89

AUTHORING AND INFORMATION SEEKING ON THE WORLD-WIDE WEB: AN EXPERIMENTAL STUDY

by

Stoyanka Atchkova Ilcheva

Submitted in fulfilment of the academic requirements
for the degree of
MASTER OF INFORMATION STUDIES
in the

Department of Information Studies
University of Natal

Pietermaritzburg
1997

ABSTRACT

The research reported in this thesis is focused on some of the implications for Library and Information Science of the development of the **Internet**, and the **World-Wide Web** (**WWW**) in particular. It highlights the need for a wider involvement of this profession in the better management, organisation, and ultimately - usability, of the global information system.

An extensive study of the information seeking process was carried out on a collection of interlinked hypertext documents constituting a PC-based model of the World-Wide Web. Special attention was paid to the correlation of WWW document authoring characteristics and the success rate in browsing sessions performed by novice, casual and experienced Internet end-users. The analysis of the results from a number of browsing sessions on the model led to some specific recommendations for comprehension-oriented WWW authoring.

PREFACE

The research described in this thesis was carried out in the Department of Information Studies, University of Natal, Pietermaritzburg, from January 1994 to June 1997 under the supervision first of Professor W. Horton and Dr D. Petkov, then of Professor A. Kaniki and Dr D. Petkov and since November 1996 of Dr D. Petkov alone. Part of the research formed the basis of an article titled:

> A UNIVERSITY ON THE WORLD-WIDE WEB: THE CASE STUDY OF THE UNIVERSITY OF NATAL, PIETERMARITZBURG (co-authored with Professor W. Horton)

The article was published by the **African Journal of Library, Archives and Information Science** in October 1995.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any university. Where use has been made of the work of others it is duly acknowledged in the text.

ACKNOWLEDGEMENTS

I am greatly indebted to Professor W. Horton for his untiring, inspiring, enthusiastic support and assistance at all stages of the project. Without that I would never have completed this thesis. Many thanks too to my supervisor Dr D. Petkov for his helpful suggestions, guidance, patience and understanding throughout the research process.

I am grateful to all who willingly sacrificed of their time to participate in the experimental part of this research.

Thanks to Roy Tennant, manager of WEB4LIB, and the other members of this electronic discussion forum for generously allowing me to use their contributions for my studies.

CONTENTS

List of Appendixes

List of Figures and Tables

List of Appendixes

List of Figures and Tables

Introduction

The old problem of matching resources to ever-growing needs in providing, disseminating, and retrieving information today is aggravated by the progressively heavier information overload brought on by the Information Age. Just as growing needs press for an increased availability of resources, so the expansion and upgrading of information resources brings about greater, deeper, and more refined information needs. The electronic revolution, leading to the exponential spread of the planet's "network of networks" - the Internet, has created the basis for a much speedier and efficient information dissemination and communication. This medium has also the potential to provide an easier and more democratic way of sharing the most fundamental resource to present-day societies - global, up-to-date information.

The Internet, and its most popular information delivery system, the World-Wide Web (WWW), is gaining importance in the efforts to solve the host of economic, social and educational problems of developing countries and make them competitive in the knowledge society created by globalization. The South African Department of Arts, Culture, Science and Technology (1996: 13; 9) has embarked on a policy which regards science, engineering and technology as "absolutely vital components of economic and social progress" and "central to the empowerment of all citizens". As scientific publishing is gravitating to electronic archives and digital libraries, it is evident that access to the Internet is necessary for scientific research and development to flourish.

However, building the infrastructure to provide nationwide access to the Internet

in South Africa is no more than a good start. Of paramount importance is a matching increase in the number of users and the level of sophistication with which they approach the world-wide electronic communication system. The majority of South African information workers and educators, as well as end-users need help to progress from a simple use of the e-mail facility to tapping into multimedia resources and global databases and benefitting from all the advantages of finding information and learning with the World-Wide Web.

How the information and training needs of the day are approached does not determine only the present, but sets the trends for the future as well. Technology cannot fulfil its potential without competent intermediaries to make it work for the end-user. Teachers, educators and librarians should be in the forefront of the communities driving the transformation and development process in South Africa.

Introducing the latest technology in libraries will open the door to a wider unimpeded access to a quick, up-to-date and exhaustive reference service and the material basis for the modern methods of teaching and learning. The Internet, reinforced by the WWW, is a great equalizer and an effective distance education and independent discovery tool. Teachers, librarians, educators and trainers who can make full use of all the services on the Internet, will pave the way towards redress and integration in education, the bridging of the gap between the urban and rural communities and economic growth and competitiveness for South Africa.

The World-Wide Web blurs the boundaries between writers, readers, publishers and learners by turning all parties into equally active participants in the communication process. In this new information environment a large number of

problems have evolved, all of them related to the effects of information overload. The solution of these problems will eventually lead to a considerably more effective form of information organisation and exchange.

