INTELLIGENT AGENT BASED COLLABORATIVE CONSTRUCTION INFORMATION NETWORK

M. Sun, N. Bakis, I. Watson[1]

Abstract. In recent years, the World Wide Web (WWW) has emerged as the most popular vehicle for disseminating information through computer networks. More and more construction organisations are beginning to make use of Intranets and the Internet. While the amazing growth of the WWW is an indication of its strength, it also makes the task of finding the right information more difficult. Most of the current solutions such as information gateways and search engines are "single user mode" and "domain independent" which limit their effectiveness. This paper describes a knowledge-based Collaborative Construction Information Network (CCIN) which supports collaborative information searching and retrieval for construction related topics on Intranets and the Internet. The system is developed based on software agent technology. It consists of CCIN clients and CCIN servers configured in a three-layer architecture with the CCIN servers playing the intermediary role between the information users and information suppliers.

Keywords: Internet search engine, collaborative searching, information retrieval, intelligent agent, information network, user profile

INTRODUCTION

The construction industry is an information intensive economic sector. Many activities involve the creation and consumption of massive amount of information ranging from product data to technical publications, from building regulations to best practice guides. With the increased use of computers, much of the information is generated in digital form. There is a growing demand for exchanging construction information over computer networks electronically. In recent years, the World Wide Web (WWW) has emerged as the most popular vehicle for disseminating information on the Internet. It is estimated that the volume of information on the WWW is doubling every fifty days. The amazing growth of the WWW is an indication of its strength. However, as more and more information becomes available, the task of finding the right information becomes more and more difficult.

There are two common types of solutions to the information discovery and retrieval problem on the Internet. The first is a "yellow-pages-like information gateway" in which the on-line information is organised in logical categories. The user can locate the right information by browsing through the listings. The second type of solutions is the use of Internet search engines. A search engine is a special WWW server that gathers information about documents located on other WWW servers. The user can locate the required information by supplying keywords to the search engine which in turn finds the documents by matching these keywords. Both types of solutions enjoy a certain degree of success. However, they both have serious limitations.

By and large, the current WWW network can be described as a two-layer structure with direct interface between the information users and suppliers. Many people argue that it is no longer sufficient to manage the continuous growth of the WWW. A three layer structure (Daigle, 1996) is required consisting of users, intermediaries, and suppliers. The function of the intermediary layer is to match users' demand and providers' supply in the best possible way. To achieve this, the intermediary layer needs to know what is available on the Internet and also to have knowledge of the users' information needs, domain specific knowledge and the ability to

[1] M. Sun, Lecturer, School of Construction and Property Management, University of Salford, Salford, M7 9NU, UK, Email: m.sun@salford.ac.uk

N. Bakis, Research Associate, School of Construction and Property Management, University of Salford, Salford, M7 9NU, UK, Email: n.bakis@salford.ac.uk

I. Watson, Senior Lecturer, Department of Computer Science, University of Auckland, New Zealand, Email: ian@cs.auckland.ac.nz

learn. Although the existing search engines, e.g., Infoseek, Lycos, try to provide some of the intermediary functions, they do not yet offer any value-added services.

Given the shear size of the WWW, it is impractical to build a single system that combines all the information resources. A more realistic approach is to build specialised information retrieval systems that provide access to information resources of a particular domain, i.e., construction. In the longer term, these domain specific information systems can collaborate by sending requests to each other to provide a unified service to the users. This paper describes a Collaborative Construction Information Network (CCIN) project which addresses the issue of collaborative information search and retrieval for construction related topics using the World Wide Web (WWW) on Intranets and the Internet. Its key features include: (1) a one-stop-shop information search service for both Internet and Intranet users; (2) support for user collaboration; (3) application of construction domain and user profile knowledge to improve the information search.

Before the introduction of the proposed CCIN system, the current solutions for Internet information searching are reviewed.

REVIEW OF CURRENT SOLUTIONS

When a collection of information reaches a certain size, there is a need to find ways to organise the information and retrieve it when required. In the paper-based information age, card catalogues and classification systems like the Dewey Decimal System (Fowler, 1997), are effective information retrieval tools. Today, the amount of digital information is growing exponentially and much of it is available through the Internet. The World Wide Web has become an expanding hypermedia database where information in various formats can be found on many related and unrelated topics. The traditional methods alone are no longer able to manage this information growth. A large number of new tools have emerged aimed at helping users to locate information on the WWW. These tools can be broadly divided into two types:

- Yellow Pages type of information gateways.
- Robot based Internet search engines.

