A Context Based Approach to Acquisition and Utilization of Personal Knowledge for WWW Browsing

Toru Takashiro

RACE, The University of Tokyo
Komaba 4-6-1, Meguro-ku,
Tokyo 153-8904, Japan

takasiro@race.u-tokyo.ac.jp

Hideaki Takeda

National Institute of Informatics
Hitotsubashi 2-1-2, Chiyoda-ku,
Tokyo 101-8430, Japan

takeda@nii.ac.jp

Abstract: *Our personal knowledge plays a vital role in our daily intellectual work. This paper proposes a method for acquiring and utilizing our personal knowledge in computer systems. We developed a system called MindHeap that helps us to acquire personal knowledge by browsing WWW (World Wide Web) hypertexts. Our approach for acquiring a user's personal knowledge is to use situational context which has important effects for acquiring and reminding knowledge. Situational context means information which characterizes her/his situation of information activities. MindHeap can help the user to organize knowledge on WWW pages by finding and recording related words as situational context for topics specified by the user. The system extracts a list of related words as situational context, and records it with the topics. Thus, topics and their contexts are accumulated, then they can work as knowledge for the user, e.g., she/he can retrieve pages of which situation is similar to the current page.*

1. INTRODUCTION

In various situations in our everyday life, we rely on knowledge. We have some common knowledge, and more important part of knowledge in daily life is peculiar knowledge for everyone. We call this type of knowledge as "personal knowledge". That holds our personal character peculiar to each of us. If a keyword is given to some different people, they would associate the keyword with different things. For example, the word "agent" is given to us, we may associate it with various things such as "ad agent", "distributed computer systems with agent technologies", "spy novels". The difference among these associations is derived from the difference among our memory of experiences. A person developing computer system is likely to remember "computer agent" for the keyword "agent", and a person who have read a lot of spy novels may remember "spy agent" for it. In this paper, we define knowledge that is based on our experience and peculiar to each of us as personal knowledge.

Many cognitive psychologists such as Bruner[1] and Norman[2] assert that there are two modes of cognition. Bruner calls one as the paradigmatic mode, and the other as narrative mode. According to Bruner, the paradigmatic mode is the mode of explanation. In this mode, we tend to explain things logically and abstractly. On the other hand, the narrative mode is the mode of understanding. We can understand things with peculiar ways based on our experiences that are different from each other. Both modes are important for our intellectual work.

It is therefore needed to treat knowledge in both mode of cognition to deal with knowledge on computer systems. Our way of treating knowledge in the paradigmatic mode resembles that of computer systems, since computer systems have well-defined knowledge bases and reason logically. However, such computer systems cannot treat the mode of narrative, and ignore personal dependent properties of knowledge. These systems postulate that knowledge is shared by all the people in common. Therefore, users are often dissatisfied with the systems because these systems cannot work as their expectations. Consider the case in which we use search engines on WWW. When a user sends a query to search engines, they retrieve web texts that match with the query from WWW. The same results are given for another user who sends the same query to search engines. It is undesirable behavior with respect to the narrative mode of cognition because expectations of different users should be different even if they ask the same query.

If it is possible to let computer systems know our personal knowledge, we are able to overcome such disappointment. Our final goal is acquiring a user's personal knowledge in computer systems, and utilizing the personal knowledge for the user.

2. CONTEXT AND KNOWLEDGE

According to Rumelhart[3], there are the contextual

effects in remembering. These effects play an important role in our remembering and recalling knowledge.

When we store our remembrances in memory, they are kept with information about the situation in which the remembrance is acquired. This situational information is seemingly vain for our memory, but it is conceivable to work as the key for recalling the remembrance.

We define this situational information in which we acquire remembrance as situational context. The word "context" can refer to many things. In the narrow sense of the word, the context of a word or sentence consists of the words or sentences before and after it. In addition, it is often used to mean the general situation in which something exists or occurs. In this paper, we concerned with the later meaning.

The situational context makes our personal knowledge peculiar to each of us. As we mentioned in section 1, our personal knowledge derives from our each individual experiences. Such remembrances are accumulated in our memory, i.e., they are associated to each other with situational context. Since our knowledge consists of such associated remembrances, it has a proper structure to each of us.

