ELSEVIER

Automation in Construction 14 (2005) 45–58

AUTOMATION IN CONSTRUCTION

www.elsevier.com/locate/autcon

Learning search keywords for construction procurement

Ren-Jye Dzeng*, Shih-Yu Chang

Department of Civil Engineering, National Chiao-Tung University, 1001 Ta-Hsieu Road, Hsinchu, Taiwan 30050, R.O.C.

Accepted 25 June 2004

Abstract

Seeking information from websites has become an essential part of a contractor's procurement undertaking, as more and more procurement websites become available on the Internet. Websites host extremely large amounts of information; a keyword search, therefore, is often more efficient than browsing via an index. However, in order to find the desired information, it may be necessary to enter keywords using a trial-and-error process. This research recognizes that professional procurement experience can help users search website information more effectively, by using fewer keywords, and so proposes a learning model and suggestion model that can capture such experience, thus guiding inexperienced users in their search. Experiments, evaluating the performance of the system, were also conducted.

Keywords: Procurement; Information search; Machine learning; E-commerce

1. Introduction

Since the proliferation of Internet users in the second half of the 1990s, almost all of the document management systems of electronic projects have migrated to using the Internet as their physical network; project-specific Web servers are used as the storage medium and Web browsers as the main platform for building buyer interfaces [4]. Several e-markets, specifically for construction, have also been established, including AEC Info [1], BuildPoint [5], bLiquid.com [3], Citadon, [9], ProcureZone [15], and PrimeContract [14]. The proliferation of Web-based project management platforms and e-markets has provided contractors with more business opportunities and a wider selection of suppliers; at the same time, it has created a challenge for contractors in managing the flood of electronic information.

Commercial construction procurement websites have attempted to provide all relevant procurement information on a single website, to attract buyers with one-stop shopping and creative business opportunities. However, most websites provide only two primary ways of searching for information, namely

* Corresponding author. Tel.: +886 3 5733529; fax: +886 3 5745074.

E-mail address: rjdzeng@mail.nctu.edu.tw (R.-J. Dzeng).

0926-5805/$ - see front matter
doi:10.1016/j.autcon.2004.06.003

by index/menu or by keyword. The keyword search method is probably most often preferred by users not familiar with the content or indexing scheme of a website. A keyword search may also be more efficient, when there are either too many indexes to deal with, the website hosts an enormous amount of information, or a buyer is not aware of the specific procurement terms used by the website. Nevertheless, it may still be necessary for most buyers to input keywords using a trial-and-error procedure, in order to narrow down the search to find the desired information.

Instead of relying on the primitive search engines found in most procurement websites, a search guide system could help a user's keyword search by reducing the number of keywords required to find the desired information. Qiu and Frei [18] assumed that each keyword could be expanded to include related keywords, just as "exercise" is related to "basketball" and "track and field". Analysis of these keywords can help clarify the search target and increase user success in attaining the desired search results. Balabanovic and Shoham [2] proposed a scheme to calculate the "interestingness value" of each word, based on how frequently it occurs in a document, and to extract the most likely words from each document, to incorporate into a representation of user interest. Several concept-based query models for general-purpose search have also been developed for deducting information-seeking behavior (e.g., Ref. [20]) or for interactive information retrieval (e.g., Ref. [12,17]).

Kasbah [7] is a prototype of a Web-based marketplace, where buyers are able to create autonomous agents that buy and sell goods on their behalf. These agents help with the major steps of buying and selling by locating prospective buyers. Cheung et al. [8] developed intelligent agents to assist a user in locating relevant documents by uncovering access patterns (a series of keyword inputs) from user access logs. Carlyle [6] proposed a clustering method that divided documents into groups, and based on the keywords and attributes of the documents read by the users, determined their interests.

Finding information related to construction procurement is a more complex process than for consumer goods or general-purpose documents, which was the primary focus of the aforementioned research works. The higher complexity and scale of construction projects, emerging construction materials and technologies, and changes in building codes and regulations make the preparation of accurate tendering specifications a challenge for contractors. For example, different specifications are necessary for different types of construction materials: to order steel, accepted industry standard specifications may be required (e.g., the American Society for Testing and Materials (ASTM)), for the grade, diameter, strength and weight of the steel; to order ceiling boards for a building project, it may be necessary to specify length, width, thickness, material type, fireproof certification, texture and color. A predetermined or fixed search guide, therefore, seems impractical.

Our research recognized that professional procurement experience helped users more effectively carry out website information searches, by using fewer keywords. We planned to capture such experience in order to guide inexperienced users in their search. The proposed learning network required experienced users to annotate the relationships among the series of keywords entered during their web search, and then corrected the user's keyword input or prompted subsequent keyword candidates in order to help inexperienced users reduce the number of keywords required during their search.

Twelve professionals, using 14 procurement packages, with 64 items in total, evaluated the proposed framework. It was demonstrated that the proposed learning keyword guide facilitated a dynamic, customized menu and indexing system, and reduced the number of keywords required for the professionals to find the information they desired.

2. Guiding the search

The goal of this research was to improve search effectiveness by guiding the user's search using three approaches; namely correction, specification and extension. The correction guide corrects misspelled or misused keywords. The specification guide constrains the search space by adding more "AND" words to a keyword phrase or by replacing the keyword with a more specific term. For example, a keyword string which includes “400×400 13 mm 21 mm” and “H-type steel” results in more specific search results than “H-type steel” alone. The extension guide extends the search space by adding more “OR” keywords, by suggesting keywords the user may need for subsequent searches; for example, when purchasing curtain walls, detailed specifications are required because several patented technologies are available. Thus, while searching for suppliers of curtain walls, the user may also want to include specifications and performance evaluations for curtain walls.

Based on these three approaches, this research applied the following guides: *correction*; *specification-by-equivalence*; *specification-by-detail*; *extension-by-time*; *extension-by-location*; *extension-by-team*; and *extension-by-component*. These guides are abbreviated as *correction*, *S-equivalence*, *S-detail*, *E-time*, *E-location*, *E-team* and *E-component*, respectively.

