

# Singapore Management University Institutional Knowledge at Singapore Management University

---

Research Collection School Of Information Systems

School of Information Systems

---

1-1997

## Managing world wide web publications

Schubert FOO


*Nanyang Technological University*

Ee Peng LIM

*Singapore Management University*, [eplim@smu.edu.sg](mailto:eplim@smu.edu.sg)

**DOI:** <https://doi.org/10.1108/09685229710167980>

Follow this and additional works at: [https://ink.library.smu.edu.sg/sis\\_research](https://ink.library.smu.edu.sg/sis_research)

 Part of the [Databases and Information Systems Commons](#), and the [Numerical Analysis and Scientific Computing Commons](#)

---

### Citation

FOO, Schubert and LIM, Ee Peng. Managing world wide web publications. (1997). *Information Management and Computer Security*. 5, (1), 11-17. Research Collection School Of Information Systems.

**Available at:** [https://ink.library.smu.edu.sg/sis\\_research/156](https://ink.library.smu.edu.sg/sis_research/156)

This Journal Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [libIR@smu.edu.sg](mailto:libIR@smu.edu.sg).

# Managing World Wide Web publications

**Schubert Foo**

Division of Software Systems, School of Applied Science, Nanyang Technological University, Singapore

**Ee Peng Lim**

Division of Software Systems, School of Applied Science, Nanyang Technological University, Singapore

The surge of interest in the World Wide Web (WWW), with its potential commercial payoff, has resulted in an explosion of information as organizations join in the fray to publish and do business on the Internet. In addition, Intranet, the private corporate WWW-based network, has also emerged to provide cost-effective and efficient groupware and information management solutions for organizations. As a result, the HyperText Markup Language used for constructing WWW documents has become a contender for authoring future office documents. Examines the current practices of WWW publishing and highlights its inadequacies and drawbacks. Demonstrates the need and advantages of using a hypermedia database system to manage Web publications and briefly describes the design and prototyping of the D4W3 hypermedia database system.

## Introduction

Internet, the world's largest network inter-connecting networks across continents, is poised to be the information technology (IT) phenomenon of the 1990s, especially with the introduction of the World Wide Web (WWW) and commercial facilities. The popularity of Internet is derived from a number of contributory factors:

- *Easy accessibility.* Individuals and organizations with computing resources can gain easy access to Internet by acquiring an off-the-shelf computer, a low-cost modem and an account with a commercial Internet service provider (ISP) at competitive rates.
- *Ease of use.* The development of better human-computer interfaces and improvement in the level of computer literacy in many developing nations around the world has resulted in more people finding the Internet technology accessible and easy to use.
- *Global connectivity.* By virtue of Internet's global interconnectivity, all that is required is the knowledge of another user's "Internet address" before communication can take place between two parties from anywhere around the world. Participation in the World Wide Web and USENET Newsgroups (a worldwide distributed discussion system) guarantees a worldwide audience.
- *Speed of access to Internet facilities.* Access to facilities is almost instantaneous. An electronic mail message sent anywhere round the world reaches its destination within minutes. Downloading a freeware or shareware program from a file transfer protocol (ftp) site and getting it up and running is achievable in a short time. Likewise, browsing and searching for information at WWW sites can also be very fast.
- *Commercialization opportunities.* Most importantly, the advent of the WWW has transformed Internet into an environment which can support commercial interest and form business-oriented networks. From an organization's standpoint, the Internet offers a novel and attractive feature (unlike normal mass media tools such as radio, television or newspapers) to allow customers to "interact" with multimedia

information provided by the organization and, in some cases, to carry out business transactions on the spot.

- *Scale and distance independence.* Given the technical know-how of the technology, it takes little more effort and cost to reach out to thousands and even millions of users than to reach out to an individual. Additionally, the cost to reach someone at the other side of the world is typically the same for someone in the same country.
- *Low barriers to entry and equal opportunities.* As a result of the ease of access and an open and extremely democratic environment, equal opportunities exist for all in this business community. Large organizations investing more financial resources into making more impressive WWW pages may not necessarily exhibit a distinct advantage over an individual or a small team who may effectively focus on better communications and support for its niche-market customers.