The studies in this field have a short history but even at this stage they reflect a broad spectrum of research perspectives. As will be shown in the literature review of this thesis, Library and Information Science (LIS) has its own contribution to make to the concerted effort in improving communication on the WWW. One way in which this can be achieved is by highlighting the effect of the new technology on the information interaction activities of the user.

This thesis deals with some aspects of the author-reader interactions on the World-Wide Web. The particular focus is on the interrelationship between document authoring characteristics and the user's comprehension in information seeking on the Web. Several variables inherent in this relationship are monitored and analysed in the results and conclusions from an experimental study on a PC-based model of the WWW.

The structure of a Web document, and especially the way it is segmented and linked, internally and externally, has a marked impact on the meaning conveyed by the document to the user. It is crucial that the information seeker is always kept well-oriented and provided with the necessary context for proper micro and macro information retrieval. This will contribute to the ease of choosing and traversing meaningful and economic paths through the multi-dimensional, complex WWW environment.

Therefore, the solution of the above-mentioned problems of authoring style is

decisive for the successful author-user communication. It is this structural aspect of WWW authoring which is of paramount significance to information intermediaries and end-users whose prime need is to relate other authors' content units into coherent entities.

The problem of the usability of the WWW is interdisciplinary in nature and for its solution research methods are borrowed from computer science, human-computer interaction, cognitive psychology, library and information science, communication, education. In particular, the research presented here is to a large degree based on simulations on a model of the WWW.

The research problem and objectives

The broad aim of the research is to reveal the importance of local and global structural coherence of documents as a factor facilitating the effectiveness of information seeking on the WWW, thus improving the usability of the system and enhancing access to its resources.

The objectives of the research are:

- To study the new information environment of the WWW with a view to identifying its relevance to and effects on the theory and practice of Library and Information Science. This is pursued through the following sub-objectives:

  1. Investigate the impact of the Internet on the information needs and practices of library and information workers and end-users.
  2. Explore the advantages of the WWW over the other global information

delivery systems.

3. Focus on WWW problem areas that LIS specialists are best equipped to tackle.

- Establish the degree to which the length and modularity of hypertext documents on the one hand, and the configuration of internal and external hypertext links on the other, influence the outcome of browsing for information on the WWW. To achieve this, the following sub-objectives were identified as appropriate:

  1. Create a functioning model of the WWW, henceforth referred to as **MicroWeb**, which reflects the salient characteristics of the WWW.

  2. Incorporate various WWW real authors' document structuring and organisation patterns into an architecture of the model which would allow for the testing of the research problem.

  3. Formulate suitable tasks for the browsing simulations.

  4. Select suitable criteria and techniques for collecting and recording data from the observation of the simulations.

  5. Analyse the results and compare the conclusions to guidelines recommended in WWW authoring style manuals and the findings of browsing strategies research.

Scope and delimitations of the research

Modelling is a particularly suitable approach for analysing a new and complex system and discovering the relationships between its components. Yet, unlike the popularity of using the WWW for other subject-related simulations, no research has been found by this author on modelling the WWW-mediated communication

process.

Information seeking, which is a part of the author-reader communication process, has been extensively studied in other information systems. However, all of those have been closed, single-author systems, unlike the WWW which is an open, dynamic and multi-author one. Studies of WWW browsing and navigation exist, but all that have been found are in the field of human-computer interactions (HCI), which centres on the interactions of the information seeker with the computer and is aimed at improving interface support.

In contrast, the approach adopted here puts the human, rather than the technology, at the centre of attention at all stages of the research process. A justification for this can be found in the collaborative nature of the WWW environment in which the human author's input can no longer be regarded as a constant in the information seeking equation. Therefore the **MicroWeb** simulations were designed to examine the dynamic interdependence between the communication context created by intermediaries and general users in their capacity as authors and their readers' information seeking outcomes.

While the construction of **MicroWeb** was based on the major principles established in hypertext system design, it did not borrow or adapt other methodologies because the diversity of presentation and authoring styles is an essential feature of the WWW. Since the model was created as a tool to examine relationships in the whole WWW system as it is, and not in a single-author, single application context as it should be, the preservation of the above-mentioned multi-author diversity was mandatory. Therefore, a bottom-up design approach was adopted to allow the incorporation of other authors' documents complete

with their original local and global organizational structure.

It was also necessary to go through a genuine WWW intermediary authoring process to be able to compare the effects of authoring decisions on information seekers' comprehension and preferred browsing strategies. Therefore, the study could not make full use of existing data representation guidelines which are mostly interface oriented and set the stage for insights based on the relationships between user interface level moves, such as selection of anchors, backtracking, open URL, Hotlist - Go to, Forward, Open Local File, Home Document, Window History.