The *correction* guide suggests the proper words for misspelled keywords, suggesting, for example, “diaphragm wall” for “diahpagm wall”. The *S-equivalence* guide suggests equivalent words used by the website for the keywords entered and may suggest: “diaphragm wall” for “slurry wall”; “rebar” for “reinforcing bar”; or “premixed concrete” for “ready-mixed concrete” and “ready-mix concrete”. The *S-detail* guide suggests more specific words for the input keywords. For example, the guide may suggest “hot coal-tar waterproof felt” for “felt”. The *E-time* guide suggests keywords for procurement items required by the procedures that normally follow the activity to which the input keyword relates. For example, the guide may suggest “steel rebar labor” for “3500psi type II Portland cement concrete” because construction activities requiring these two items are next in sequence for typical reinforced concrete building projects. The *E-location* guide suggests keywords for procured items whose construction normally occurs adjacent to the item to which the input keyword relates. For example, the guide may suggest “reverse circular concrete pile” for “diaphragm wall” because the construction activities requiring these items often occur at adjacent locations. The *E-team* guide suggests keywords for procured labor or other resources relating to the type of labor the entered keyword represents. For example, the guide may suggest “high-rise welding labor” for “steel erection licensed labor” because the two types of labor are often utilized simultaneously, as a team, in building projects. The *E-component* guide suggests keywords of procured items that are normally embedded within the item to which the entered keyword relates. For example, the guide may suggest “#4 deformed rebar” for “type C reinforced concrete pipe” because the former item is often embedded in the latter.

Procurement websites can be divided into two categories based on the type of buyers participating, i.e., closed market and open market. Only a single contractor or an alliance of contractors can exist as a primary single buyer in a closed market, while any qualified contractors may participate as buyers in an open market.

For procurement in a closed e-market, the buyer is usually familiar with the procurement system. Therefore, the primary guide objective is to extend the search; i.e., to predict any additional information the buyer may need according to initial keywords entered. Correction of misspelled keywords is also useful. For procurement in an open market, the buyer may not be familiar with the procurement website, and the search result may often be different from what is anticipated. In this case, the specification guide and correction of equivalent keywords are especially useful, as are both the correction of misspelled keywords and the extension guide.

3. User differentiation

Depending on both their professional procurement and web-search experiences, construction procurement engineers may differ in their search behaviors and require different types of search guides. As shown in Fig. 1, this research divided users into four types: Type I—users with procurement experience and familiarity with the website; Type II—users having procurement experience but who were unfamiliar with the website; Type III—users with little procurement experience but familiarity with the website; and Type IV—users with little procurement experience who were also unfamiliar with the website.

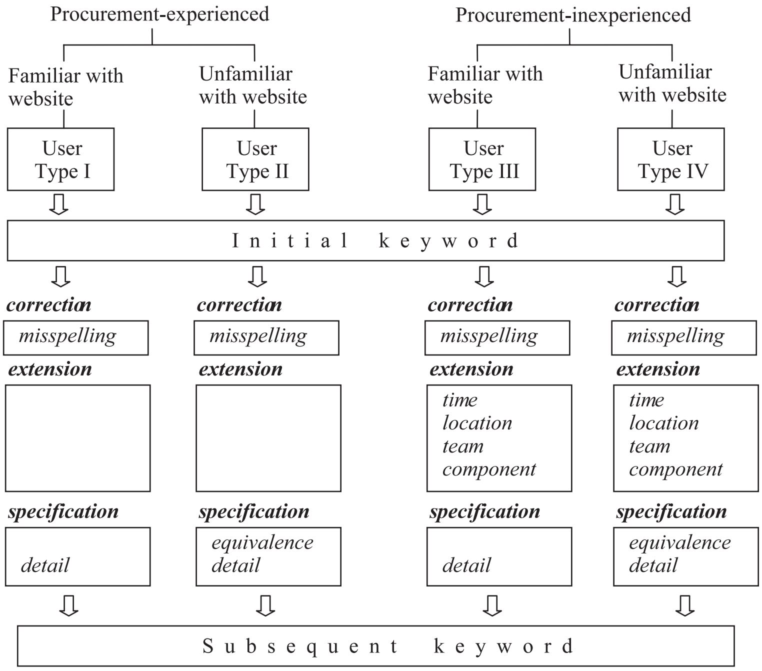

Fig. 1. Primary guides for different types of users.

Fig. 1 also emphasizes which primary guides are most useful to each type of user; other guides, not mentioned here, may also be useful. The guides assist the users' search by predicting possible subsequent keywords based on the initial keyword entered. The *correction* guide is useful for all types of users, because everyone is liable to misspell words now and again. The *extension* guides are most useful for user Types III and IV, as they have less procurement experience and may not know all of the available items required to fulfill a procurement package. The *S-equivalence* and *S-detail* guides are most helpful for user Types II and IV, as they are unfamiliar with the website and may well know nothing about the terms used by the website's database. Nevertheless, even though user Types I and III have adequate knowledge of the website and need not use the *S-equivalence* guide, they still may obtain some help from the *S-detail* guide, as few people can memorize the details of all the specification for each type of construction material; in this aspect, the guide quickly helps to narrow down the search.

This research study proposes a learning network model that can capture the search knowledge of procurement- and website-experienced users to guide inexperienced users in their search.