## The impact of the Internet on businesses

Internet has indeed revolutionized business in that it redefines the methods used in traditional business practice and offers another important channel for mass communication which is likely to grow even further in future. Depending on the type and nature of the business, Internet provides a platform to carry out a host of potential business applications, such as:

- a public relations tool to establish a global presence and heighten public interest;
- a marketing tool to advertise goods and services, and to open up international markets;
- a marketplace to sell goods and services;
- an information kiosk to provide up-to-date business information, to answer frequently asked questions and to release time-sensitive information;
- an alternative support tool to answer customers' queries and solicit feedback;
- a research or information-gathering tool for market surveys, product launches, and

even to obtain solutions to problems encountered from around the world;

- a human resource tool for staffing and recruitment purposes;
- a support tool to serve mobile employees or telecommuters;
- a computer-supported co-operative work tool (CSCW) to facilitate groupwork and communication in an organization or across organizations, locally and globally.

The WWW in its current form has already made many such applications realizable. WWW-based business computing is already seen to be a new competitive business weapon. With the move towards the era of global information technology, regionalization and globalization of businesses, and the ongoing enhancement to Internet's technology (system and data security, encryption techniques, and communication protocols), the WWW is set to take off further in the future. It is predicted[1] that Internet commerce is likely to grow to over US\$100 billion by the end of the year 2000. In addition, the number of devices accessing the WWW will grow from 12.6 million worldwide in 1995 to 233.3 million in the year 2000. During this period, the number of Web users will jump from 16.1 million to 163 million and the percentage of "buyers" on the Web will grow from 24 per cent to 28 per cent.

In addition, Intranets are proving to be an exceptionally cost-effective way to distribute reports, track assets, improve employee communications, provide access to diverse corporate databases, distribute and run applications and enhance collaboration at a fraction of a cost of groupware solutions. An Intranet has the same look and feel as the Web. It basically uses the same architecture and standards of Internet and the Web. If necessary, Intranets can be closed off from the public Internet through software programs known as firewalls. Organizations thus have a choice to use Intranet as a closed system or to open part or all of it for public access.

While the Internet represents an opportunity for organizations to enhance their business profile and revenue, the Intranet represents the opportunity for organizations to enhance employee productivity. The extended Intranet (which denotes the partial opening of Intranet to the public) represents the opportunity for organizations to communicate with their customers and suppliers, and for those customers and suppliers to effect transactions with corporations. In fact, it is predicted[2] that the Intranet market alone will be worth US\$28 billion by 1999. With such staggering figures, Web technology and publications will become an important dimension

which almost no organizations can afford to ignore.

## World Wide Web publication

The World Wide Web is a wide area hypermedia information retrieval initiative aimed to give universal access to a large universe of documents. It is organized as a set of HyperText Transfer Protocol (HTTP)[3] servers designed specially for rapid distribution of hypermedia documents. Hypermedia uses a set of notes and links to organize information. This results in a hypergraph (or hyperstructure) that can be represented as a network as shown in Figure 1.

As hypermedia systems are often used to structure information contained in pre-existing documents, the same document can be referenced by different nodes. Thus, the hypermedia model allows the sharing of node contents. Users navigate around the hyperspace by moving from link to link, or by specifying direct links.

The HyperText Markup Language (HTML)[4] is a hypermedia language used to construct Web documents. It is designed to specify the logical organization and formatting of general text documents, with extensions to include inline images, audio, video clips, fill-in forms and hyperlinks to other HTML documents and other Internet resources. As a result, HTML is not only applicable to Web documents alone, but can be applied to the environment of office documents within an organization. It is anticipated that HTML will emerge as the main contender for office documentation and that future office documents will be authored using authoring tools developed for the WWW. Each WWW resource has a unique address known as the uniform resource locator (URL).

In order to publish and read information on the WWW, three basic components are required: access to a HTTP (Web) server, an HTML-authoring tool for composing Web documents, and a Web browser for navigating and reading the Web pages stored on various Web servers on the Internet.