Because of its different focus, this thesis investigates only the two types of navigation-related user events of hyperlinks selection and backtracking, established by previous research as accounting for over 90% of all navigational activity. These are studied, however, at a level of fine granularity and in connection with other outcome measures, such as time, users' inter-documment and across-document browsing behaviour patterns and searching success rates.

Importance of the research

The literature review in this thesis seeks to cover the fundamental implications of the global electronic information revolution on the LIS world. Having a clear picture of these implications is a prerequisite to the redefinition of the role of libraries and librarians, which is crucial for their survival in the context of immediate accessibility of information on the Internet and other intermediaries' competition. The current debate on how to transform the theory and practice of LIS, reflected in the review, also maps out the directions for the future

development of research in this discipline.

As an original research tool, the **MicroWeb** model can make an important contribution. To the best of this author's knowledge, it is the first model created to investigate the WWW-mediated communication process in its systemic entirety of input, process, output, and feedback.

A large number of the authoring principles underlying the print publishing world, as well as lessons from information-seeking in pre-Internet electronic environments are relevant to the WWW as well. However, radical changes are introduced by the new communication medium and the non-sequential organisation of information by hypertext. A number of WWW style guides have appeared trying to capture the implications of these changes and build on the experience accumulated in print publishing in the new environment. It is necessary to test how the implementation of the various recommendations in these guides affects the effectiveness and efficiency of users' browsing strategies.

The holistic approach of the model constructed for this study creates the unique opportunity for direct feedback from information seekers that can be compared to the way the Web space author's intended purposes and meanings, as well as his/her perception of other authors' intended purposes and meanings, are interpreted and acted upon by the user.

The model, which is available on the Internet, can be used by other researchers to replicate or modify the study of the full chain of the human WWW-mediated communication process. The simulations have pointed to correlations between user interface events and behaviour patterns which cannot be monitored

automatically, thus laying the ground for log-file analysis of the problem under study in this research. This is conducive to the improvement of both expert and end-user authoring skills and information-seeking strategies. The accumulation of research results in this area will contribute towards the development of generally accepted WWW authoring standards. In this way the homogeneity, seamlessness and thus - accessibility of the information resources on the WWW will be increased.

There is also a practical spin-off from the work done for this thesis: the **MicroWeb** is valuable in itself as an introduction to the understanding and practical skills necessary for digital librarianship. It can be accessed at **http://zorba.phy.unp.ac.za/microweb.htm**.

Research methodology

An extensive literature survey was conducted on the trends accompanying the emergence of library sites on the WWW and of digital libraries. Traditional LIS, as well as innovative and radical approaches to the arising issues were compared. The need for librarians to focus on improving all aspects of communication on the WWW was identified.

Communication in a non-sequential and interactive information system such as the WWW is an extremely complex process influenced by a great number of various factors at all of its stages. In particular, for the study of the author-reader interaction aspects of the communication process, the researcher has two alternative options in the treatment of the author component. The researcher can adopt a passive role with regard to the WWW authoring process and concentrate

on the user-information interactions. This approach has two advantages: firstly, the research can be performed on the actual WWW, and secondly, the data-gathering process can be automated. The use of log files to record user navigation patterns can generate massive data and convincing statistics. However, it also introduces a substantial number of uncertainties and erroneous assumptions concerning unmonitored human behaviour, since the researcher has no control on the contextual changes in this collaborative author environment.

Alternatively, the researcher can decide to take over some of the authoring functions, thus assuming control over the shifts in Web documents' contexts related to the various authors' communication purposes and changing hypertext structures.

The present study needed to take into account factors which are of great significance to the information seeking process and which in principle cannot be reflected in a log file. These factors are related to the user's comprehension of the meanings conveyed and contexts created by the authors on the one hand, and searching strategies on the other. They are reflected in user behavioural patterns, such as the manner in which documents are read - are they scrolled up and down or read systematically, or skimmed, or scanned; the way the user looks and sounds - confident, purposeful or lost and disoriented; the preferred approach to selecting links - analytical, looking for clues to match the information need, or systematic, or serendipitous.

Therefore the method of computer-aided simulations on a model was the best choice for the investigation of this process for the following reasons: the method is powerful enough to provide insights into the behaviour of complex systems; it

allows the controlling or elimination of extraneous factors. It was decided to record manually all the details of the browsing sessions, thus obtaining information that cannot be obtained from a log file.

The application of the computer simulations method requires the existence of a functional model of the real-life system. For the needs of this study it was imperative to place the model in an environment that would allow a detailed monitoring of the information seeking process.

