A DISTRIBUTED ENVIRONMENT FOR EFFECTIVE INTERNET SEARCH USING INTELLIGENT PERSONAL AGENT AND DISTRIBUTED KNOWLEDGE BASE

Chun Che FUNG[1] Pornpit WONGTHONGTHAM[2]

[1] *School of Electrical and Computer Engineering*
Curtin University of Technology, GPO Box U 1987, Perth, Western Australia 6845
Tel: +61 8 9266 2575, Fax: +61 8 9266 2584, Email: tfungcc@cc.curtin.edu.au

[2] *Mathematics Department, Science Faculty*
Naresuan University, Muang, Phitsanulok,Thailand 65000
Tel: +66 55 261 000, Fax: +66 55 261 062, Email: pornpitw@nu.ac.th

Abstract : With the rapid increase in the available information over the Internet, it is important to develop an efficient search and retrieval mechanism. In particular, the use of distributed search will be beneficial to many users. This paper presents two proposals to implement such approach. The first one is based on an intelligent personal agent to assist in finding pages relevant to the behavior and characteristics of the user. Secondly, distributed knowledge base situated across the network will lighten the query-processing loads. Distributed knowledge bases on separate issues located at different servers can be treated as integrated domain knowledge. The integrated knowledge is then used in an inference process at the main server. In this paper, the technical issues on the above model will be addressed.

Keywords: Distributed knowledge base, Intelligent personal agent

1. Introduction

As evidenced from the federal initiatives in National Information Infrastructure (NII), there are consistent huge increases in the number of users of the Internet. The number has grown from 1 million to 25 million in the past decade [1]. The trend is expected to continue in the coming years. According to the result of a report published by the Institute of Science, the World Wide Web (WWW) is estimated to contain 320 million pages as of 1998. The number is continued to grow exponentially at a rate of doubling every few months [2]. With many more users and the extremely rapid growth of information on the WWW, centralized global search engines such as Alta Vista, Excite, Lycos, and HotBot require huge single site network and hardware resources to handle index construction and query processing loads [3.1]. Due to the high query processing load and the huge quantity of information on the Internet, search results may be of low quantity and outdated if limited single-site resources. The problems is also compound by the growth in the number of users on the Internet. Hence, the use of distributed systems seem like a logical solution to this problem.

One of the causes of the difficulty in searching on the WWW is the large amount of unstructured documents. It can be said that better search requires better organization. Thus, by using some forms of commonly agreed structure to organize the knowledge base, this will enable much more effective retrieval of the information from the Web. The system can also reap the benefits of a distributed system by having the inference process and the distributed knowledge bases treated as separate entities located on different network nodes. This behaves as a distributed expert system.

In the coming years, it is expected that information retrieval should be based on semantic rather than syntactic. The searches should be based on concepts rather than words [1]. For example, the focus of the Extensible Markup Language (XML) is on describing the content rather than presentation. This opens up new possibilities in using the language to provide more effective organization. However, XML is still only a syntactic and it is likely that many incompatible tag sets will evolve to describe similar concepts. In addition, XML promotes interoperability between organizations that agree on standards ahead of time but the divergent vocabularies will make it of little use for general searches on the Web [4]. Therefore, to perform more effective Web searches, a semantic markup language has been chosen to add markup to the Web pages.

On the other hand, personalization is one of the features for improving the Web searching performance. This is achieved by finding more relevant pages that relate to individual's behavior and preferences. Presumably, information retrieval will be more effective if the individual user's idiosyncrasies are taken in the form of an intelligent personal agent. This agent will acquire and refine its knowledge according to its environment and records of the user's behaviors and characteristics.

0-7803-7490-8/02/$17.00©2002 IEEE.