Prior to publication, the intended information to be published must be made available and arranged in some structured hierarchical fashion. Information that is cross-referenced or related to other information is identified and linked. By carrying out this process, the hypergraph structure of Figure 1 will be realized. With this, it becomes an easy task to compose the Web documents and link them together using HTML conventions (i.e. anchors and URLs).

As information is contributed from various sources within the organization, there will generally be an overall administrator (or chief information officer) who is in charge of the overall project with various responsibility centres set up for actually defining and maintaining the information. Guidelines and rules of thumb are distributed to the various people involved to ensure homogeneity and conformity in presenting information. A schedule of regular updates and maintenance, and a system to handle proposed changes to existing information, is desirable to ensure overall data integrity and accuracy of information of the system.

### Current problems in WWW practices

Many organizations throughout various industries have entered into the Internet environment and started to use Web technology to publish on the WWW. Companies without computing resources will have to rely on commercial vendors to put up and maintain the information on their behalf. However, publishing in the WWW and using it in the present manner can lead to a number of problems in the future:

- *Maintainability of publication.* As organizations realize the ease of putting information on the WWW, there will be a tendency to put on more rather than less information which can result in the number of Web pages growing rapidly. Ultimately, it will reach a point where maintaining and

ensuring the accuracy of information becomes a difficult task.

- *Integrity of data.* The chances of information duplication or inaccuracy can arise as different people contribute to publishing. The integrity of data can be compromised as different authors are responsible for their information domain. Such information could be cross-linked to information provided by other authors without their knowledge. Hence, deleting documents will cause integrity problems with related documents. In addition, information updates can be a problem if such information is shared among different information domains. For example, a change in the pricing of goods or services may require corresponding changes to a number of Web pages which are maintained by different people. Any loss of data integrity of the system can have potentially severe consequences, especially if the system is used for direct business transactions.

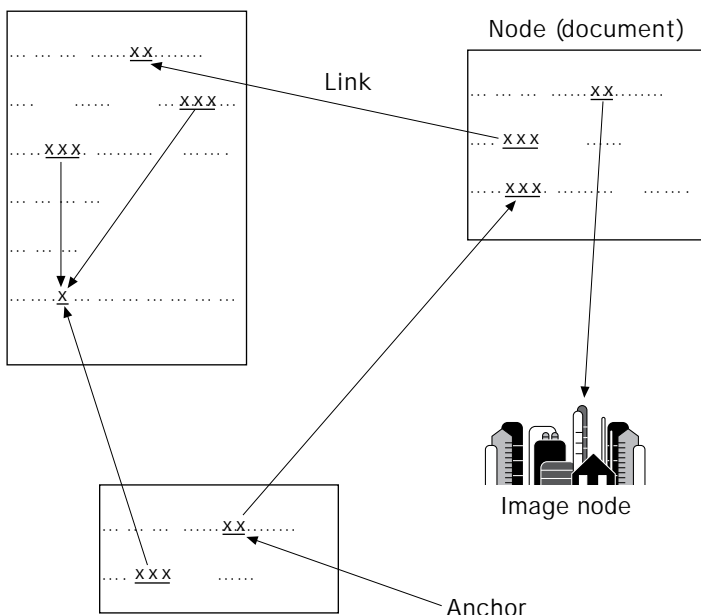
In order to ensure that the two problems are not present, there is a need for a structure, a change mechanism and manpower resource for the overall management of the system. Such a scenario may not be feasible in the case of organizations with limited financial and manpower resources. Additional costs are also incurred for having such a system in place.

- *Inadequate search facilities on the WWW.* There is currently no high-level query language for locating, filtering and presenting WWW information. Searching is achieved by using any of the many existing Internet search engines.

Search engines work in many different ways: some search titles of headers of documents, others search the documents themselves, and still others search other indexes or directories. Some are specially dedicated to the WWW (e.g. Lycos[5] and Infoseek Guide[6]), while others can also support searching of other Internet facilities such as USENET Newsgroup, ftp (file transfer protocol) – a file transfer facility for uploading and downloading files around the world – and Gopher (a globally searchable collection of menu-based hierarchical information resources). Examples of such search engines include Magellan[7] and Alta Vista[8]. However, searching in this manner usually results in a long list of matches (with possible duplicate matches if indexing is used) which must be further explored one at a time.

