DAS-1-1

On a Knowledge Enhanced Search Method for Personal Academic Information Collections

Baoning Li[1] Toru Abe[2] Yukio Iwaya[3] Tetsuo Kinoshita[2]

Graduate School of Information Science, Tohoku University[1] Information Synergy Center, Tohoku University[2]
Research Institute of Electrical Communication, Tohoku University[3]

1 Introduction

In order to efficiently search personal academic information collections, it is necessary to properly consider the characteristics of this kind of collections. Particularly, the following two types of knowledge should be effectively used to enhance a search process:

- **K_R** Knowledge on Relations among the personally collected academic materials. When building a personal academic information collection, a researcher often filters and gathers academic materials that are related to each other in various ways[1]. Consequently, a strong relation between these filtered materials often implies a high possibility that they are relevant to same search requests.
- **K_U** User's Knowledge about the personally collected academic materials. When collecting and using academic materials, a researcher usually makes various evaluations on them. This kind of accumulated knowledge can be used to improve search.

Based on this consideration, we propose a knowledge enhanced search method for personal academic information collections.

2 Use of the AIR Scheme

When a knowledge enhanced search is performed in the distributed environment, K_U and K_R should be maintained and used efficiently. We adopt the Active Information Resource (AIR) scheme [2] to achieve this. As depicted in Fig. 1, an AIR is an information resource enhanced with proper knowledge and functions. Moreover, distributed AIRs can autonomously interact with each other to perform information cooperation.

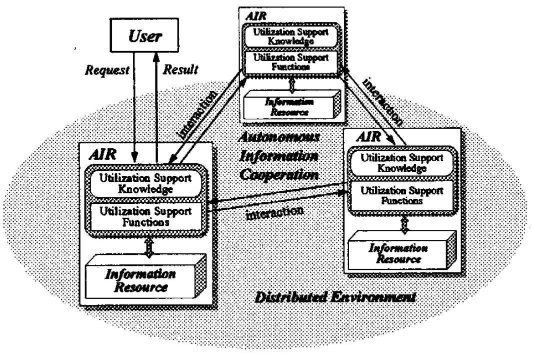

Fig. 1 Concept of AIR

Using the AIR scheme, we can extend a piece of personally collected academic material into a PerC-AIR (Personally Collected academic material AIR). In a collection of PerC-AIRs, knowledge including K_R and K_U can be maintained and used autonomously and effectively, and the proposed knowledge enhanced search can be conducted actively and efficiently.

3 Design of a PerC-AIR System

3.1 Construction of PerC-AIR

Main knowledge items of a PerC-AIR include *title*, *subject*, *creator*, *information about related PerC-AIRs*, and so forth. In addition, a PerC-AIR also keeps knowledge about communication and cooperation. Moreover, a PerC-AIR has proper functions to maintain and use its knowledge, and to communicate with other PerC-AIRs or applications via messages.

3.2 Autonomous Maintenance of K_R and K_U

- When a new PerC-AIR is added into a collection, all PerC-AIRs will interact with each other to discover their relations and update their AIR-K_R.
- Each PerC-AIR keeps track of its user's operation on a corresponding academic material, so that it can record the user's evaluation on that material, and update its AIR-K_U.

3.3 K_R Enhanced Search in PerC-AIRs

A process of K_R enhanced search in PerC-AIRs derives from the spreading activation approach [3]. As depicted by Fig. 2, the PerC-AIRs that match the request are regarded as initial relevant ones. From relevant PerC-AIRs, the relevance spreads through related PerC-AIRs during continuous iterations. In addition, a threshold is set to filter those PerC-AIRs that get low relevance values.

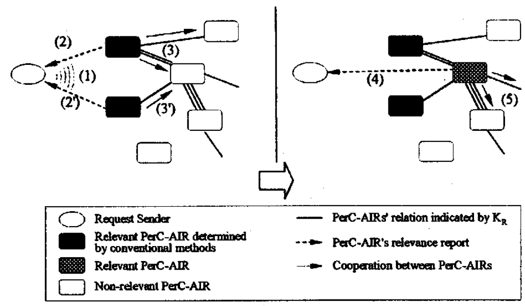

Fig. 2 K_R Enhanced Search in PerC-AIRs

The influence of a relevant PerC-AIR on a related one is determined by the following factors:

- **Relevance value of the relevant PerC-AIR.** A PerC-AIR with higher relevance value should have more influence on its related PerC-AIRs.
- **Strength of their relation.** It is measured as the number of relations between the two PerC-AIRs.
- **Depth of iteration.** Such influence decreases when the process iterates.