4. Learning keyword search

4.1. Search scenario

Let us consider a search scenario as illustrated in Fig. 2. A user who is an engineer is seeking supplier information for a construction project that includes several procurement tendering packages. Each package includes several procurement items. The user first enters the keyword K_{11} to search for information about the first item K_1 of the first package K. The result shows no URL (Universal Resource Locator) reference because the user misspelled a word. The user corrects the mistake by inputting another keyword, K_{12}. There is still no result for URL references. Then user tries another equivalent term, K_{13}. This time, however, the result contains too many URL references. Thus, the user begins the search again, by inputting a more specific keyword, K_{14}, and finally obtains satisfactory references. The process, from

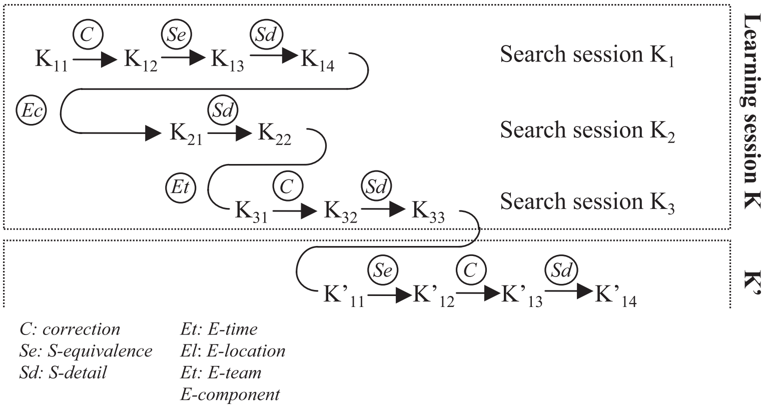

Fig. 2. Search sessions and learning sessions.

entering K_{11} to the final entry of K_{14} is the search for a specific item, and is termed a search session.

In order to search for another item in the same package, the user continues the search by inputting another keyword K_{21}. The result, in this case, contains too many URL references. Thus, the user redoes the search by inputting a more specific keyword K_{22}, which locates satisfactory references. This process repeats until the user inputs keyword K_{33} and finishes the search for all three items in the first package. To search for information related to another package, the user continues the search by inputting K'_{11}. The process continues, in a similar fashion, until the user finishes searching for information required for all the tendering packages.

The process from K_{11} to K_{33} represents a series of keywords input to search for tendering package information and is termed a learning session because the keywords used in this session are related. With appropriate annotation by experienced users, the relationships between these keywords can be learned in order to support the corresponding proposed guides. In some situations, the user may want to define a single learning session as covering the search keywords for several related tendering packages, if the items belonging to these packages are considered to be related.

4.2. Annotation links

At the end of each learning session, the experienced user teaches the guide system by annotating appropriate links for each pair of consecutive keywords entered during the search. The available links correspond to the aforementioned guides and include *correction*, *S-equivalence*, *S-detail*, *E-time*, *E-location*, *E-team* and *E-component*.

Annotation links can only be added to two consecutive keywords within the same learning session. A learning session may include several search session. Within each search session, no *extension* link should be annotated. (Extension links should only be annotated between two keywords in different search sessions, in order to target different procured items.) Other links can only be annotated between keywords within a search session.

Fig. 2 shows the links that may be annotated for the scenario previously introduced. The links are denoted by circled, abbreviated link names above each arrow between the keywords. *Correction* is annotated between K_{11} and K_{12} because K_{12} is the corrected phrase for K_{11}. *S-equivalence* is annotated between K_{12} and K_{13} because K_{13} is the correct website term that is equivalent to K_{12}. *S-detail* is annotated between K_{13} and K_{14} because K_{14} details and narrows the search space of K_{13}. *E-component* is annotated between K_{14} and K_{21} because they search for information of different procurement items where one item is a component of the other. Links are annotated in a similar fashion for search sessions K_2 and K_3.

4.3. Learning model

In a learning session, an experienced user annotates the relationship between two consecutive keywords via the links. Each pair of linked keywords constitutes a keyword search pattern. Each pattern includes the preceding keyword, the succeeding keyword, the type of link connecting the two keywords and other log information such as number of recording times and the time from the last recording of the pattern.

Fig. 3 illustrates links that are conceptually possible between the two keywords K_i and K_j after a period of learning. It contains four keyword search patterns: (K_i, L_1, K_j), (K_i, L_2, K_j), (K_i, L_3, K_j) and (K_j, L_1, K_i). It is possible that two keywords may have a link recorded multiple times, as does L_1 from K_i to K_j, which was recorded twice by a Type I user on Oct. 1, 2003, and three times, also by a Type I user, on Dec. 23, 2003. Two keywords may also have multiple links of different types, like L_1, L_2 and L_3, which were annotated from K_i to K_j. They may also have bidirectional links of the same type, as where two L_1 links are pointing to K_j and one L_1 link is pointing to K_i.

Fig. 4 shows an example of a linked keyword network. As can be seen, "premixed concrete" and "#6 deformed annealing rebar" have bidirectional *E-location*, *E-time* and *E-component* links. That is, premixed concrete and rebar are often used at locations adjacent to one another. Premixed concrete may come to mind when searching for rebar. Rebar can be embedded in concrete as a reinforcing component, so premixed concrete has a *component* link to rebar. Pouring concrete is a typical activity after rebar has been assembled, thus rebar has a time link to premixed concrete.

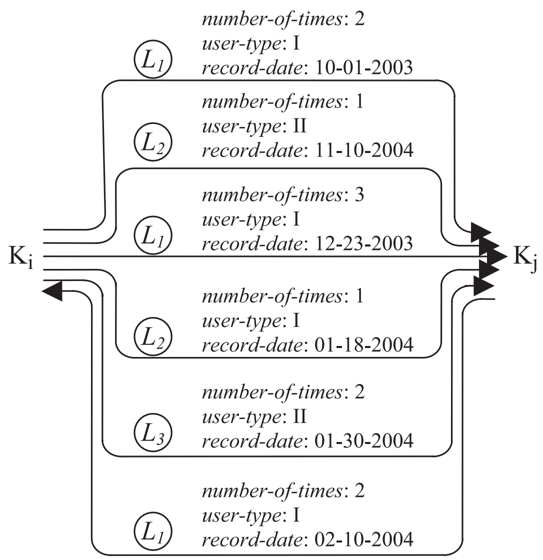

Fig. 3. Conceptual illustration of search patterns that contain links between two keywords.