Furthermore, the home page of the document is usually presented so that navigating around the hypermedia space via hyperlinks is still required until the desired

**Figure 1**  
A hyperstructure or a hyperdocument



information is located. Such a form of searching is suitable for Net surfing to look up general information from various sources. However, it is inefficient, time-consuming and unsuitable for cases when a specific and detailed query is known beforehand.

- *Degradation of performance as number of users increase on WWW* As the number of users grows on the WWW, it will eventually overload the underlying network capabilities and result in a degradation of performance. Such observations are already apparent when accessing the more popular sites where network traffic is extremely high. Depending on the severity of the problem and expectations of users, this could lead to users' frustration and subsequent loss of interest in using the WWW. For organizations which have invested heavily in doing business on the Internet, it becomes a real threat which they can do nothing or little to solve.

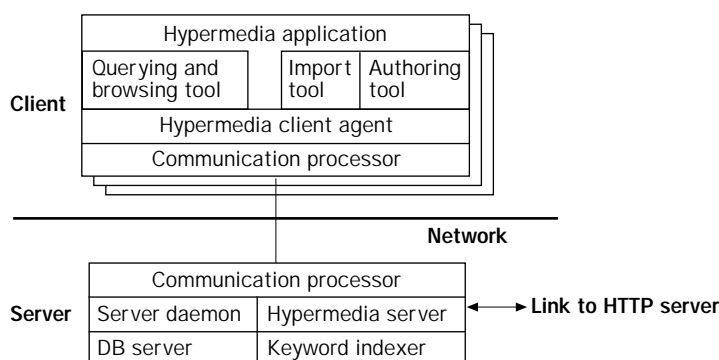
### Using a hypermedia database system for managing WWW publications

It should be apparent that there does exist a number of real concerns in current WWW practices which, if left unchecked, will lead on to even more severe problems in future. We propose that a hypermedia database system be used to manage HTML-based documents (including Web documents). In doing so, we will be able to resolve many of these cited problems.

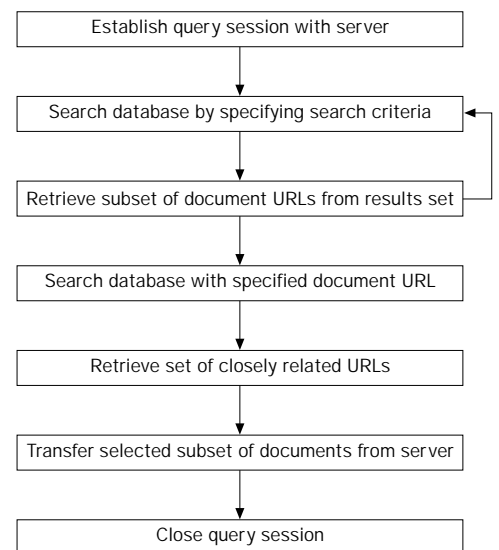
A general overview of the proposed hypermedia database architecture is shown in Figure 2. It utilizes a client-server architecture to support multiple users in a networked environment and comprises two main components:

- 1 *Client component.* This is the front-end of the system that comprises three main applications: an authoring tool, import tool and querying and browsing tool. Other hypermedia applications can be added in the future if necessary. The authoring tool facilitates the creation of new documents and modification of existing documents. Using the import tool, collections of hypermedia documents can be inserted into the hypermedia database in a batch mode. The querying and browsing tool allows users to formulate content-based searches and to view the hypermedia documents. At a lower level, the hypermedia client agent provides a common client interface to the hypermedia database server modules which reside on a server host. The communication processor on the client machine is responsible for sending and receiving messages to the hypermedia database server.
- 2 *Server component.* Like the clients, the hypermedia database server requires a communication processor to interpret incoming messages and to construct outgoing messages. To handle concurrent multiple client-server sessions, a server daemon is required to list to incoming server connection requests. For each client-server connection, the daemon creates a hypermedia server process which handles all queries from the client within the session established by the connection. To support flexible queries, some important attributes of hypermedia documents are extracted and stored in a database server. The hypermedia documents are

**Figure 2**  
Hypermedia database system architecture



**Figure 3**  
Query formulation steps



stored as files fully indexed using a keyword indexer. The server can be linked to an existing HTTP server so that any changes or additions of HTML documents in the server will be correspondingly reflected in the HTTP server.