There are two very important conditions which must be fulfilled so that the results from a simulation experiment would be relevant, reliable and reproducible. First, it must be ensured that all the relevant characteristics of a simulation session can be monitored accurately and in full detail. And second, it must be ensured that all simulation sessions are performed under the same conditions.

In order to meet these two methodological requirements it was decided that a functional downscaled model of the actual WWW must be constructed and installed on a personal computer. Most of the model's interlinked hypertext documents were selected from the living World-Wide Web. The collection was organized in such a way that it not only reflects the characteristic features of the Web but it also facilitates the detailed monitoring of the effect of modularity and structural document characteristics on the success of browsing sessions.

The PC-based model makes it possible to control to the necessary extent the choice of paths, so that the effect of structural document characteristics on the information seeking process can be monitored in detail and compared, thus meeting the first methodological requirement mentioned above. Furthermore, the

model eliminates differences in document downloading times, delays, dead links and server failures. The removal of these disturbances, typical of an on-line environment, meets the second methodological requirement of conducting the simulations under the same conditions.

Layout of the thesis

The first two chapters of this thesis review the history of the global information infrastructure from its emergence to its present state. Special attention is devoted to the creation of the World-Wide Web. The relevant terminology is explained. The overview part of the thesis focuses on the effects of the World-Wide Web on library and information science and the development of digital libraries.

The main features and organization of **MicroWeb** - the PC-based model of the WWW, used for the computer simulations of information seeking, are described in Chapter 3. The results from these simulations are analysed and from these conclusions are drawn in the form of concrete recommendations.

Due to the nature of the research area, the greater part of the sources referred to in the thesis exist or are accessible only in digitised format and are entered in the List of References with their corresponding universal resource locators (URLs). Lastly, eleven figures, three tables and eleven appendixes are included in the thesis.

Note: Quotations in the main text are delimited by inverted commas. Quotations in paragraphs are italicized.

1

The Global Information System and the World-Wide Web

1.0 Introduction

A revolution often compared to the one initiated by Gutenberg's printing press is taking place in the world today. To understand the magnitude and significance of its far-reaching impact, it is necessary to look into the social needs and technological progress that caused it to happen. This is a broad subject with too many aspects, widely discussed from various perspectives, to be examined in any detail for the purposes of this study. Nonetheless, a brief account of the events which led to the establishment of a global computer network, the Internet, is in order.

Similarly, a historical background to the emergence of the World-Wide Web (WWW) is also essential to reveal the nature and interplay of the factors behind break-through advances in communications and the improved systems for the access to, organization, retrieval and utilization of information produced in their wake. This process culminating in the synergy of the Internet and the World-

Wide Web is among the forces driving mankind to the threshold of a new era.

Although the World-Wide Web came into existence as recently as the beginning of the 1990's, it has its roots in more than 50 years of scientific and technological development preceding its invention. An attempt will be made to capture the highlights - from a Library and Information Science (LIS) point of view - from the wealth of this background research and from the new ideas implemented in the WWW.

A vast new world like the WWW generates its own rich store of terminology. Some of it goes beyond the LIS scope envisaged in this project and some will be discussed in greater detail later. This chapter will provide a summary of the WWW's salient features and key concepts which are of paramount importance to the library profession because of their embodying the essence of the World-Wide Web as a radically new and transforming information provision system. The chapter will end with an outline of the WWW's strengths.

1.1 The Internet - a Revolutionary Communication Medium

There is a general agreement that the current emergence of the global information infrastructure is marking a new era in the same way that the invention of print technology did in the 15th century. Even if not directly involved in the Internet, everyone is affected by the sweeping changes brought to life by its ascendance. As Thompson (1994: 34) points out, the introduction of new means for information delivery and communication also leads to the establishment of "new forms of interaction and new kinds of social relationships between individuals". Melody (1994) highlights the interdependence between the effectiveness of the

prevailing information flows and communication patterns on the one hand and the pervasive spread of knowledge throughout society on the other.

It was the thirst for knowledge during the Renaissance that spurred the search for a technology capable of rapidly producing cheap and multiple copies of the scarce existing information record and coping with the growing information output at that time. The advent of moveable type thus opened the door to a broad dissemination of information with profound social, political, economic and cultural implications (Keys 1995). As a result, over the following five centuries the production of information was stimulated to such an extent that eventually it became too fast and copious to be effectively used and managed in the old ways. A large-scale search then began for new means of communication and information exchange. The birth of the digital networked universe today is part of mankind's efforts to conquer information overload (Melody 1994).

According to statistics from the Matrix Information and Directory Services (1996) the Internet's intercontinental electronic grid has grown from its 1969 modest beginnings of 4 mainframe computers in the USA, interconnected in the ARPANET (Chamberlain and Mitchell 1996), to the present 13.4 million individual servers in more than 100 countries around the world. The growth rate of networked computers is illustrated on **Fig. 1-1**.