In order to acquire personal knowledge in computer systems, we need to know situational contexts in which the knowledge is acquired, then to make use of these contexts as a cue to construct personal knowledge in the systems.
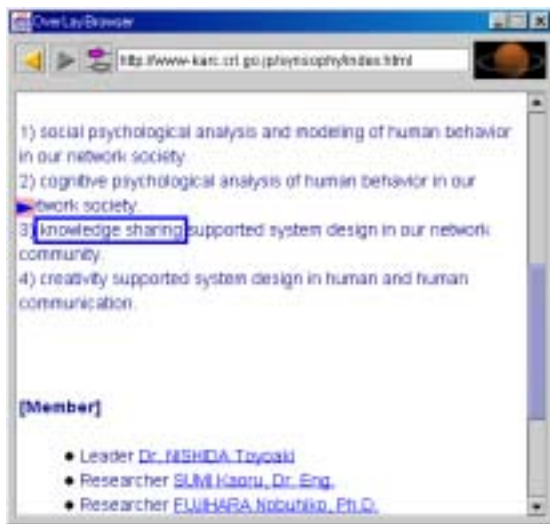

Figure 1: A snapshot of MindHeap

3.   BEHAVIOR OF THE SYSTEM

We developed a prototype system called MindHeap, that can assist human behavior to acquire, store and utilize personal knowledge on WWW browsing. A user on WWW browsing can acquire concepts or facts through text reading on

pages. But this acquisition is often done implicitly. What s/he remembers explicitly is that s/he read some words interesting to her/him. Even with such conscious remembrance, s/he can retrieve what s/he read from various hints because of unconscious part of remembrance, i.e., situational context.

Our system works as "intelligent memo system" by simulating the above process. The user just points out her/his attention by drugging some words or phrases as "memo item", then the system stores the memo item with its context. The stored information is used when s/he encounters similar contexts.

3.1 Acquisition of Context

When the user point out their attended phrases, the system acquires these phrases as the input. And then the system attempts to extract the situational context behind the inputted phrase from the document that is displayed on the browser at the time. In MindHeap, situational context is represented as list of extracted words appendant to the inputted phrase (see Figure 2).
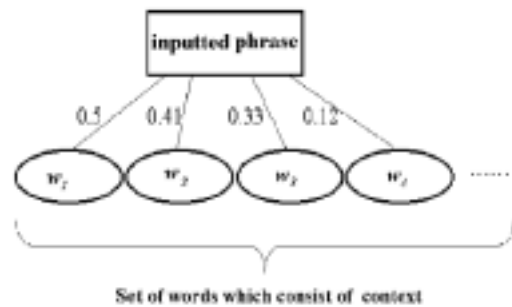

Figure 2: Unit which composes of personal knowledge.

The extracted words that are considered to consist of situational context are linked with the inputted phrase, and each link has a value which represents degree of relation between the extracted word and the inputted phrase. This pair of inputted phrase and extracted context words is the unit which composes of the user's personal knowledge.

To extract these context words, we need to consider two following effects.

a) The information about situation on the time in which the user inputs phrase: As we have described in section 2, when we remember something explicitly, we also implicitly remember the information about the environment with the user's attention. In browsing WWW, we can assume that the document which the user browses on the time is an environment of the user's attention.

Therefore, the system attempts to extract situational context words from the document showing on the time.

b) The influence of the user's browsing history on the extracted context words: If a user browses same document, different browsing history makes different browsing experiences for the user. Therefore, we must consider historical influence on the context.

The context is calculated as follows;

Firstly, the system picks up nouns I_N in the inputted phrase I.

$$I_N = \{ w{:}w \quad noun, w \quad I \}$$

Secondly, it collects a set of words that would be associated to the nouns above by calculating occurrence in the given document. Suppose $co(w, w')$ is occurrence degree of word w and w' in the given document D. Then the degree of association to the inputted phrase is defined as follows;

$$C_{w'} = \sum_{w \in I_N} co(w, w') \quad (w' \quad noun, w' \quad D)$$

In the current implementation, $co(w, w')$ is a number of occurrence where word w and w' are appeared in at most seven consecutive words in the document. We adopt top 30 words in Cw as context words and call them Wc.

Thirdly, we obtain "non-historical" context $C'w$ by normalizing Cw.

$$C'_w = \frac{C_w}{\sum_{u \in W_c} C_u} \quad (w \in W_c)$$

In order to consider browsing history, we modify the context by emphasizing it with contexts of the past five pages. If a word in the "non-historical" context is often appeared in the past pages, the word should be more important.

We just regard context of a page as occurrence pattern of words in that page. Suppose Fw is a number of occurrence of word w in the past page. Then we can obtain its normalized version $F'w$ according to the set of words in the "non-historical" context.

$$F'_w = \frac{F_w}{\sum_{u \in W_c} F_u} \quad (w \in W_c)$$

"Historical" context $Ch'w$ is defined as the exaggerated version of "non-historical" context by the past pages in the following way.

$$Ch_w = C'_w \ (F'_w + 1)$$

$$Ch'_w = \frac{Ch_w}{\sum_{u \in W_c} Ch_u} \quad (w \in W_c)$$

Due to memory limitation, the system can only record values of Fw for a limited number of words. Then we may encounter un-recorded Fw in calculation of Chw. We assign 0 for Fw if Fw is not found in the record.