The learned outbound links from a particular keyword point to other keywords the user may need for subsequent searches when the user inputs that keyword. Therefore, when a user inputs a keyword, the guide system considers all succeeding keywords within the search patterns whose preceding keywords match that keyword, to be candidates for subsequent keyword suggestions. To limit the number of suggested keywords and present the most useful keywords first, a suggestion model incorporating the mechanisms of ranking, indexing and setting a threshold of keyword candidates is required.

4.4. Suggestion model

This section describes the suggestion mechanism in terms of ranking, indexing and threshold. Inspired by the concept of Google's PageRank [10] for ranking Web pages, the ranked value of each possible keyword candidate is calculated by considering the number of links emanating from the keyword input, the number of links pointing to the keyword candidate and the experience level of the user who has annotated the link between the keyword input and the keyword candidate. Given the keyword input as K_i, the rank value of K_j from the perspective of the *link* type is calculated via Eq. (1).

$$
\begin{aligned}
R(K_i, link, K_j, user, \Delta t) &= w_1\left(S_{\text{pattern}}\left(K_i, link, K_j, user, \Delta t\right)\right) \\
&\quad + w_2\left(S_{\text{keyword}}\left(link, K_j\right)\right) \\
&= w_1\left(p\sum_{(K_i,\text{link},K_j)}(user) + q(1/\min\Delta t)\right) \\
&\quad + w_2\left(\frac{Q_{\text{link}}\left(link, K_j\right)}{Q_{\text{link}}(link)}\right) \qquad (1)
\end{aligned}
$$

where user: the experience level of the user (the default setting is 1 for Type I, 0.7 for Type II, 0.4 for Type III, and 0.1 for Type IV user); Δt: (the current

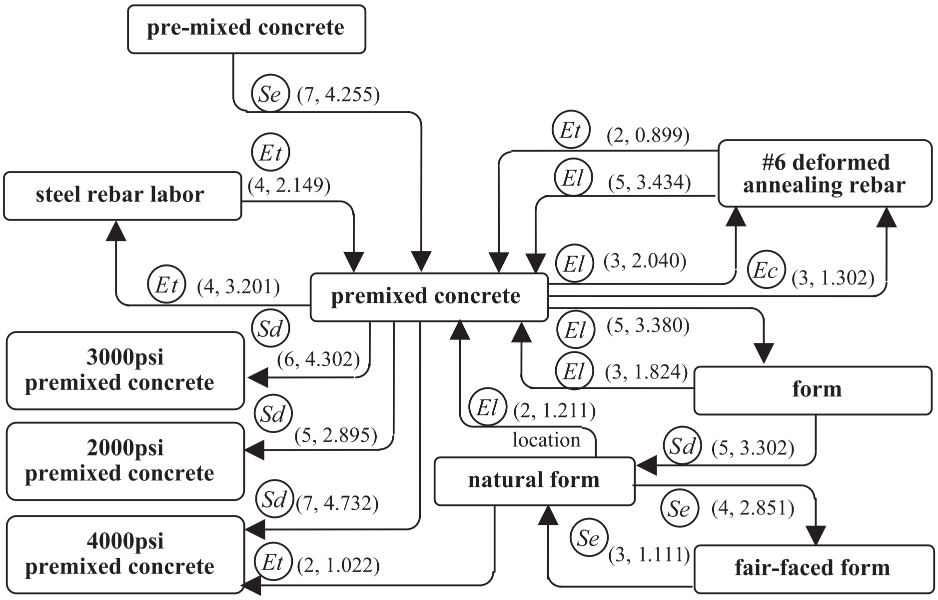

Fig. 4. Example: keyword network.

date)−(the last date when the link was recorded), w_1, w_2: weights determined by the system manager, $w_1+w_2=1$, $0\le w_1$, $w_2\le 1$; $S_{\text{pattern}}()$: score for the importance of the pattern (K_i, link, K_j); $S_{\text{keyword}}()$: score for the importance of K_j; p, q: weights determined by the system manager, $p+q=1$, $0\le p$, $q\le 1$; $Q_{\text{link}}(\text{link}, K_j)$: the number of links of the link type entering K_j; $Q_{\text{link}}(\text{link})$: the total number of links of the link type in the keyword network; Eq. (1) shows that the rank value is a weighted average of two parts, namely S_{pattern} and S_{keyword}. S_{pattern} concerns about the importance of the pattern and how recently the pattern was last learned. In Fig. 4, the left number in parentheses above each link represents the number of times the link has been learned, and the right number represents the S_{pattern} for the search pattern. Suppose that the user has just entered the keyword “premixed concrete” and needs help on the subsequent *location*-related keywords. Only two search patterns found in Fig. 4 start with “premixed concrete” and also include an *E-location* link. These two patterns are (“premixed concrete, *E-location*, “form”) and (“premixed concrete, *E-location*, “#6 deformed annealing rebar”), whose log information is detailed in Fig. 5, which shows the number of times the link has been learned, the type of user who has annotated the link, and the date recorded for each of these eight links (noting that the bottom-left link has been learned twice).

Assuming that p=0.7, q=0.3, 1.0 for Type I user and 0.7 for Type II user, and the current date is Feb. 10, 2004, S_{pattern} equals $p\times\sum\text{user}+q\times(1/\min\Delta t)=0.7\times(1+0.7+1)+0.3\times(1/2)=2.04$ for (“premixed concrete, *E-location*, “#6 deformed annealing rebar”), and $0.7\times(1+0.7+0.7+1\times 2)+0.3\times(1/1)=3.380$ for (“premixed concrete, *E-location*, “form”).