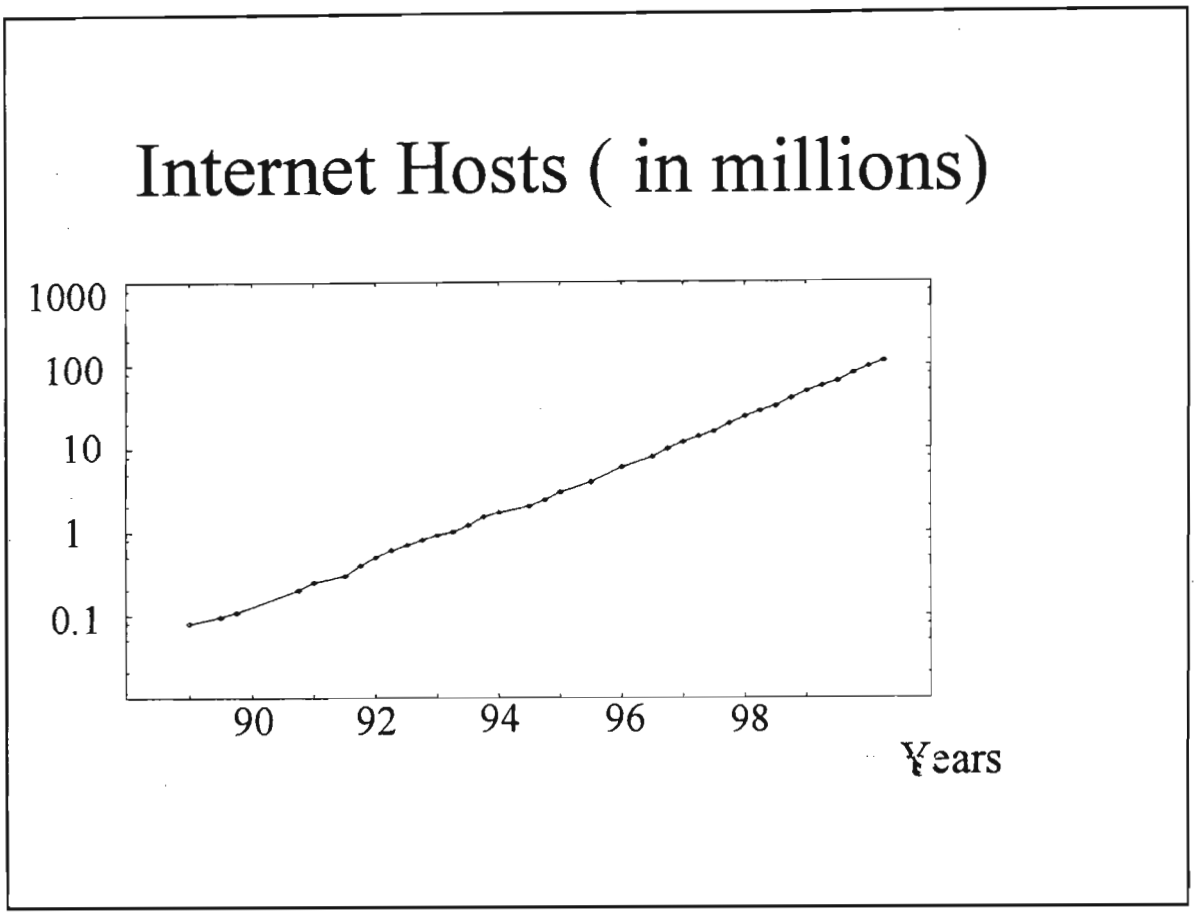


Fig. 1-1 Historical and extrapolated growth of the number of computers (hosts) connected to the Internet: 1989 - 2001 .

The values of the data (taken from Rutkowski 1996) plotted on the vertical axis of **Fig. 1-1** are the logarithms of the actual numbers. The fact that the points of the graph which represent the actual numbers of networked computers in the period 1989 - 1997 fall very closely to a single straight line on the logarithmic scale plot means that the growth of networked computers is exponential. The straight line drawn for the period after 1997 and well into the 21st century is an extrapolation of this exponential function. Far from being accidental, this proliferation is the logical result of the never-ceasing quest of mankind for more effective communication media: from speech - to writing - to printing - to

electronic networking.

A myriad of networks joined together to allow people all over the world to talk to each other and exchange data through their computers, the Internet combines all the achievements of the communication media known before and adds some radically new ones to them. The Internet can be:

- As fast as speech and the telephone
- As reliable as writing in storing and spreading the record
- Much more accessible, unrestricted, indestructible and cheaper than printing and teletyping
- As multi-sensory, powerful and effective, but far more interactive than the other electronic media.

