

Toward Personalized Digital Library for Providing “Information JIT”

Hisashi Ikeda, Naohiro Furukawa, Katsumi Marukawa, and Hiromichi Fujisawa

Central Research Laboratory, Hitachi, Ltd.,
1-280 Higashi-Koigakubo Kokubunji, Tokyo, 185-8601, Japan
h-ikeda@crl.hitachi.co.jp

Abstract. The personalized digital library for providing users with proper information just in time (JIT) is proposed. This digital library can support not only document retrieval task, but also users’ reading and comprehension task. This function can be realized by recording the person-dependent information, which is about human behavior toward documents, such as highlighting, marking, annotating, etc. By utilizing the digital pen and paper with dot-pattern, handwritten annotations on paper documents can be simultaneously digitized and stored in the corresponding e-document. Furthermore, hybridization of real world and e-world in digitizing personal annotations is proposed. This information on users’ annotations makes it possible to retrieve document based on the implicit memory and to provide information support appropriate to the situation with users. The personalized digital library proposed in this paper is to be developed.

1 Introduction

What do we search documents from Digital Library (DL) systems for? What are we going to do after retrieving documents? In general, we access DL systems in order to acquire information which is novel for us or more detail than what we have. We read the retrieved documents and try to comprehend the contents of them. After that, the documents are interpreted in some users’ manners, and they are memorized as our knowledge. In other words, our knowledge acquisition process consists of (1)to search or to retrieve, (2)to read and to comprehend, and (3)to organize and to store.

Many of the DL systems which are currently available support the search and re-trieval of documents. Users of DL systems can retrieve documents relevant to their needs by providing keywords with DL systems, and browse retrieved documents by referring document clustering results or summaries of documents[1]. Result of document retrieval is, however, usually independent of users, and organization of the library is usually stiff.

On the other hand, current DL systems didn’t seem to support users’ reading, comprehending and knowledge organizing task. It is because human knowledge acquisition is dependent on person, and on situation. Summary of a document, for example, is different from person to person.

The target of this research is to build up a personalized digital library, in other words, a personal information environment(PIE), which can support the reading, comprehension, and knowledge organization task by extending human memory. This personalized digital library stores not only documents but also human behavior toward documents, such as how problems are solved, and which part of the document is focused under current situation. By using this kind of information, the system can provide more appropriate information or suggestion, in advance, like which kind of query should be input, that is, substantial information “just-in-time.”

As a first step toward the personal digital library, in this paper, we pay attention on annotations on documents, and consider a method of digitizing annotation for recording human action to documents.

Following chapters describe personalization of digital library as a mean of extending human memory at first, hybridization of real world and e-world for making personal annotations computer-accessible as human actions to documents, and for providing information JIT.

2 Personalized Digital Library as Human Memory Extension

There are two possible situations in accessing documents in DL; the first one is the general document retrieval based on similarity between a query and documents in the document vector space, the second one is the subjective retrieval based on a user’s memory which is not only explicit but implicit. The precision of document retrieval by the user is sometimes higher in the second situation than in the first situation, because the user’s memory directly expresses how the user interpreted the documents and which document he/she associated one document with under some specific situation, which straightforwardly correspond to the user’s needs in similar situation.

Personalization of DL is regard as to store user-dependent or situation-dependent information in addition to documents themselves. This means that human memory extension, in that human memory is expressed on the computer more explicitly.

Now, what we do to comprehend documents is considered as follows. In order to remember the documents more deeply, we often make paper copies to read because paper is easy to glance over and suitable for careful reading[2]. Furthermore, we (1)highlight passages by underscores and marks, (2)make annotations on the margins, (3)bookmark the pages, (4)take notes, (5)make summarization, (6)refer other information sources (documents, web pages), (7)file the document into a categorized folder, and (8)rearrange the hierarchical folder structure.