S_{keyword} represents the importance of the keyword candidate K_j from the perspective of the link type of interest. Considering all links of that type, the more links there are to the keyword, the more likely it is the subsequent keyword the user needs. Given the input keyword “premixed concrete”, the network in Fig. 4 comprises 18 (=3+5+3+5+2) *E-location* links, of which three go to “#6 deformed annealing rebar” and five go to “form”. Thus, “form” (S_{keyword}=5/18) is more likely than “#6 deformed annealing rebar” (S_{keyword}=3/18) to be next location-related keyword the user requires.

Assuming w_1=0.7, and w_2=0.3, R is $2.04\times 0.7+(3/18)\times 0.3=1.478$ for “#6 deformed rebar”, and $3.380\times 0.7+(5/18)\times 0.3=2.449$ for “form”. Therefore, “form” will be ranked higher than “#6 deformed rebar” in the suggested list of location-related subsequent keywords with respect to “premixed concrete”.

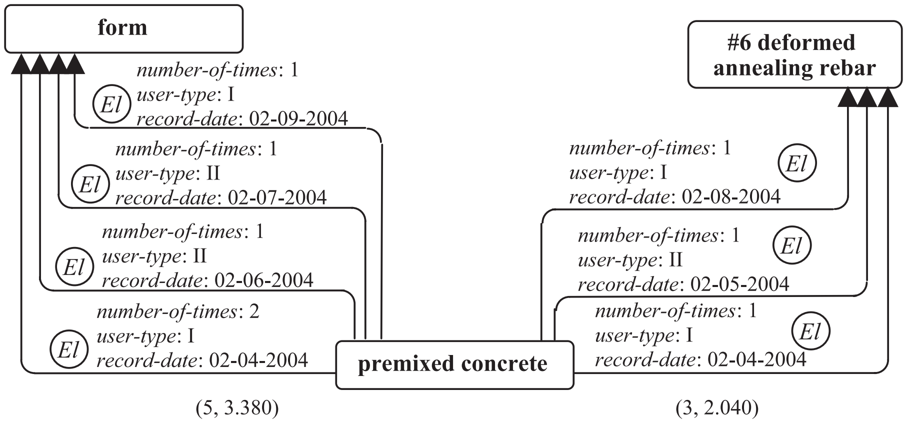

Fig. 5. Example: link records for two search patterns.

Next, keyword candidates are indexed based on the types of the links pointing to them. These indexes include *correction*, *S-equivalence*, *S-detail*, *E-time*, *E-location*, *E-team* and *E-component* and allow the user to select from a reduced set of candidates based on his desired subsequent search direction. The ranked values calculated for keyword candidates enable the user to set the numeric threshold for candidates to appear in the suggestion list.

5. System implementation

A web-based procurement prototype system incorporating a keyword learning network was developed to evaluate the extent to which the guide improves the user's search. The development was based on three-tier distributed client/server architecture. This architecture comprises a user interface system as the top tier server, a process management system as the middle tier and a database management system as the bottom tier [19]. The top tier server provides user services such as session, text input and dialog and display management. The middle tier server provides process management; it executes business logic and rules and can accommodate multiple buyers. The bottom tier server is dedicated to data and file services [11]. The system was implemented on the Microsoft Windows 2000 Server platform with Internet Information Server v5.0, using PHP [13] and MySQL [16]. The system database consisted of three categories of procurement-related data, including suppliers, supply catalogs and specifications.

Fig. 6 shows the flowchart of the system, in which rounded rectangles represent user actions and regular rectangles represent system actions. First, the user logs in and the system identifies the user type according to the user profiles maintained by the system manager. The user begins the search for information pertaining to a procurement package. If learning is desired, the user must activate a learning session before starting the search so the system can monitor the keyword inputs. After completing the search for a procurement package, the user must end the learning session. The system will then present pairs of consecutive keywords entered during the learning session. The user at this time may annotate an appropriate link for each pair of keywords as shown in Fig. 7; the system then updates the learning network accordingly. If learning is not desired, the user can search directly by inputting keywords without defining a learning session, and the keywords entered will not be recorded.

When the search guide is activated, for each keyword that is entered, the system will suggest subsequent keywords, appearing in one window and the search results (a list of URL references) in the other window, as shown in Fig. 8. The user may choose one subsequent keyword, combine several from the suggested list or input another.

The user may also configure the suggestion mechanism in order to determine when, and what, to suggest. For example, one may ask the system to