To further organize the hypermedia documents in the system, a number of related documents are grouped together to form a publication. Every publication is owned by some user (owner) who is responsible for authoring and updating documents belonging to the publication. A list of owners and passwords is maintained by the system. This

ensures that only allowed authorized users have access to their own set of publications.

### Implementation issues

We have designed and implemented a prototype hypermedia database system, D4W3, using the proposed architecture for the management of HTML-documents. The authoring tool, import tool and server modules have been implemented on the SUN SparcStation running SunOS4. As most D4W3 users are expected to be PC users, a graphical query front-end has been developed on the IBM PC running Windows to allow users to perform query and browsing tasks. An extended relational database management system known as POSTGRES[9,10] is used for the database server.

In order to benefit and support a wide range of existing authorware which exists currently, a flexible authoring tool utilizing the “plug-and-play” paradigm is chosen. In doing so, users can use their own preferred authorware to create and modify HTML documents. Flexibility is achieved as users can switch and upgrade authorware without affecting the functionality of the overall authoring tool. For the same reason, the “plug-and-play” paradigm has been extended to the browser so that users can use their preferred browser to view the documents.

Users of the system can query hypermedia documents by specifying search criteria on selected attributes and keywords. The attributes that can be queried include author, title, subject and header. A complex search criterion can be constructed by a conjunction of multiple simple search criteria. To handle potentially large query result sets for any kind of search queries, the system is designed to keep the result sets temporarily at the server so that subsets of result sets can be returned to the client. Figure 3 illustrates the steps a client has to perform in order to search and retrieve documents from the server.

A query session has to be established before any query can be submitted to the server as shown in Figure 4. A user first specifies search criteria on the documents’ attributes and keywords as shown in Figure 5. Since every document is assigned a unique URL, the query result of the search request will be represented as a set of document URLs and attribute values as shown in Figure 6. The user can choose to perform one or more retrieval requests in order to obtain subsets of result from the server. Based on the summarized information about the documents which satisfy the search request, the user

Figure 4  
Client opening screen with request for server connection

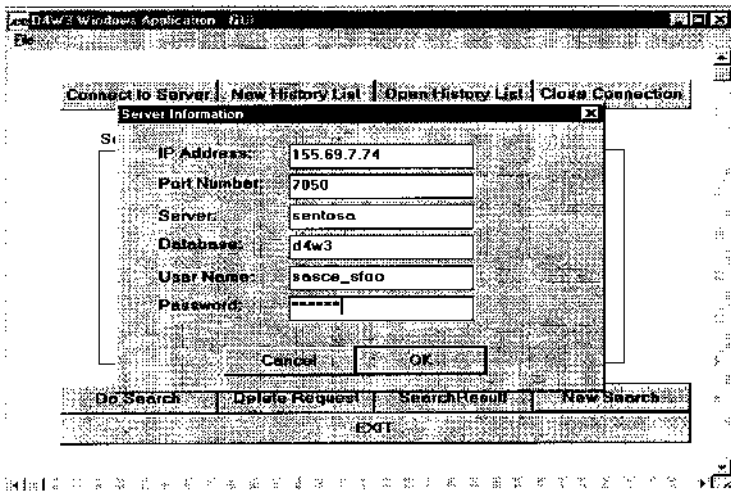
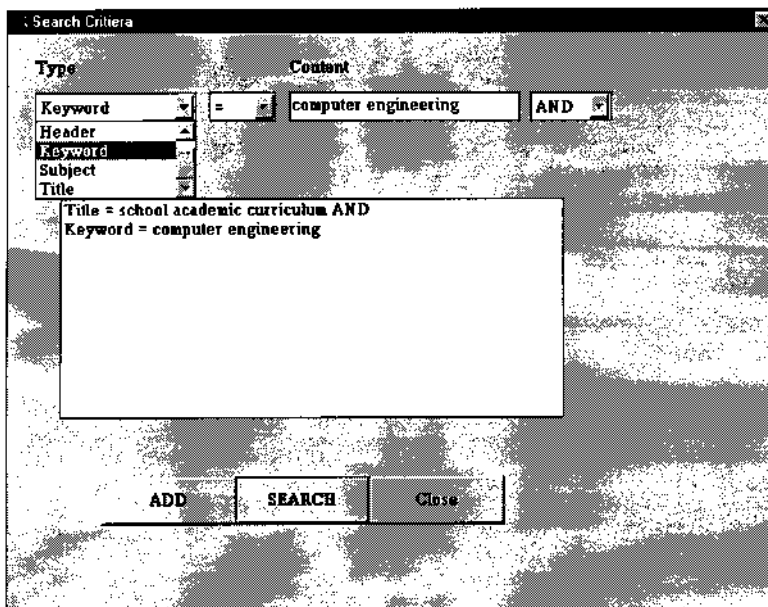