Furthermore, an accumulated relevance is normalized using a hyperbolic tangent function, so that it is limited between 0 and 1.

3.4 K_R and K_U Enhanced Search in PerC-AIRs

The process for the K_R enhanced search is conducted, except that K_U is used in each step of the K_R enhanced search. At the beginning of a search, PerC-AIRs that were previously evaluated as relevant for a same request will be initial relevant PerC-AIRs. On the other hand, PerC-AIRs that were previously evaluated as nonrelevant for a same request will judge themselves as nonrelevant for current search. During the later iterations, these PerC-AIRs will not accumulate their relevance values.

4 Experiments

4.1 Experimental Setting and Implementation

In the experimental system, 110 PerC-AIRs were built, each for one piece of academic material such as a paper, a common article, or a presentation. The PerC-AIRs are randomly located in 3 networked computers, so that the experiments can be conducted in the distributed environment.

We applied the agent-oriented approach to implement the experimental system [2]. One agent is used to enhance and extend a piece of academic material into a PerC-AIR. A user agent is also developed to provide an interface between the user and PerC-AIRs.

4.2 Experiment Procedure

30 most popular keywords used by the 110 pieces of materials were chosen to be used as requests. Before the experiments were conducted, the user of the materials determined relevant materials for each keyword.

For each request, the following three search strategies were used to perform the experiments. The average results for all 30 requests are used for evaluation.

S0) Baseline A conventional keyword matching method is used as a baseline.

S1) Using K_R All initial relevant PerC-AIRs are determined via the same process as in S0. Subsequently, K_R enhanced search is conducted.

S2) Using K_R and K_U Firstly, top 10 results retrieved with S1 are evaluated to update AIR-K_U of corresponding PerC-AIRs. Next, a process similar to S1 is conducted, with the newly obtained AIR-K_U used.

4.3 Experiment Results

In order to evaluate the results of search, we calculated the following values for each search strategy: recall, precision and F-Measure, which is a weighted harmonic mean of recall and precision.

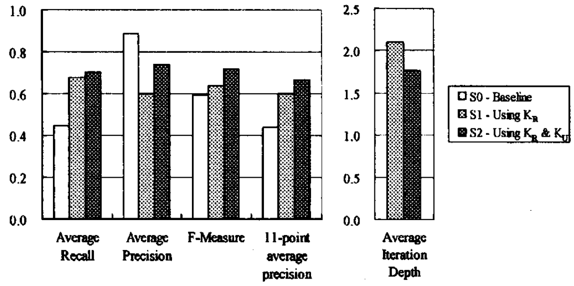

Fig. 3 Experiment Results

As shown in Fig. 3, with the knowledge enhanced strategies, average recall, F-Measure and 11-point average precision can increase apparently. This confirms that the use of knowledge can improve the performance of search.

Comparing the results of the two knowledge enhanced search strategies, i.e. S1 and S2, we verified that the search with S2 can get more relevant results and less nonrelevant results. Furthermore, the average final iteration depth with S2 is approximately 15.9% lower than that with S1, indicating that the use of K_U can improve the efficiency of search.

5 Conclusion

In this paper, we proposed a knowledge enhanced search method for personal academic information collections in the distributed environment. Particularly, we demonstrated that K_R and K_U can be used to effectively improve the performance of a search. Moreover, we presented a design based on the AIR scheme, which was verified to be appropriate for enabling the proposed search method in the distributed environment.

References

[1] C.M. Brown, "Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists", *Journal of the American Society for Information Science*, Vol.50, No.10, pp.929–943, 1999.

[2] B. Li, T. Abe, K. Sugawara, and T. Kinoshita, "Active Information Resource: Design concept and example", in *Proc. 17th Int'l Conference on Advanced Information Networking and Applications*, pp.274–277, 2003.

[3] G. Salton, and C. Buckley, "On the use of spreading activation methods in automatic information retrieval", in *Proc. 11th Int'l Conference on Research and Development in Information Retrieval*, pp.147–160, 1988.