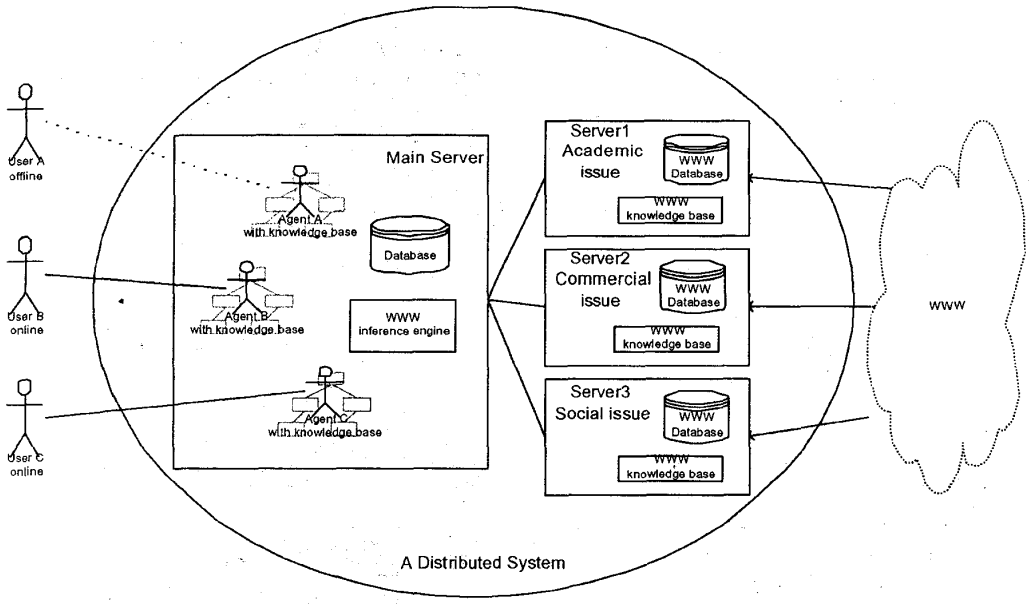

Figure 1: The System Architecture

At present, most of the search services just present the information to the user who makes the queries. The user then browses the results and the search session is considered completed. The results will also be discarded. A phase of ongoing searching while the user is offline is likely to improve the search results and keeping the search results up-to-date. The agent may continue to search in accordance to the user's specific queries irrespective of whether the user is online or offline. The results will be made available for the next search exercise, or, until the agent was instructed to stop searching.

In this paper, we present an approach of using distributed knowledge base. This incorporates an intelligent personal agent. The personal agent contains personal knowledge of the user. The objective of this proposal is to improve Internet searching under a distributed environment. We suppose that the Web pages are added using a semantic markup language thereby the information will be automatically categorized on specific issues or on dedicated servers.

The organisation of the paper is as follows: Section 2 describes the architecture of the system and Section 3 gives a brief overview of the prototype implementation. The conclusions and discussions are then summarized in Section 4.

2. Architecture

This paper presents a distributed system of Internet search engine. An example architecture of the system is shown in Figure 1.

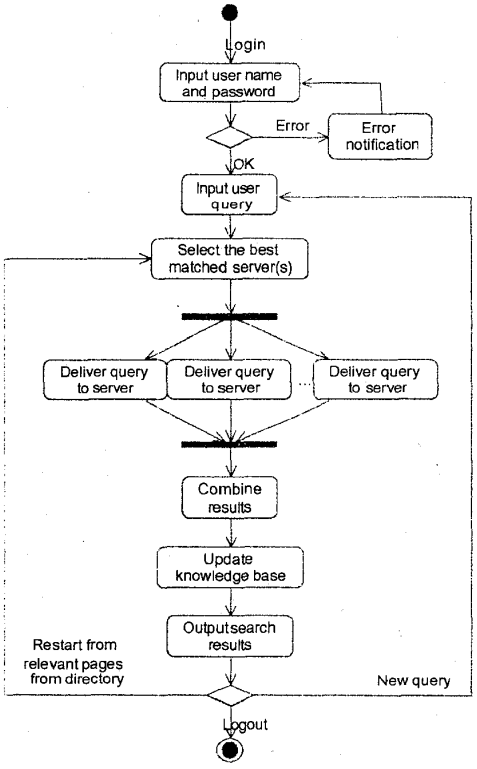

Figure 2: Foreground process of system

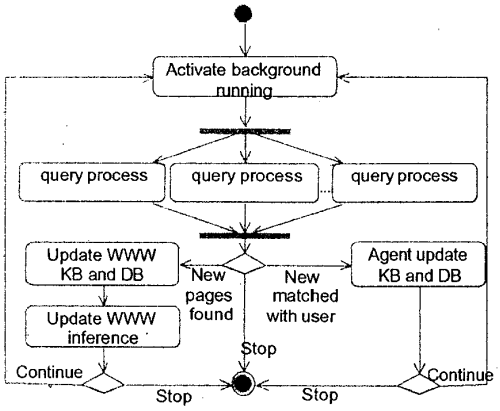

Figure 3: Background process of system

An online user can send a query to his/her own agent in the system. Attributes of the agents are corresponded to the profiles of the individual users. The query is then processed locally within the main server. The agent's query is then automatically delivered to the best-matched servers in the system. The server should hold information which are only related to the issues that the user is interested in. Each server is specialized in a particular issue. It is likely to have more than one server that overlaps the same topic area.

While the user is offline, the agent may continue to work in the background. This may be happening whenever the load of the server is low. For example, the agent might regularly check specified web sites or particular queries for any changes and modifications. When these occur, the user may either be notified if he/she is online or through the user's e-mail.

The activity diagrams for both foreground and background processes of the system are presented in Figure 2 and Figure 3 respectively. We suppose that the Web pages have been developed in a unified standard based on semantic markup language, or, the system could have been developed by the same organization.

For knowledge representation, a frame-based approach has been chosen. In the main server, agent approach is adopted in the inference process based on the knowledge base situated within the same machine. Unlike the normal approach in accessing information from the Web, domain knowledge acquired from separate knowledge bases on different issues are now integrated and being used in the inference process. This provides a centralized approach in the inference process yet the information sources are accessed in a distributed manner. This provides a more efficient search and more relevant results will be returned.

3. Implementation Details

The components of a prototype server implementation are briefly overviewed as follows.