Thus, annotation on paper documents (including underscore or mark) is one of key clues for revealing what the user thought, and how the user interpreted in reading and comprehension process. Therefore, to store information about annotation as metadata of documents in the library is effective for user-dependent document retrieval and for providing information JIT. To sum up, digitizing

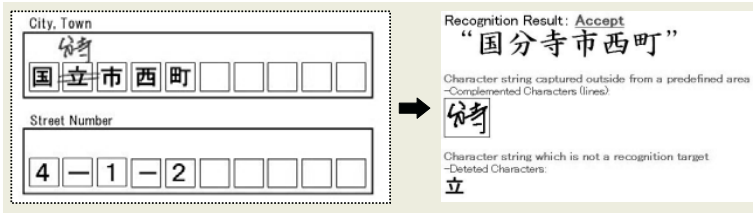


Fig. 1. An example of the personal annotation by Digital Pen

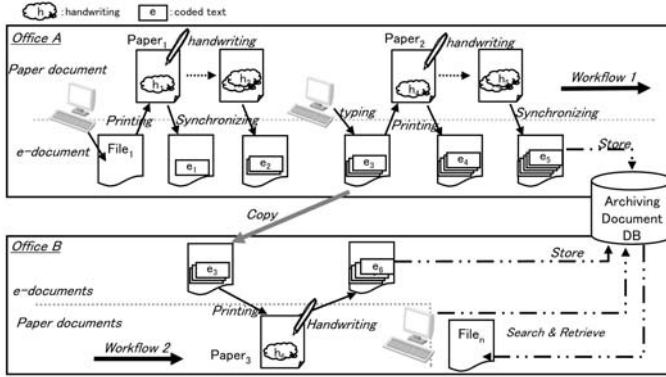


Fig. 2. Integration of real-world annotation with e-world annotation

handwritten annotations on paper documents is an essential function in the personalized digital library proposed in this paper.

3 Hybrid of Real and e-World for Digitizing Personal Annotation

To realize the digitization of personal annotations, we adopted the digital pen[3]. This device generates pen tracks data by sensing dotted patterns pre-printed on paper. By assigning unique dot pattern for each paper, the digital pen can identify on which paper and where in the paper annotation was made. Therefore annotated strokes and/or texts can be associated with the words or sentences originally printed on the paper.

Fig.1 shows an example of a personal annotation. This annotation means correction of a character string. In this example, some characters in boxes are replaced into the characters written above the boxes. As a result of interpretation of the annotation, character deletion is detected, the deleted characters are identified, and the remaining characters including inserted characters are collected and recognized correctly.

In this way, handwritten annotations are interpreted and simultaneously stored in the personalized digital library.

On the other hand, annotations are not only handwritten but also typed. As described before, when referring other web site, or associating other document files with the current one, it is convenient to make annotations electronically on computer. Furthermore, annotated URL string written by using the digital pen should be linked to the specified web page on computer. Such web page could be printed for careful reading, and handwritten annotations should be made on paper.

Thus, in order to record users' reading and comprehending process naturally, annotations both on paper and on e-document should be digitized. As shown in Fig.2, every e-document can be printed paper document at any time. Once handwritten annotations are made (on Paper 1), they are simultaneously digitized, and new version of the e-document is generated, while old e-document is labeled as expired one. When annotations are made on computer, new version of the e-document is also generated, and at the same time, old paper document (Paper 1) should be flagged as an invalid document. It is realized by the paper and dot-pattern management software. If several paper documents are printed from the single e-document, e-document is prepared respectively by duplicating the original e-document file.

In this way, personal annotations can be digitized and the history of the annotations is stored as metadata of the document in the library (in Fig.2, we call it "Archiving Document") for later use, that is, for providing appropriate information JIT.

4 Conclusion

We proposed a personal information environment, featuring the combination of e-world and real-world. Retrieved documents can be printed, highlighted, marked, and annotated. All the added information is simultaneously stored in the computer for later use. The personal digital library we would like to construct will also feature reading assistance, contents analysis, and intelligent knowledge management by utilizing the information on human action in reading and comprehension process. According to this proposal, we are going to implement a prototype system.

References

1. William C. Janssen and Kris Popat: UpLib: a Universal Personal Digital Library System. Proceedings of the 2003 ACM symposium on Document engineering. (2003) 234-242
2. Abigail J. Sellen and Richard H. R. Harper: The Myth of the Paperless Office. MIT press, (2001).
3. <http://www.anoto.com>
4. Francois Guimbretiere: Paper Augmented Digital Documents. Proceedings of the 16th annual ACM symposium on User interface software and technology. (2003) 51-60