
Copyright © 2002 IEEE

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
A DISTRIBUTED ENVIRONMENT FOR EFFECTIVE INTERNET SEARCH USING INTELLIGENT PERSONAL AGENT AND DISTRIBUTED KNOWLEDGE BASE

Chun Che FUNG¹ Pornpit WONGTHONGTHAM²

¹School 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: thungc@cc.curtin.edu.au

²Mathematics 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.
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.
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.

![Frame-based knowledge representation of agent](image-url)
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 path. The <directory> specifies the locations where relevant documents are stored. During the search process, the list 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