The main server consists of intelligent personal agents, a database which keeps the users' information, and Web information inference. The agent contains a knowledge-base, it mediates between the individual user's query and the distributed set of available servers. The agent will process the query related to the issues that the user is interested in. The main objective of the agent is to link the user's query, Web inference and the user's idiosyncrasies from the database. This will enable the query to be delivered to the distributed servers related to specific knowledge sources. The agent's tasks also include the selection of an appropriate set of servers. These are the processes that agent is expected to perform when the user is online. When user is offline and subject to command from the user, the agent may be set to run in the background in order to satisfy the user's request if there are any. The Web inference can be considered as a form of global contents of the available servers. It assists the agent in the selection of a set of servers for query delivery.

Frame-based knowledge representation of agent is shown below in Figure 4.

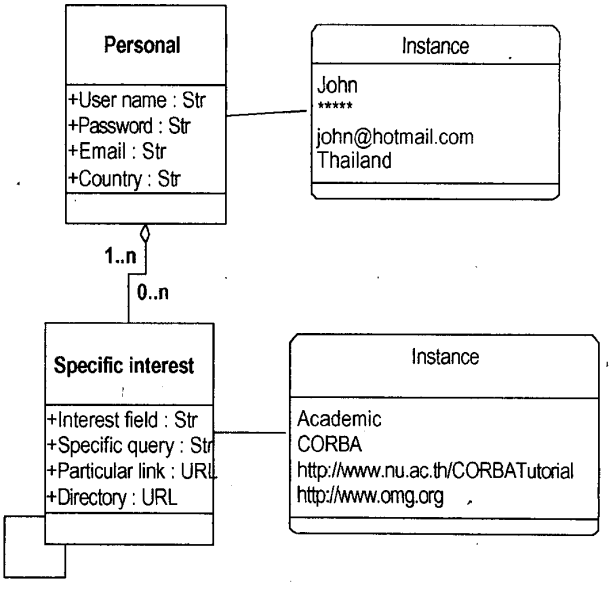

Figure 4: Frame-based knowledge representation of agent

Most of the fields shown in the example are self-explanatory. The *<particular link>* is for the agent to keep running the query in a particular link. The *<directory>* specifies the locations where relevant documents are stored. During the search process, the users can keep some documents from the search results if they are relevant. The *<directory>* is useful due to the agent uses it to restart the search that would help in finding the more relevant pages related to the history of relevant documents.

The distributed servers contain history profiles on particular issues. The issue might be for example academic issue, commercial issue, social culture issue, etc. It is likely to add more servers on the list as more issues are needed. The system adopts an open system architecture in order to support scalability characteristic as required by a distributed system.

4. Conclusion

A distributed environment for effective Internet search using intelligent personal agent and distributed knowledge base is presented. By using the agent, it is expected to find the most relevant pages relating to the behaviours and characteristics of the user. By using distributed system, there are many advantages. For example it will lighten the query-processing load and it facilitates system extension. Furthermore, the query-process is concise because it only processes the information on particular issue. The proposed work is currently being developed and further results will be reported in the future.

5. Acknowledgement

The authors would like to thank Naresuan University, Thailand for the support given to this research work.

References

[1] Bruce, R. S., 17 January 1997. *Information Retrieval in Digital Libraries: Bringing Search to the Net.* Bioinformatics Science, vol. 275, pp.327-334.

[2] Kaushik, P.R., 1999. *Personal Search Assistant: A Configurable Personal Meta Search Engine.* In proceeding of 5th Australian World Wide Web Conference, NSW, Australia.

[3] Koch, T., Ard, A., Bremmer, A., and Lundberg, S., 1998. *The building and maintenance of robot based internet search services: A review of current indexing and data collection methods.*

[4] Heflin, J., and Hendler, J., 2000. *Searching the web with shoe. In Artificial Intelligence for Web Search.* Papers from the AAAI Workshop. WS-00-01, pp.35-40. AAAI Press.

[5] Bessonov, M., Heuser, U., Nekrestyanov, I., Patel, A., April 27-29, 1999. *Open Architecture for Distributed Search Systems.* In proceeding of 6th International Conference on Intelligence in Services and Networks, Barcelona, Spain.

[6] Shang, Y., Shi, H., and Chen, S., 2001. *An Intelligent Distributed Environment for Active Learning.* Journal on Educational Resources In Computing.

[7] Maedche, A., Staab, S., Stojanovic, N., Studer, R., and Sure, Y., 2001. *SEmantic PortAL - The SEAL approach. to appear: In Creating the Semantic Web.* Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., (eds.) MIT Press, MA, Cambridge, 2001.

[8] Sukarno, M., J., and Lin, W., 1996. *Distributed Knowledge-Base: Adaptive Multi-Agents Approach.* In proceeding of the 1996 IEEE International Joint Symposia on Intelligence and Systems (IJSIS).

[9] Soshnikov, D., 1999. *An Approach for Creating Distributed Intelligent Systems.* In proceeding of the Workshop on Computer Science and Information Technologies CSIT'99, Moscow, Russia.