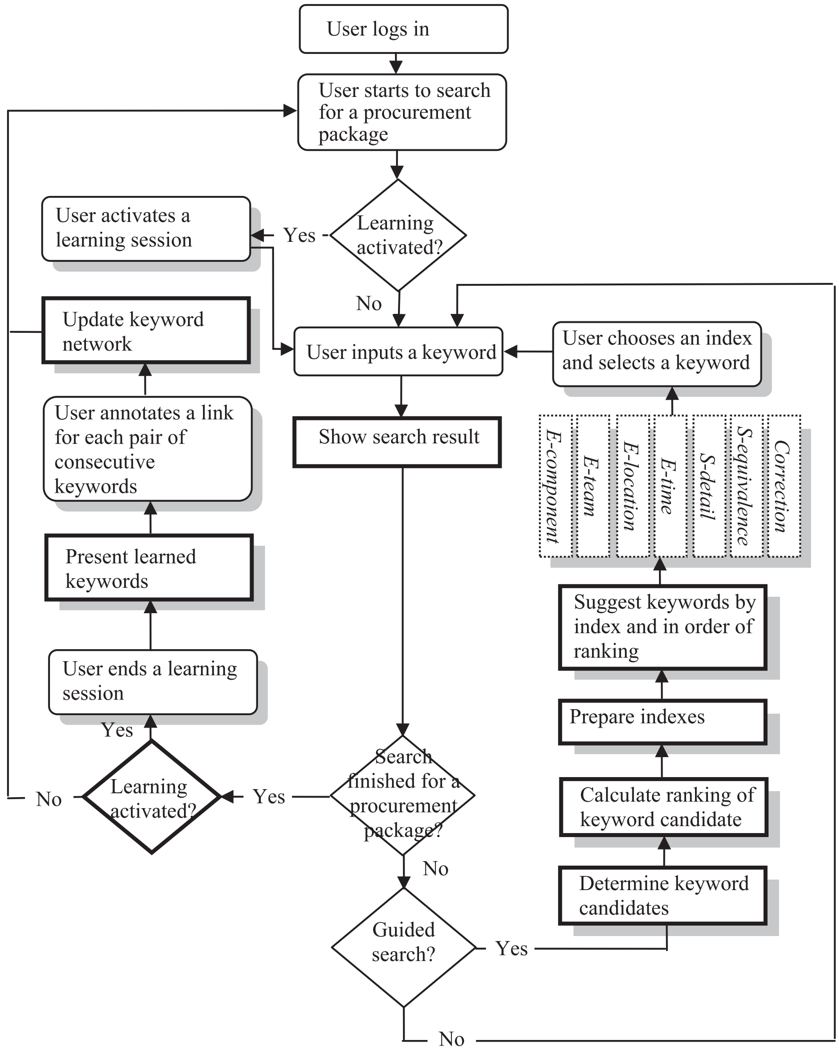

Fig. 6. System flowchart.

suggest subsequent keywords only when the number of resulting URL references is less than 3 or greater than 20, and keywords to be suggested only if their rankings are greater than 10.

6. Experiments

T1he goal of the experiments described here was to evaluate the reduction in keyword inputs due to the guide. The experiment involved monitoring 12 procurement engineers performing procurement tasks on a single project—a US$ 9,000,000 construction project of school buildings consisting of 14 procurement packages with 64 items, as listed in Table 1.

The experiment allowed participants to perform a procurement task twice: one without the guide and the other with the guide activated. To reduce the learning effect, a pretest was performed to divide the project into two “equivalent” subprojects so that they both required approximately the same amount of search work (i.e., same number of keyword inputs). Fifteen graduate students participated in the pretest. Each student was required, for each procurement item, to

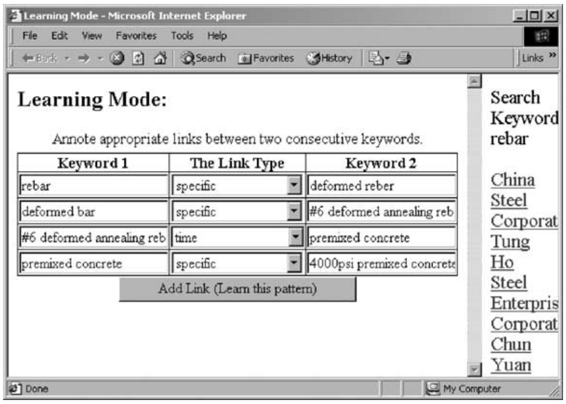

Fig. 7. Dialog for annotating a link between each pair of keyword inputs.

complete the procurement task that included finishing specifications and Request For Quotations (RFQs), and finding contact information for three prospective suppliers. The quantity takeoff for each procurement item was given. Each participant could find information only from the website by inputting search keywords and browsing the resulting URL references. The average number of keyword inputs used for each procurement package was recorded. The packages were then divided into two groups, Subproject I and Subproject II (as shown in Table 1), so that the average of the total number of keywords entered was approximately the same for each group. Thus, the two subprojects of procurement items were assumed to require the same amount of search work.

A questionnaire was sent out to engineers working for the contractor who provided the experimental project data, to ask if they would be interested in participating in the experiment. Of the 15 engineers who agreed to participate, we carried out a further screening process to determine their procurement and web-search background experiences, selecting three to represent each of the four types of users previously described in this paper. In addition, Type I users (with procurement experience and familiarity with the website) and Type III users (with little procurement experience but being familiar with the website) were given a 1-h tutorial and practice session on the website without the guide. Table 2 lists the backgrounds of the 12 engineers who participated in the following experiment.

In this experiment, the system was allowed to learn first from Type I users and then the other types of users were required to complete the procurement task using the procurement site with and without the guide. Type I users were asked to input keywords in the procurement site to search for procurement information associated with any three projects they had recently been or were currently actively involved, with the learning mechanism activated. This learning

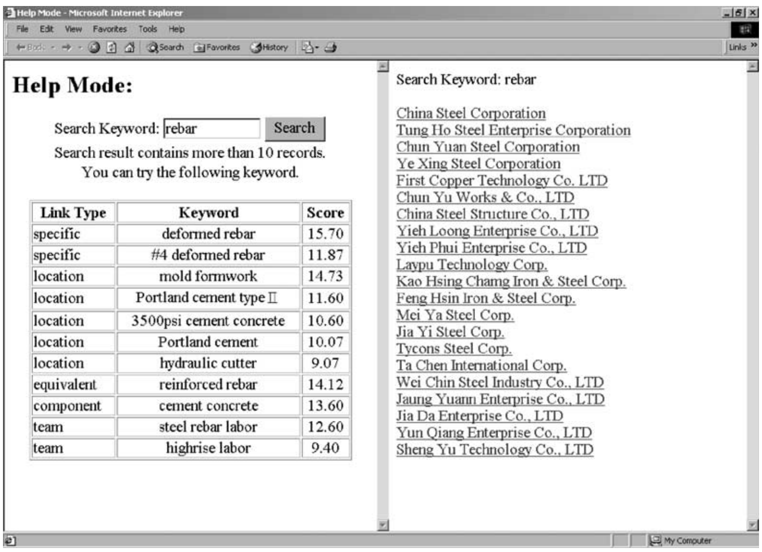

Fig. 8. Left window showing suggested subsequent keywords and right window showing search result.