Figure 5  
Search screen for query formulation



may wish to browse some of these documents and continue to navigate to other documents. Thus, the user will highlight a particular document and selects the “retrieve document” option and specifies the number of levels of related documents which should be retrieved. This is accomplished by the system by performing a breadth-first search on a user-selected document, and retrieving its close-related documents. These are documents that can be reached by direct or indirect links. When all the documents are transferred to the client, the user can call up any Web browser to display the HTML document using the “load browser” option. The system uses Netscape’s Navigator[11] as the default

browser. This can be changed for use with any other Web browser. Figure 7 shows the retrieved document being displayed using Microsoft’s Internet explorer[12] browser. The user will use the normal navigation techniques to browse the document or go to other related documents. On exiting from the browser, the application returns to the query result display screen of Figure 6.

## System features and advantages

The use of a hypermedia database system for the management of WWW publications offers a distinct number of features and advantages over traditional WWW publication. In particular, it:

- provides an integrated system for the management of HTML documents from authoring to publication;
- allows users to log into the system directly to carry out query and retrieval operations. It offers an enhanced query and retrieval facility not normally found in WWW search engines. Users can prescribe the amount of search results to be brought back to the client machines, as well as the amount of browsing information. As the central repository of information is smaller, standalone and specific to the organization, it results in superior efficiency of searches;
- can be configured to act as a mini-Web site to store an exact copy of the organization’s publication on the WWW, thereby allowing remote users to log-in and visit the local site and use normal hyperlink navigation to search for information;
- enhances quality of service (e.g. speed of access), as the organization provides the direct computing resources over which it has full control;
- is easy to define an access control structure to control ownership of documents and to access the information domain within the information profile. Such a structure provides a framework to control the information resources of the organization;
- preserves data integrity while editing and updating documents. It eliminates the chances of deleting linked documents by mistake. The locking mechanism of the database system ensures document integrity during updates;
- ensures integrity of the system as the database is responsible for the management of the overall system. Each time information is updated on the system, it can be triggered to automatically update the same information in the actual Web and local sites;

