as the user browses entries to read more terms, an unread entry containing several identical terms becomes less fresh. On the other hand, if an entry is almost composed of unseen terms, it is highly interesting, or very fresh, to the user. The freshness of a session, represented in colors and sizes in the topic map, is obtained by averaging the information freshness of all entries in the session. In addition, relevance strength between two sessions as calculated in Eq. (3) is represented by the link thickness.

According to the system scheme above, the user interface of a prototype system is shown in Fig. 8. A topic map is shown over all the entries, and a visualized indicator showing the information freshness value is shown to the right of each entry. A user can shift between topics by simply clicking on the corresponding node in the topic map. The corresponding search results are presented accordingly. The concept behind such an interactive interface is the transformation and visual representation of the information cues accumulated during the informational web search. The user can then focus on moving forward in the search or recalling topics explored in his or her browsing history.

The following example based on the information in Fig. 8 illustrates this process.
Suppose a user is interested in international conflicts in Asia. Two events may be noticed: the missile conflict between Taiwan and China and the sinking of Chenoan. As he or she reads a few pages on the topics of “South Korean Navy,” “North Korean Navy,” and “Chenoan,” the topic map shows this familiarity in terms of node sizes and degrees of color lightness. This user then proceeds with his or her search, reading more pages about “Taiwan missile defense” because the corresponding node is larger and brightly colored.

4.2 Experimental Results

We devised an experiment to evaluate the feasibility of using our proposed scheme for supporting informational web search. We asked 25 volunteer participants to answer 7 open questions. They were randomly asked to use either our prototype system or a conventional search engine, e.g., Google, to obtain the required information before answering these questions. These questions and reference answers are detailed below.

• Question 1: Specify the differences between a co-constructed station and a union station. (Answer: Several mass transit systems that share an identical station structure are called co-constructed station; the systems having their own stations that are connected with each other are called union stations.)
• Question 2: Specify the differences of color tones for Nikon and Canon DSLRs (digital single-lens reflex cameras).
  (Answer: Nikon DSLRs have colder color tones, and Canon DSLRs have warmer color tones.)
• Question 3: Specify the relationships between the following four names, Qi Bai Shi, Xu Bei Hong, Johann Ambrosius Bach, and Johann Christoph Bach.
  (Answer: The former two are friends and both are artists. The latter two are siblings and both are musicians. There is no relationship between the two pairs.)
• Question 4: List four union stations.
  (Answer: They could be Kansas, St. Louis, Chicago, and Washington, D.C., etc.)
• Question 5: What is independence movement? List three countries that have successfully achieved in independence movement.
  (Answer: Independence movement is some movement whose goals are primarily to create national identity and to be recognized as a sovereign state. Countries that have successfully achieved in independence movement include India, the U.S., Ireland, etc.)
• Question 6: List two ongoing military conflicts and corresponding belligerents.
  (Answer: They could be Mexican Drug War and War in Iraq, etc.)
• Question 7: List four government organizations of the U.S. in Washington, D.C.
  (Answer: They could be the White House Office, Federal Bureau of Investigation, United States Capitol, etc.)

Fig. 9. The average time required by the participants.
Fig. 10. The average number of web pages visited by the participants.

The required time and the number of web pages visited by the participants are shown in Figs. 9 and 10, respectively. Fig. 9 clearly shows that our approach is generally more efficient in completing a question. Moreover, as shown in Fig. 10, participants using our approach can obtain correct answers by reading fewer web pages. A participant can only read snippets to conjecture his or her next step when using a conventional search engine, which may result in reading redundant pages containing nothing helpful when constructing answers. Also, the required time for digesting both useful and redundant information tends to be longer with conventional search engines. On the other hand, by utilizing our approach with the implemented prototype, both the required time and the number of visited web pages can be significantly reduced.

Question 3 is a representative example to illustrate the feasibility of our approach, because the two pairs of names can be seen in the topic map in our approach. The partici-
pants then must only read a simple profile of each name to answer completely. In contrast, our approach does not outperform a conventional search engine in Question 6. By observing the browsing logs of all the participants, the main reason is that some well-organized web pages on the corresponding topic can be found at Wikipedia, a well-known online encyclopedia. Therefore, participants do not need the indication of either the topic map or the freshness indicator as proposed in our scheme.

5. CONCLUSIONS

In this work, we have generally explored problems encountered in an informational web search task. We have thus proposed to utilize and evaluate information cues obtained during the search process. Specifically, we have defined an indicator to reflect the user interest in a specific web page, and we have utilized the topic map to represent the relevance between search sessions. Empirical studies have shown that our approach can help users conduct informational search tasks efficiently and effectively.

REFERENCES

11. B. J. Jansen, D. L. Booth, and A. Spink, “Determining the informational, navigational,


**Wei-Guang Teng** received B.S. and Ph.D. degrees from the Department of Electrical Engineering at the National Taiwan University, Taipei, Taiwan in 1998 and 2004, respectively. He is currently an Assistant Professor in the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan. He was a senior researcher at Groundhog Technologies from 2004 to 2005. His research interests include data mining, multimedia networking, and mobile computing.
Wei-Hsun Wen (文偉勳) received a B.S. degree from the Department of Computer Science, National Taipei University of Education, Taipei, Taiwan in 2009, and an M.S. degree from the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan in 2011. His research interests include web searching and location-based services.

Yue-Cen Liu (柳岳岑) received a B.S. degree from the Department of Computer Science, Hsuan Chuang University, Hsinchu, Taiwan in 2007, and an M.S. degree from the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan in 2010. His research interests include web searching and human-computer interaction.