Thompson (1994: 35 - 37) distinguishes between three types of interaction: "face-to-face interaction", "mediated interaction" and "mediated quasi-interaction" depending on the degree of spatio-temporal closeness and active involvement of the participants in the communication process. Face-to-face is the most immediate, both spatially and temporally, but also the most constraining of the three types of interaction. Mediated interaction adds the new element of an external carrier that makes the message transportable over space and time. Publishing and broadcasting industrialize the process of communication by their ability to simultaneously reach large numbers of recipients. At first glance their message might seem like a "monologue" addressed to a faceless multitude,

suggesting the above-mentioned classification as "quasi-interaction", but in actual fact it is prepared on the basis of carefully studied responses from well defined target audiences. However, because the "mass" media have managed to come across distance and time so close to their recipients, they have stimulated the need for a more active and immediate role for the latter as well.

The Internet has made this possible. Users are no longer mere recipients of information, they are participants in the information sharing and provision process. The Internet is the latest, and most sophisticated of McLuhan's (1964) "extensions of man", augmenting not only all the senses, but man's cognitive processes as well. It empowers people to form teams and groups of common interest in a co-operative and unifying international environment.

1.2 The History of the Internet

During the 1950's, when the cold war was at its height, the US Department of Defence embarked on a multiple-route highway building campaign with the intention of providing alternate roads for the army in the event of a military attack and the destruction of any of the major cities in that country. When this was completed in the late sixties, the Department set itself a similar goal of creating an indestructible, no-single-point-of-failure system for the national transport of data. For this purpose, the predecessor of the Internet, the ARPANET, was launched in 1969 by the Department's specially set-up Advanced Research Projects Agency (ARPA), to serve as an experimental base for its research into networking (Chamberlain and Mitchell 1996). The inevitable association of the interconnected electronic network with the highway network metaphor which led to the coining of the term "the information superhighway" in 1971 (Fain 1995)

has persisted. Ultimately, it was turned into the watchword of the day after US Vice President Gore's "Information Superhighway" speech in 1991 (Chamberlain and Mitchell 1996).

From the very start, the ARPANET implemented the idea of packet switching under which data sent across networks do not travel in whole units along lines dedicated solely to them. What happens is that different units of information coming from various sources are broken up into small pieces, called packets, and the whole bundle of packets is carried from router to router (computers doing the job of postal substations) until each one of them reaches its destination. This allows users to share connections, which makes the operation of the network cheaper. It also ensures that if any part of the network fails or is destroyed the remainder would still manage to do the job (Krol 1994: 13, 24 - 26).

Research on the ARPANET focused initially on the user's ability to log in to a remote computer and pass commands to it as if it were on his desk (this programme was later called **telnet**) and the improvement of the software rules that govern the operation of the Internet - the data communication protocols. However, the immediate by-product benefit of developing electronic mail (e-mail) for research collaboration support in "one-to-one (personal) and one-to-many (broadcast) communication" (Huston-Somerville and Kreitz 1995), was soon to make a major impact (Bowman *et al* 1994, Krol 1994: 25).

Computer specialists at universities and other research centres in the USA had steadily been getting connected to the ARPANET and in 1973 it started its extension to the country's allies and overseas military bases. In the same year another important step was made in improving collaboration support by the

introduction of the file transfer protocol (ftp), which enables the user to access remote databases and instantly download or upload files of any size and type. Six years later USENET, the world-wide Unix users' network carrying e-mail and newsgroups (messages by people discussing topics of common interest), was established (Segal 1995).

Similar in their intent to the USENET newsgroups are electronic discussion lists (or mail lists, or listservs) and bulletin board systems (BBS): the users can tap into "the human resources that are available - the people who are often specialists in a particular field of interest" (Benson 1995: 272).

In 1982 the Transmission Control Protocol (TCP) and Internet Protocol (IP) were established as the protocol suite for ARPANET. This led to "one of the first definitions of an 'internet' as a connected set of networks, specifically those using TCP/IP, and 'Internet' as connected TCP/IP internets" (Zakon c1993-6).

The appearance of desktop workstations in 1983 gave a boost to the development of local area networks (LANs). These workstations employed the UNIX operating system which includes the Internet Protocol and created the opportunity for LANs to be connected to the Internet (Krol 1994: 14).

In 1986 another US government agency, the National Science Foundation (NSF), built the academic network backbone to which regional academic institutions' networks from all over the country as well as other national and international networks were gradually linked.

By 1990 the ARPANET was retired (Chamberlain and Mitchell 1996) and the

Internet had transformed itself into a powerful means for democratic access to information and sharing of ideas. Not surprisingly, within one short year, a whole host of resource discovery and information delivery tools was born to meet the needs for effective collaborative work from afar of the increasingly specialized research communities, scattered over the planet. Among these tools were Archie - a locator of files on anonymous ftp archive sites, WAIS - a wide-area searcher of information in indexed databases, Gopher - a menu-based system for the exploration of networked resources, and the World-Wide Web - the synthesis of all Internet-based resources and utilities (Zakon c1993-6).