Figure 6  
Query result display

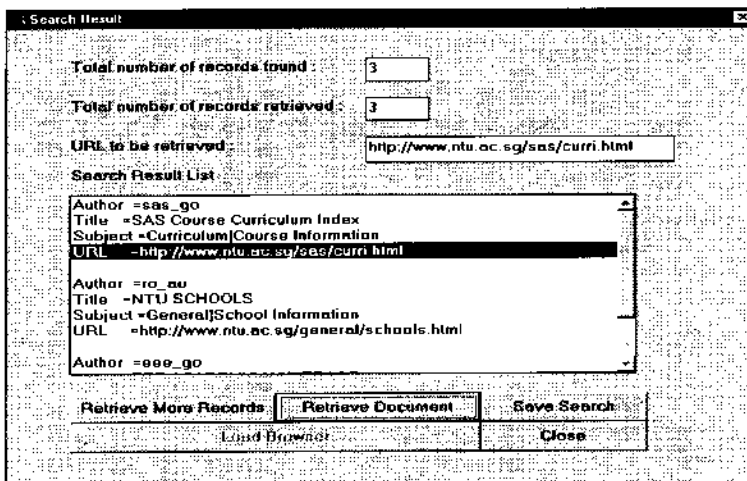
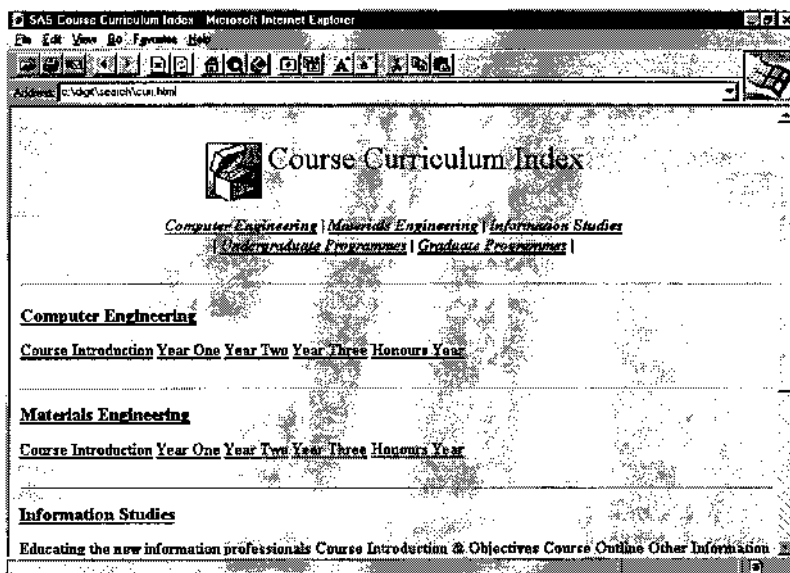


Figure 7  
Browsing the retrieved home document



- provides flexibility, configurability and upgradability as users select their own HTML authorware and browsers;
- minimizes duplicate information as the query engine is a convenient facility for checking prior to publication.

## Conclusion

This paper has demonstrated the clear need as well as the advantages in having a hypermedia database system to maintain and manage office or Web documents in the HTML format. A prototype system had been implemented to demonstrate the research idea. This system provides a structured and controlled means to publish, manage and query HTML documents which is currently lacking. The design is generic and employs a plug-and-play paradigm so that users can choose their own HTML editors and browsers for authoring and browsing.

---

## References

- 1 International Data Corporation, *Unprecedented Growth in Global Internet and World Wide Web Usage*, <URL: <http://www.idcresearch.com/HNR/cpr4ic.htm>>, 1997
- 2 Zona Research Inc., *Internet and Intranet: 1996 Markets, Opportunities, and Trends* (2nd ed.), <URL: <http://www.zonaresearch.com/>>, 1996
- 3 World Wide Web Consortium, *HyperText Transfer Protocol*. <URL:<http://www.w3.org/hypertext/WWW/Protocols>>
- 4 World Wide Web Consortium, *HyperText Markup Language*. <URL:<http://www.w3.org/hypertext/WWW/Markup>>
- 5 Lycos, Inc. & Carnegie Mellon University, *Lycos*, <URL:<http://lycos.cs.cmu.edu/>>
- 6 Infoseek Corporation, *Infoseek Guide*, <URL: <http://www.infoseek.com/Home>>
- 7 The McKinley Group Inc., *Magellon* <URL: <http://www.mckinley.com/>>
- 8 Digital Equipment Corporation, *Alta Vista*, <URL: <http://www.altavista.digital.com/>>
- 9 University of California, *The POSTGRES User Manual*, Berkeley, CA, 1994.
- 10 University of California, *The POSTGRES Reference Manual*, Berkeley, CA, 1994.
- 11 Netscape Communications, Inc., *Netscape Navigator 2.02 WWW Browser*, <URL: [http://home.netscape.com/comprod/products/navigator/version\\_2.0/index.html](http://home.netscape.com/comprod/products/navigator/version_2.0/index.html) >
- 12 Microsoft Corporation, *Internet Explorer*, <URL:<http://www.microsoft.com/ie/ie.html>>