Information Gateways

The Yellow Pages is a well-established business directory publication. Using the Yellow Pages, the user can locate correspondence details of a firm by looking through the classification hierarchy. At the initial stage of the WWW development, surfing and browsing were the main navigation methods. Yellow Pages type of information gateways were popular, ranging from personal hotlists to comprehensive lists of international services covering multiple publication-types and subjects (Brümmer, 1997). Good examples of this type of information gateways include the WWW Virtual Library (Manning, 1999) and Yahoo (1999). There were also subject specific gateways, for example, the Construction Industry Gateway in the UK an initiative to provide a gateway service for construction specific information on the WWW (Lockley 1998).

There is no doubt about the usefulness of these systems. However they also have some major weaknesses. Each of these services often only covers a very small fraction of the WWW resources, only a couple of hundreds or thousands of WWW sites with Yahoo as an exception. There are no widely accepted standards for classification. Usually each information supplier decides the structure and content of its materials governed by chance, occasional decisions and staff responsible for the implementation. The consequence is a lack of consistency and reliability and a lack of independence from the individuals performing the task. Another weakness of information gateway solutions is the difficulty of keeping information

updated. Due to the rapid changing nature of the WWW, many resources are quickly becoming non-usable. Because there are no effective automatic updating mechanisms it is very difficult for the gateways to follow the swift changes of their contents, addresses, appearing and disappearing of documents on the sites they cover.

Internet Search Engines

The search engine solution is based on the principle of keyword matching. A user describes his/her information needs using a number of words. If these words are found in a document then there is a higher probability that this document is relevant to this particular user's information needs. Although in reality the keyword matching methods are much more complicated and sophisticated the principle remains the same. An Information Retrieval System (IRS) iterates through a document collection and builds an index of the most important words found in the collection. It associates each word with the locations of all the documents containing it. When a user specifies a query with a set of keywords, the system returns the documents containing all or some of the words in the query. Internet search engines are programs that roam the Internet (with flashy names like *spider, worm* or *searchbot*) to build up an index of meta-information about everything available on the net (Gilster, 1996). The gathered information, characterised by a number of keywords (references) and perhaps some supplementary information, is then put into a large database. The database is increasing all the time as more information becomes available on the net. Anyone who is searching for some kind of information on the Internet can then try to localise relevant documents by giving one or more query terms (keywords) to such a search engine. Since their first appearance in 1994, the number and diversity of Internet search engines have increased rapidly and continue to do so. The most popular existing Internet search engines include Infoseek, Excite, HotBot, Lycos, Alta Vista, etc. Each of these engines covers a percentage of the Internet. There are some overlaps between their coverage. There are now meta-search engines available, which pass the user’s queries to multiple search engines in order to perform searching in a wider scope.

These search engines are still providing a useful service in the face of information overload on the WWW. However, they have many weaknesses which become more and more evident (Hermans, 1996). All these search engines rely on the user being able to formulate queries effectively, an assumption that is not always valid. There is no good ranking mechanism to rank the many hundreds or even tens of thousands of *hits* in response to each user’s query. In addition, most search engines are domain independent and used in a single user mode.

LATEST DEVELOPMENT

In recent years, a number of techniques have emerged which are aimed at improving information retrieval on the WWW. Amongst these techniques are information filtering, collaborative browsing and information agents.

Information Filtering

The goal of an information filtering system is to sort through large volume of dynamically generated information and present to the user those that are likely to satisfy his or her information requirement (Newell, 1997). Information filtering was originated from the manual alerting services that brought new information to the attention of users of special interests. With the growth of the Internet and other networked information, research in automatic information filtering has been on the increase in recent years. Using a filtering system, a user does not need to search for information using queries. Instead, the user specifies his or her interests

and the system monitors the information sources to inform the user when information of the required nature becomes available.

Collaborative Browsing

Various studies show that when people search for information, the first thing they usually do is to ask other colleagues who have done similar searching before (Lockley, 1998). The existing browsing and searching systems focus on technical information discovery methods while neglecting perhaps the single most important method of discovery that people rely on - other people. Collaborative browsing, sometimes also know as social resource discovery, assumes the existence of other users of similar interests who have located and evaluated relevant resources. The goal of collaborative browsing systems is to aggregate and share the fruits of the individual activity and knowledge of Internet information retrieval (Twidale, 1998).