Another watershed year in the history of the information superhighway was 1994. The US government expanded its "planned **National Research and Education Network** into an all-inclusive information policy spanning all parts of society: industry, research, education and the public at large" (Voss 1994). This made it necessary that the problem of the Internet's financial viability should be addressed. The NSF, which in the 1986-1994 period had been chiefly responsible for funding the net-backbone, had enforced an "acceptable-use policy" restricting the commercial use of the Internet. A Commercial Internet Exchange (CIX) was set up in 1991 but it failed to attract much traffic. It was in 1994, when a new Internet architecture defining the areas for the different role players went into effect that the scene was set for a fully-fledged commercial provision and thus availability of the Internet to the commercial and public sectors (Ricart 1994).

Irrespective of fears of rising costs to the users, the large-scale introduction of commercial provision has triggered a dramatic rise in Internet growth rates since 1994 (see **Fig. 1- 2**). This graph represents the same data as the one on **Fig. 1-1**, only this time on a linear-scale plot which makes it easier to observe the

characteristic features of the Internet growth, namely:

- The time dependence of the growth rate is indeed exponential
- The period of doubling the number of networked computers is approximately one year
- The onset of the sharp rise in the number of newly-connected hosts happened about 1994

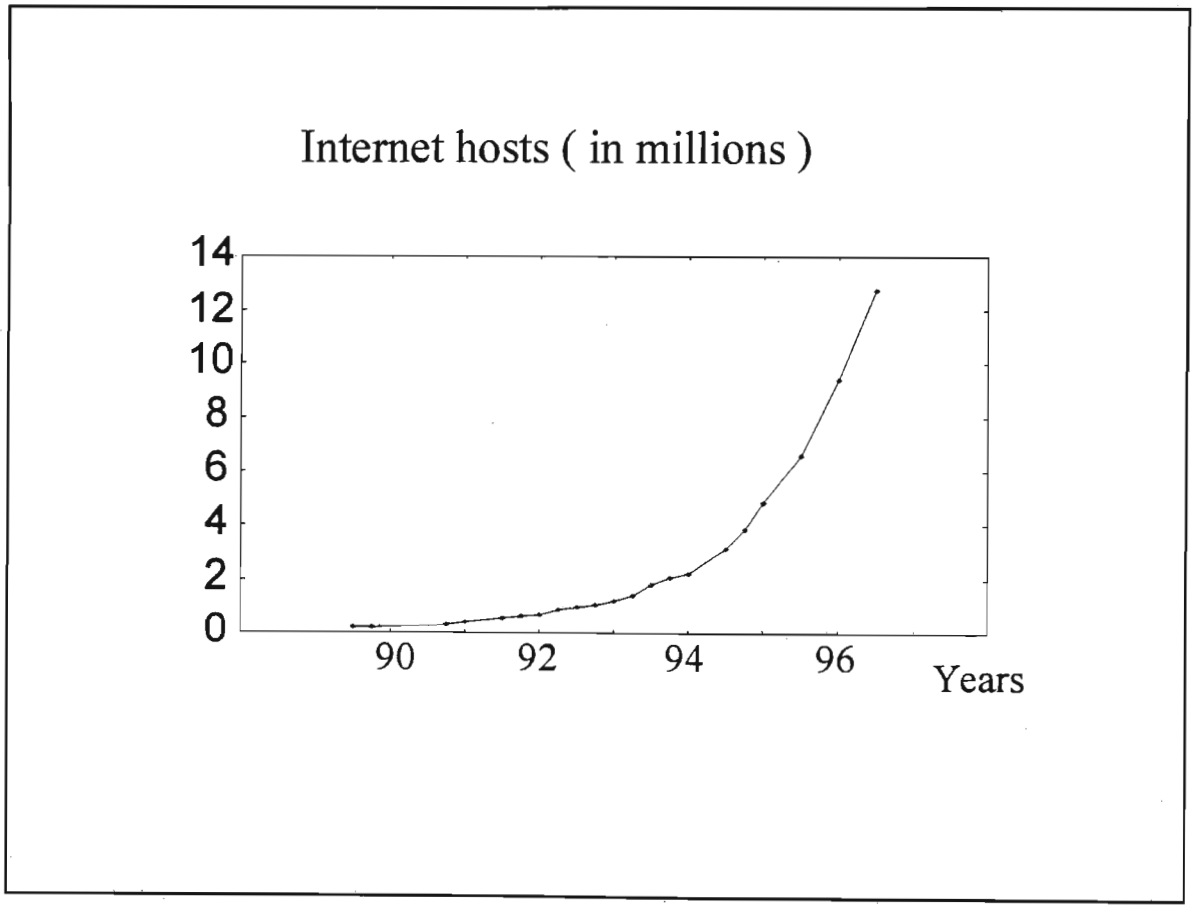


Fig. 1- 2 A linear scale plot of the rise in the number of computer hosts connected to the Internet: 1990 - 1996.