Table 1
Procurement items for the experimental project

No procurement package	Items		Average number of keyword inputs
Subproject 1 (subtotal)			121.62
1	Pebble pavement	Pebble, sealant material, unskilled laborer	11.67
2	Backfill and compaction	Compactor, unskilled laborer	3.35
3	3200 psi premixed concrete	Type II Portland cement, concrete additive, coarse aggregate, coarse sand, technician, concrete vibrator, concrete conveyor, concrete culvert	17.42
4	Formwork	Formwork 2.5cm, formwork laborer, unskilled laborer, #22 wire, nails	23.67
5	Floor tiling (silica brick)	Tile(30×30cm), tile(15×15cm), 1:3 cement mortar, cement grout filler, tiler, unskilled laborer	16.92
6	Wall finishing and painting	1:3 cement mortar, cement, paint, concrete laborer, painter, unskilled laborer	17.50
7	Wall tiling	Tile(16×8 cm), 1:3 cement mortar, cement grout filler, tiler, unskilled laborer	17.67
8	Roof waterproofing	1:3 cement mortar, asphalt, waterproofing material, technician, unskilled laborer	13.42
Subproject 2 (subtotal)			120.00
9	Site preparation	Excavator, bulldozer, operating driver, unskilled laborer	18.58
10	Survey	Surveyor, carpenter, unskilled laborer	15.67
11	Excavation	Excavator, operating driver	10.00
12	2500 psi premixed concrete	Type IV Portland cement, concrete additive, coarse aggregate, coarse sand, technician, concrete vibrator, concrete conveyor, concrete culvert	31.17
13	Brick laying	Red brick (23×11×6), 1:3 cement mortar, brick laborer, unskilled laborer	17.08
14	Rebar assembly	Rebar, rebar assembly laborer, unskilled laborer	27.50

phase took about 3 weeks due to the busy schedules of the participants. When the learning was completed, each of Types II, III and IV users were asked to complete the procurement tasks for Subproject I without the guide, and for Subproject II with the guide activated. The scope of the procurement task was the same as the one used in the pretest.

Table 3 shows the average number of keyword inputs used to complete the procurement tasks by Types II, III, and IV users for Subprojects I and II. The results indicate that procurement experience and web-search experience both helped in reducing the number of keyword inputs when the guide was deactivated. In other words, participants having more procurement or web-search experience used fewer keywords than those with less experience. It was further found that procurement experience had a larger effect on the reduction than web-search experience, resulting in 89.00 keywords vs. 139.67 without the guide and 70.00 vs. 104.33 with the guide, respectively. It may be that because of their procurement experience, the participants knew what was in the marketplace and which specific terms to enter as keywords. Web-search experience allowed participants to more quickly manipulate the search; the time spent during the search was not measured in this experiment, however, so that part of the effect was not seen.

Table 3 also shows the savings on the number of keyword inputs used with the guide activated (Subproject II) compared to those without the guide (Subproject I). It was found that the guide always reduced the number of keyword inputs for all types of participants. The more keyword inputs the participants needed to use, the more keyword inputs the guide reduced. The percentage savings ranged from 21.3% to 38.8%.

Table 2
Categorization of participants in the experiment

User type	Participant no.	Years of procurement experience	Years of web-search experience	Experience in using the proposed system
Type I	1	4	3	Yes
	2	3	5	Yes
	3	3	2	Yes
Type II	4	11	0	No
	5	6	0	No
	6	5	<1	No
Type III	7	0	7	Yes
	8	0	7	Yes
	9	0	6	Yes
Type IV	10	0	0	No
	11	0	<1	No
	12	0	<1	No

Participation in this experiment took 8 h (4 h for each subproject) and as only 12 engineers participated, the results presented previously need more validation in the future. Nevertheless, the initial findings have shown that the proposed guide system helped engineers reduce the number of keyword inputs required. In addition, knowledge surrounding the magnitude of the reduction of keyword inputs used by each type of user was discovered. However, it must be emphasized that this experiment was mainly concerned with improvement in search effectiveness rather than search efficiency; time, therefore, was not measured. The use of fewer keywords during a search does not necessarily imply that the desired information will be obtained sooner. Many factors contribute to search efficiency, including the design of the user interface that retrieves keywords from the suggested list, the system configuration (e.g., the number of keywords in the suggestion list), the communication speed for accessing URL references, the ease of finding the desired information on suppliers' Web pages and the speed for writing procurement specifications. All of these factors added to the difficulty of conducting an experiment to reach a conclusive outcome with a prototype system.

7. A comparison with other commercial procurement websites

Table 4 compares the information content and search functions of the proposed system with other commercial procurement websites. These websites may contain information about the buyers and/or sellers, procurement announcements, technical specifications, standards and codes regarding the products,

Table 3
Average number of keyword inputs used by each user type

Procurement package	No. of items	II	III	IV
Subproject I (without guide)				
Pebble pavement	3	7.33	12.67	18.67
Backfill and compaction	1	1.00	4.67	5.00
3200 psi premixed concrete	8	8.33	16.67	27.67
Formwork	5	20.67	32.00	28.00
Floor tiling (silica brick)	6	12.33	19.33	23.00
Wall finishing and painting	5	14.67	18.33	20.67
Wall tiling	5	14.33	20.33	19.00
Roof waterproofing	5	10.33	15.67	16.00
Subtotal	38.00	89.00	139.67	158.00
Subproject II (with guide)				
Site preparation	4	12.33	15.67	13.33
Survey	3	6.33	16.33	12.33
Excavation	2	4.67	7.67	7.33
2500 psi premixed concrete	9	22.00	26.33	26.67
Brick laying	4	11.00	13.33	12.67
Rebar assembly	4	13.67	25.00	24.33
Subtotal	26	70.00	104.33	96.67
Saving on number of search keywords		−19.00	−35.33	−61.33
Percentage saving on number of search keywords		−21.3%	−25.3%	−38.8%

Table 4
Comparison of the proposed system and other commercial procurement websites

Information search mechanisms	Proposed system	AEC Info	BuildPoint	bLiquid.com	ProcureZone
Information content					
Buyers and sellers	•	•	•		•
Announcements	•		•	•	•
Technical specifications	•	•		•	•
Standards and codes	•				•
Miscellaneous news	•		•		
Search functionality					
Index	•	•	•	•	•
Keyword search (single field)	•	•	•	•	
Keyword search (multiple fields)		•		•	
Multiple choices for sorting search results	•				
Further filtering of search results			•	•	
Correction of input keyword	•				
Suggestion of potential input keywords	•				
Learning capability	•				

as well as other miscellaneous news in the construction industry.