Intelligent Information Agents

Agent technology is originated from the branch of Artificial Intelligence known as Distributed Artificial Intelligence. In recent years, it has grown into a fast expanding research area. As any other new research field, there are diverse definitions of the term *agent*. Here we want to quote a very generic definition given by Janca (1995), "an agent is a software that knows how to do things that you could probably do yourself if you had time". There is a growing consensus in the Internet community that one of the most promising solutions to the problem of Internet information retrieval is the use of software agent technology.

The main feature of an agent based WWW is that information agents perform the role of managing, manipulating or collating information from many distributed sources (Nwana, 1996). Some information agents have intelligent features for example they can be mobile, can learn and can co-operate with other agents. Mobile agents are able to roam the WWW, interact with WWW servers, gather information on behalf of their owner and return home after performing duties set by their users. The learning ability of information agents refers to the fact that they can react and interact with external environment and improve their efficiency over time. While individual agent is able to work on its own, a number of agents can co-operate with each of them performing a role to achieve a collaborative goal.

Although there is still scepticism considering software agent is just another buzzword, the initial development in this area offers promising potentials. The benefits of software agents approach include:

- More intelligent information search
- More robust and flexible system
- Opportunity to incorporate domain knowledge
- Personalisation
- Better user collaboration.

COLLABORATIVE CONSTRUCTION INFORMATION NETWORK

The Collaborative Construction Information Network (CCIN) is an on-going project carried out at the University of Salford. The aim of the project is to develop a value-added gateway to the construction information network, which facilitates collaboration between users for the benefit of improved information search and retrieval on the WWW. To achieve this aim, the CCIN project seeks:

- to examine the information needs of different types of users in the construction industry;
- to capture these information needs conceptually as user profiles and information context models;

- to incorporate construction domain knowledge into the information network;
- to improve the speed and accuracy of search for construction information by developing an information network that facilitates the sharing of search results and knowledge;

It is acknowledged that the WWW based information network will remain a distributed resource with information hosted on many servers. The gateway of this project seeks to provide a recognised starting point for the user to search for information and a repository for shared knowledge base that helps the information search. Figure 1 illustrates the architecture of the Collaborative Construction Information Network. It shows a typical three layer architecture with the CCIN servers playing the intermediary role between the users (clients) and information suppliers (WWW servers).