3.2 Representation of Personal Knowledge in the System

Phrases which a user inputs to the system on browser and context words appendant to the inputted phrases are recorded on the system's memory as units shown in Figure 2. The more the user inputs phrases which she/he attempts to, the more units are appended to the memory. When a new phrase with contexts is added, phrases which have same word in their contexts are tied up with those words (see Figure 3).

Figure 3: An example of graph structure

As a result, the phrases that the user inputted to the system are organized as a graph structure in which context words shared by some phrases mediate between the phrases. For example, in Figure 3, Phrases such as "Design as intelligent activity" and "Ontology for Design" are mediated with context words, "research" or "design". It means that concepts represented by these phrases have associative relation for the user who input these phrases with the contexts which are represented by context words

The graph structure is improved by the user's input, and it can work as personal knowledge for the user.

3.3 Utilization of Personal Knowledge

With the structure of a user's personal knowledge described above, we try to navigate the user with suggesting related pages to the user.

When the user points out her/his attended phrase, MindHeap can suggest links to related pages to the user. After context extraction process for the phrase, the system scans the graph structure described in Section 3.2 for phrases which related with the newly inputted phrase. If exists, the system displays these phrases with context words which mediate these phrases and the newly inputted phrase (Figure 4).
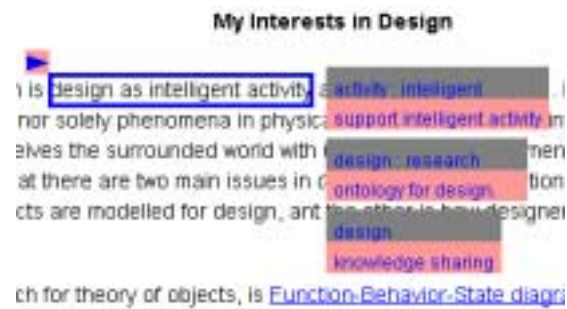

Figure 4: Suggestion of links to related pages

For example, in Figure 4, three phrases are suggested to the user for the newly inputted phrase "design as intelligent activity". In Figure 4, the user marks "design as intelligent activity". Then the system suggests their related topics, i.e., "support intelligent activity", "ontology for design" and "knowledge sharing". The lower part of a box shows the name of a topic that was marked by the user, and the upper part shows context words that mediate that topic and the original topic.

These suggested phrases work as hyperlinks. The user can link to the page in which she/he input the phrase by clicking suggested phrase. And by clicking the icon which appears at the upper left corner of the dragged phrase, the user can send the query to a search engine. The query includes not only the inputted phrase but also extracted context words, so the user can find new WWW pages which are related in her/his browsing context.

4. RELATED WORK

There found many systems to acquire and integrate personalized information for a user from WWW.

WebWatcher[4] recommends hyperlinks based on a user's interests which the system asked to the user at the beginning. Letizia[5] is also a recommendation system. It watches a user's browsing behavior, presumes the user's interest and recommends web pages to us. Margin Notes[6] works as a proxy server for WWW, and it appends

recommendations to original WWW texts. They are intended for automated browsing with fewer user actions, but our system aims to interactive knowledge acquisition system where the system and the user cooperate to improve knowledge.

Hirashima *et al.*[7] emphasized the importance of context on hypertext browsing. They use mainly a user's browsing history as context, whereas our system uses situational context.

5. CONCLUSION AND FUTURE WORK

In this paper, we discussed how personal knowledge should be acquired in computer systems with using situational context and proposed a prototype system called MindHeap that can simulate our acquiring and recalling process of personal knowledge. With this system, the user can easily capture her/his interesting topics and also easily retrieve them, because the system collects and organizes contexts of topics automatically. Of course, our current implementation is weak. We must consider possibility to manage context other than situational or historical context, and analyze the results achieved.

We are planning to apply our system to multi-user system. In the current implementation, our system can only work for a user. To expand our system into multi-user system, we will share and exchange acquired personal knowledge with context.

REFERENCES

[1]J.Bruner. *Actual Minds, Possible Worlds.* Harvard University Press, 1986.
[2]D.A.Norman. *The Psychology of Everyday Things.* Basic Books, 1988.
[3]D.E.Rumelhart. *Introduction to Human Information Processing.* Wiley, 1977.
[4]R.Armstrong, D.Freitag, T.Joachims, and T.Mitchell, "WebWatcher : A Learning Apprentice for the World Wide Web", *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, 1995.
[5]H.Lieberman. "Letizia : An Agent That Assists Web Browsing", In *Proceedings of IJCAI`95*, 1995.
[6]J.Rhodes. "Margin Notes: Building a Contextually Aware Associative Memory", In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '00)*, 2000.
[7]T.Hirashima, N.Matsuda, T.Nomoto, J.Toyoda, "Context-Sensitive Filtering for Hypertext Browsing", In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '98)*, 1998.