Most procurement websites offer an index (i.e., information categorization) and a keyword search function. Websites such as bLiquid.com emphasize a single product category (e.g., construction equipment) and offer a field of multiple keyword inputs (e.g., year and make). Some of them (e.g., BuildPoint and bLiquid.com) also offer a filtering capability by allowing the input of a second keyword to narrow down the initial search result. On the other hand, our proposed system offers multiple choices for sorting the search results, it corrects inappropriate keywords, and suggests potential input keywords for possible use in the subsequent search.

8. Conclusions

This research has recognized that professional procurement experience can help users search more effectively for website information, and has developed a learning model that captures such experience in order to guide inexperienced users in their search. The proposed learning network requires experienced users to annotate the relationships among a series of keyword inputs used during their searches, in order to suggest possible subsequent keywords to help inexperienced users conduct their own search. The guides included are correction, specification-by-equivalence, specification-by-detail, extension-by-time, extension-by-location, extension-by-team and extension-by-component, each helping engineers search for information in the desired direction. The ordering of suggested keywords is calculated by considering the number of links emanating from the keyword entered, the number of links connecting to the subsequent keywords for consideration, the experience level of the user annotating the link and the number of times a user actually adopts a suggested keyword, based on that link.

The experiments described in this paper have shown that participants with more procurement or web-search experience used fewer keywords in their searches than those with less experience. Participants having more procurement experience but less web-search experience also used fewer keywords than those having less procurement experience but more web-search experience. The experiments also demonstrated that the guides saved the number of keyword inputs used for all types of participants, ranging from 21.3% to 38.8%, when compared to the number of keyword inputs used without help from the guide.

The future direction of our research will be to improve the proposed learning mechanism, such as learning to correct misspelled keywords via a correction table rather than in the learning network, as a correction guide often has higher success rates than other types of guides in predicting subsequent keywords. Other improvements would include adding a guide for translation between different languages or industries (e.g., contractor jargon vs. terms used in the steel industry) and annotation interfaces (e.g., suggesting an appropriate link between each pair of keywords to reduce the annotation time of experienced users).

Acknowledgement

The authors would like to thank the National Science Council, Taiwan, for financially supporting this work under Contract No. NSC-92-2211-E009-064.

References

[1] AEC Info. <http://www.visualibrary.com> (last updated, 2004) (last visited, June 21, 2004).
[2] M. Balabanovic, Y. Shoham, Learning information retrieval agents: experiments with automated web browsing, Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Resources, AAAI Press, Stanford University, CA, 1995, pp. 13–18.
[3] bLiquid.com ⟨http://www.bliquid.com⟩ (last updated, 2004) (last visited, June 21, 2004).
[4] B.-C. Björk, Electronic document management in construction—research issues and results, Electronic Journal of Information Technologies in Construction 8 (2003) 105–117 (CIB (International Council for Research and Innovation in Building and Construction)).
[5] BuildPoint ⟨http://www.buildpoint.com⟩ (last updated, 2003) (last visited, June 21, 2004).
[6] A. Carlyle, Developing organized information displays for voluminous works: a study of user clustering behavior, Information Processing and Management 37 (2001) 677–699 (Elsevier).
[7] A. Chavez, P. Maes, Kasbah: an agent marketplace for buying and selling goods, Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent, Practical Application, London, UK, 1996, pp. 75–90.
[8] D.W. Cheung, B. Kao, J. Lee, Discovering user access patterns on the World-Wide Web, Knowledge-Based Systems 10 (1998) 463–470 (Elsevier).
[9] Citadon ⟨http://www.citadon.com⟩ (last updated, 2004) (last visited, June 21, 2004).
[10] P. Craven "Google's PageRank explained and how to make the most of it" Web Workshop. ⟨http://www.webworkshop.net/pagerank.html⟩ (last updated, 2003) (last visited, June 21, 2004).
[11] H. Edelstein, Unraveling client server architectures, DBMS Magazine 7 (5) (1994) 34–43 (Miller Freeman).
[12] P. Ingwersen, Cognitive perspectives of information retrieval interaction, Journal of Documentation 52 (1) (1996) 3–50 (Emerald Group Publishing).
[13] PHP, PHP: Hypertext Preprocessor ⟨http://www.php.net⟩ (last updated, 2004) (last visited, June 21, 2004).
[14] PrimeContract ⟨http://www.primecontract.com⟩ (last updated, 2004) (last visited, June 21, 2004).
[15] ProcureZone ⟨http://www.procurezone.com⟩ (last updated, 2002) (last visited, June 21, 2004).
[16] MySQL ⟨http://www.mysql.com⟩ (last updated, 2002) (last visited, June 21, 2004).
[17] A. Spink, Study of interactive feedback during mediated information retrieval. Journal of the American Society for Information Science, American Society for Information Science and Technology 48 (5) (1997), 382–394.
[18] Y. Qiu, H.P. Frei, Concept based query expansion, Proceedings of SIGIR-93, 16th ACM (Association for Computing Machinery) International Conference on Research and Development in Information Retrieval, ACM, Pittsburgh, PA, 1993.
[19] H.F. Softis, LOUIS White Paper ⟨http://www.softis.is⟩ (1998) (last visited, June 21, 2004).
[20] T.D. Wilson, Models in information behaviour research, Journal of Documentation 55 (3) (1999) 249–270 (Emerald Group Publishing).