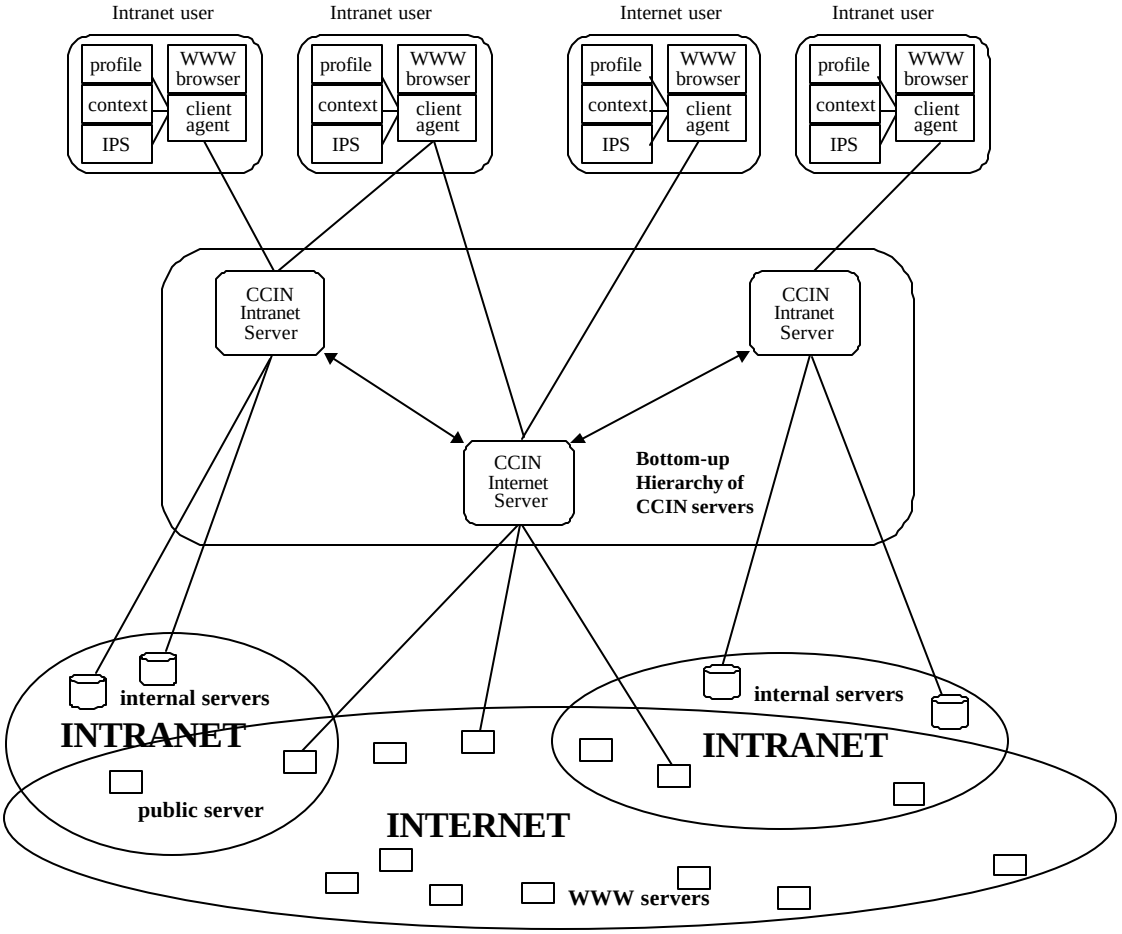

Figure 1: Collaborative Construction Information Network

The described information network consists of CCIN servers and client agents. The CCIN client agent is a plug-in package to standard WWW browsers, such as Microsoft Internet Explorer and Netscape, to provide additional functions. Once it is installed on a user's computer, the agent will facilitate the communication with the CCIN server and provide intelligent information search and retrieval services. To achieve this, the agent needs to gather information of user profile, information query context and information search history. Part of this information is communicated to the CCIN servers to improve the information search task of the user community as a whole. The CCIN server, a collection of CCIN servers, is in essence a construction oriented WWW search engine. It does

not host the original information documents. Its main function is to match information sources to users search queries. There is one central Internet CCIN server as a natural focal point for the Construction Information Network. Individual organisations can also install their own local Intranet CCIN servers. These servers are organised in a logical hierarchy and function seamlessly together to provide a transparent and one-stop-shop information search for both Intranet users within a company and the individual Internet users. The local CCIN server provides search for internal information whose circulation is confined within that particular company. The central CCIN server provides search for public accessible information sources.

CCIN SERVER

The CCIN server is a knowledge-based search engine for construction users. Its main function is to match information sources to users search queries. The key feature of the server is the framework that enables construction domain knowledge and user knowledge to be accumulated and used in intelligent information searching. The server consists of the following components (Figure 2).

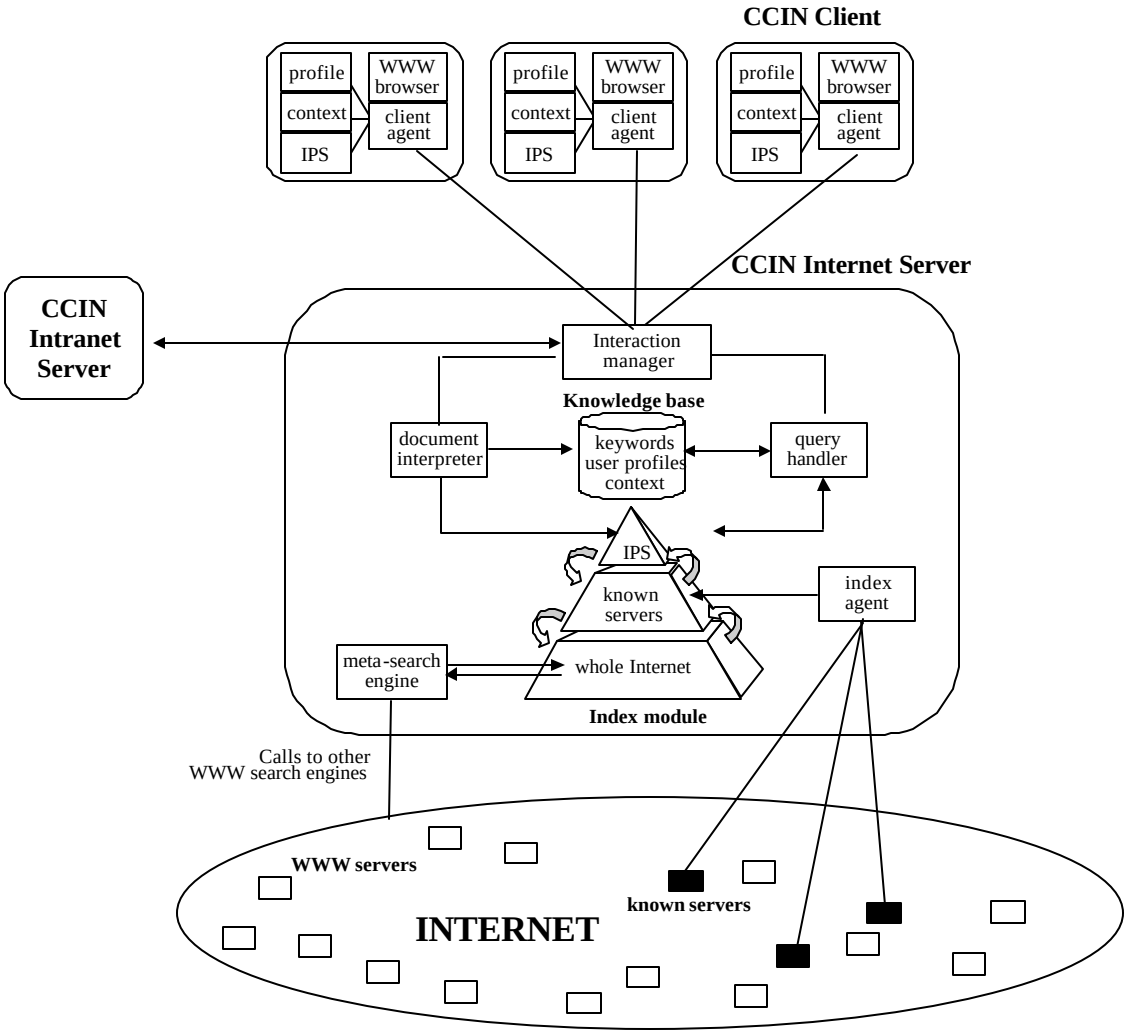

Figure 2: CCIN server and its interaction with client agents

Layered Index Module

At the centre of the CCIN gateway server is an index module that consists of three hierarchical layers. The top layer stores information of documents that have been accessed by users and communicated to the server through the CCIN client agents. The data are stored in a meta-information format similar to Jasper's Intelligent Page Store (IPS) (Davies, 1997). For each WWW page, the index module stores at least the following information gathered automatically by the client agents with a degree of user manual input:

- the document title
- a summary of the content
- a set of keywords
- a set of user profiles who recommended the page
- information type (product data, technical publication, news, project data, etc)
- users' annotations
- universal resource locator (URL)and,
- date and time of storage or update

The middle layer contains an index of documents located on known WWW servers which are more likely relevant to the construction users interests. These servers become known to the gateway server through two ways. (1) The information providers register their servers explicitly. (2) When the client agent submits a document to the top layer of the index module, the host server becomes known to the gateway implicitly. The CCIN server agent, a mobile information agent, visits all the known servers periodically and gathers information about documents hosted on these servers.

The bottom layer of the index module covers most of the accessible web hosts on the Internet. Given the constant growth of the Internet, the gateway server does not store any documents. Instead, it will rely on a Meta search engine to call upon other web search engines, e.g., Alta vista, Infoseek, etc.

The hierarchical layering implies that the amount of information accessible increases as one moves down from the top to the middle and the bottom layers, but the average potential relevance to the users' interests decreases. The advantage of the layering approach is that it allows a user to specify the scope and manner of a query. One can raise a query just for the IPS layer where more criteria can be applied apart from keywords. Alternatively, the user can make a general query using *enhanced keywords* supported by the server's knowledge base to a wider index in the middle layer or even the whole Internet.

A document's position in the *index module* is not fixed. There are migration paths through which a document can be moved from one layer to another as a result of users' search and retrieval actions. For example, when a document in the middle layer is retrieved by a user, the *client agent* will communicate that fact to the server. The server will upgrade the document to the top layer in the index module. On the other hand, if a document in the top layer has not been used by any user for a period of time (pre-defined threshold), the server will degrade it to the middle or even bottom layer. The purpose is to ensure the efficiency of the index system.

Query Handler

The task of the query handler is to handle more complex queries other than merely keyword search. The existing web searching relies only on keyword matching which often results in a large volume of documents with a high proportion not related to the user's interests. This project investigates more intelligent conversational query techniques. It aims at acquiring more knowledge about the user and the user's information needs so

that more accurate queries can be formulated. The system provides several searching methods of different levels of complexity catering for users with different levels of computer competence. It enables the end users to perform information search in a variety of ways such as using information concept, information types, user profiles, as well as the conventional keyword methods. To achieve this, the CCIN server needs to possess and accumulate knowledge about its users and construction domain specific information.

Knowledge Base

If the server knows more about a user's information needs, it can provide more accurate results for the user's search queries. The knowledge base of the CCIN server keeps information of three aspects that are construction domain specific.

Keywords: Keyword based searching remains an important method of information retrieval on the WWW. Instead of considering it as a simple syntax comparison, the CCIN server identifies a set of keywords and associates them with conceptual meanings. Relationships between these keywords are also analysed. As a result the server is able to support concept based searching. For example, when a user elect to search for information about "enclosing building element", he/she can choose to have extended the search to include *sub-types* "wall", "window", "door", "roof", or *association-types* "building material", "component product", etc. These keywords are identified based on recent data and processing modelling works of other research projects and the common classification standards used in the industry.

User profiles: A user profile articulates the features and the information needs of a distinctive type of users. The CCIN server has general profiles for the common types of construction information users, such as architects, engineers, project managers, researchers, academics, etc. The CCIN users can choose one or a combination of these profiles for their searching query. They can also derive personalised profiles based on these general ones.

Context: Context is a description of the purpose behind a user's search for information. It may be about stages of the construction process, design tasks or a research topic area. The CIG scoping study identified the information requirements of construction professionals during each stage of the construction project life cycle (PE Consulting, 1996). It can be used as a framework to organise the construction information query and retrieval.

The knowledge base is populated at the server set up stage. The server also has the ability to learn from its interactions with the client agents. The knowledge base is updated constantly.

Document Interpreter

When a user retrieves a document of interest, the client agent installed on the user's computer extracts some information about the document, such as the title and a set of keywords. This information is stored locally in an Intelligent Page Store (IPS) format. At the same time the client agent sends it to the CCIN server together with the user profile information and its context. The *Document Interpreter* on the server is responsible for processing this information and storing the document summary in the *index module* and updating the *knowledge base* if necessary.

Index Agent

This is a mobile information agent. Its function is to traverse the list of known WWW servers and build up an index for all the documents on these servers. It has the ability to handle different data formats, HTML

documents, postscript files, compressed files, databases, etc. In the case of interaction with databases on a third party server, minimum human involvement may be required for the interpretation of the data structures until industry wide data standards are universally adopted.

Meta-Search Engine

If a user wants to expand the scope of search to the whole Internet, the meta-search engine routes the query to other general purpose WWW search engines, such as Alta Vista, Infoseek, etc. In this event, the CCIN server still uses the local knowledge base in processing the search results so that the returned documents can be appropriately ranked according to their significance. In the last few years, several meta-search engines have emerged, one of them is the MetaCrawler (1999).

CCIN CLIENT

The CCIN client agent is implemented as a plug-in package to standard WWW browsers, such as Microsoft Internet Explorer and Netscape, to provide additional functions. It is available for free to the users. Its main purpose is to monitor the user's information retrieval activities and communicate with the CCIN servers with the explicit permission of the user.

User Profiler

When a user is connected to the CCIN server, he or she is presented with a collection of built-in user profiles characterising the main type of users related to the construction industry. Each profile is associated with a set of keywords and information context, and can be used as independent key for formulating queries. The user can select one or more standard profiles according to his/her interests. The user's profile is further personalised during the process of querying. The local computer stores the personal profile.

Context

There is always a reason behind a user's search for information. The reason could be performing a design task, writing building specifications, doing research on a particular topic, and so on. The *context* seeks to use a structured framework to capture knowledge of this aspect. It is another channel for collaborative information search. For example, a user can request for documents associated with a stage or related stages of the *context*. Similarly, the CCIN server has an initial built-in context model that will be modified with constant input from the client agents.

Intelligent Page Store (IPS)

This is a local repository facility for the end user providing enhanced storage which is more comprehensive than the existing WWW browser's bookmarks function. It helps the user to organise downloaded Internet documents for quick recall in the future. It adopts the same structure as the IPS layer of the CCIN server Index Module.

CONCLUSIONS AND FUTURE WORK

The emergence of the WWW as a global information network brings with it a number of challenges for helping non-technical users to retrieve information effectively. A number of characteristics of the WWW makes the information retrieval task difficult. The WWW is a very fragmented network with millions of information providers and even more users. It is very difficult to enforce any unified classification schemes over the WWW. The

contents of the WWW are constantly changing, new materials are being added and existing materials are being removed or modified. This paper reviewed the two current types of WWW information search and retrieval solutions, information gateways and search engines. As the WWW continues to expand and more and more computer naive users begin to rely on it for information, more effective information retrieval tools are required.

CCIN discussed in this paper is an on-going project. It takes advantage of the emerging technologies in information filtering, collaborative browsing and information agents. The main contribution of the project is the user collaboration aspect and the inclusion of construction domain knowledge and user profiles. For example, when a user supplies a keyword "window" to the CCIN server, the server has the knowledge to qualify the query as "window" in the building context not in the computing context as "Windows 95". The CCIN system does not seek to replace the existing information gateways and search engines. Instead it complements the current solutions and works together with them.

The system is still in its development stage. User testing will be conducted when it is fully implemented. Several non-technical issues will need to be addressed such as the user privacy, information ownership and generation of critical mass of users. Like other collaborative systems, CCIN relies on the co-operation of its users. It needs to monitor and gather information about the user's WWW query activities. Ways need to be found to overcome the common fear of intrusion on user privacy. In the construction domain, a large amount of information is owned by the traditional information providers, many of them offer commercial on-line information services. While CCIN does not need to dispute the information ownership with these providers, it needs to get access or to make interface arrangements with their indexing services. To achieve the benefits of the CCIN system, a sizeable number of users is required. It is a challenge at the initial phase to generate sufficient interests amongst the targeted users.

REFERENCES

Brümmer A., M. Day, D. Hiomet, T. Koch, M. Peereboom, A. Poulter and E. Worsfold, 1997, The role of classification schemes in Internet resource description and discovery, Research Report, NetLab, Lund University Library, Sweden
http://www.ub2.lu.se/desire/radar/reports/D3.2.3

Daigle L., P. Deutsch, B. Heelan, C. Alpaugh and M. Maclachlan, 1996, Uniform resource agents, October 1996, On-line WWW page, http://www.faqs.org/rfcs/rfc2016.html

Davies J., R. Weeks and M. Revett, 1997, *Jasper: Communicating information agents for WWW*, technical report, BT laboratories, http://www.labs.bt.com/projects/knowledge/jaspaper.htm

Fowler A., 1997, *Dewey Decimal System*, Children's Press Inc., USA

Gilster P., 1996, *Finding It On The Internet; The Internet Navigators Guide To Search Tools And Techniques*, John Wiley & Sons

Hermans B., 1996, *Intelligent software agents on the Internet*, Tilburg University, Tilberg, The Netherlands
http://www.hermans.org/agents

Janca, P., 1995, *Pragmatic application of information agents, BIS strategic decisions*, Nowell, United States

Lockley S. R., Amor R., 1998, The Construction Information Gateway, in R. Amor (ed) P*roceeding of ECPPM'98 Product and Process Modelling in the Building Industry*, BRE, UK, 337-348

Manning G., 1999, *About the Virtual Library*, On-line WWW page: http://vlib.org/AboutVL.html

MetaCrawler, 1999, *MetaCrawler Frequent Asked Questions (FAQ)*, On-line WWW page:
http://www.go2net.com/help/faq/

Newell S. C., 1997, User models and filtering agents for improved Internet information retrieval, *User Modeling And User-Adapted Interaction*, 1997, Vol.7, No.4, 239-256

Nwana H. S., 1996, Software agents: an overview, *Knowledge Engineering Review*, Vol.11 No.3 205-244, October/November 1996
http://www.labs.bt.com/project/agents/publish/papers/review1.htm

PE Consulting, 1996, *Scoping Study for the Construction Industry Information Service*, for Department of Environment, June 1996

Twidale M. B., Nichols D. M., 1998, Computer supported cooperative work in information search and retrieval, *Annual Review of Information Science and Technology*, 1998, Vol.33, 259-319

Yahoo, 1999, On-line WWW Page: http://www.yahoo.com