Another fresh driving force behind the Internet's new vigour was Europe (Handley and Crowcroft 1994-5). The Internet's development there followed a similar course to what had happened in the USA. The first pan-European backbone connecting European academic and research networks (EARN) to each other as well as to other networks on the planet was set up in 1985. The European academic networks also receive government funding and enforce a policy of acceptable use, which necessitates the expansion of commercial services to cater for the industry, business and the general public (Smith 1994: 3-4, Ewers 1994: 5).

South Africa has followed the network diffusion model typical of the advanced industrial countries by supporting the creation of a research and academic network to form a national backbone and provide gateways to other nations. Rhodes University has even helped less developed countries, such as Botswana, Lesotho, Mauritius, Mozambique, Namibia, Swaziland, Zambia and Zimbabwe establish connections to this backbone. Many networks have sprung up even in the poorest countries. Where telephone lines are lacking, connections are made by short wave radio transmissions. Fidonet allows grass-root network messages to be polled from a site in another country (Goodman *et al.* 1994).

Once the global computer communication infrastructure was established and functioning satisfactorily, the focus of attention had to shift to its effective and efficient use. It was only a matter of time before a truly universal, all-embracing, unified, versatile and easy to access and use system for finding and retrieving Internet resources was developed. Such a system by definition had to cover all the major previously established computer-based information systems, and had to allow for growth and improvement to meet the challenges of the

twenty first century. The sought-for system had to utilise all the possibilities offered by a modern state-of-the-art computer and the existing software, as well as to be flexible and adaptive. The system answering all these requirements proved to be the World-Wide Web, also known by its acronyms WWW or W3.

1.3 The History of the World-Wide Web

Among the leading institutions that pioneered the transition to the Internet protocols in Europe and thus to a global standardisation as a means to solve the problems created by heterogenous connectivity was CERN, the renowned international centre for nuclear research in Geneva, Switzerland. The adoption of the Internet there was the liberating force that was needed to enable another revolutionary vision - that of hypertext/hypermedia to live up to its promise in the birth of the World-Wide Web (Segal 1995).

1.3.1 The Hypertext/Hypermedia Vision

The idea of hypertext/hypermedia was first formulated by V. Bush in his seminal article **As We May Think** (1945). A mechanized device by the name of "memex" (from ***mem***ory ***ex***tender system) was proposed there. It was an attempt to improve the individual's use of the overwhelming amount of scientific publication by equipping the scholar's print-bound knowledge repository and workplace, the desk, with other media channels. Envisaged among the latter were pictures and photography, voice talking directly into the record, compact storage on film. This testifies to Bush's realization, later expounded by media theory, that a change in media brings about a change in the ratios of involvement of the human senses and, consequently, a change in culture (McLuhan 1962: 40 - 44).

The medium of writing, for instance, follows a linear progression dictated by the sequential sense of sight. Consequently, it slows down the grasp of the overall context because it does not allow the recipient to follow the multiple processes happening simultaneously in his environment, as is the case in oral communication (Meyrowitz 1994: 56).

But the most important of the memex's functions was to have been the ability to physically join information items that were in some way associated with each other. This would have eliminated the need of storing the association link in the researcher's memory, as well as accelerated and improved access to the relevant record. Bush believed that difficulties in getting at the scientific record were caused by the artificiality of the traditionally imposed systems of indexing under which data are filed alphabetically or numerically. Therefore he advocated the transition to associative indexing "whereby any item may be caused at will to select immediately and automatically another", thus mimicking the operation of the human mind:

> *With one item in its grasp, [the human mind] snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain.*

The concept of hypertext/hypermedia and the nature of hypertext/hypermedia as a non-constraining representation of information will be discussed in detail in **1.4** after the completion of the following account of the history of the WWW.

1.3.2 The World-Wide Web Project

Bush's vision captured the imagination of the following generations of scientists and developers who saw the computer as a suitable tool for the "augmentation of man's intellect". In 1968 D. Engelbart translated the concept of associative linking into practice by producing the first hypertext system, the NLS/Augment, and T. Nelson clad it in the term of "hypertext". Nelson had also started work at that time on Xanadu, an ambitious hypertext project (Conklin 1987). His goal was to create a "docuverse" (document universe) which would preserve the whole of mankind's literary output forever and make it available to everybody (Balasubramanyan 1995).

Hypertext implementation had to wait, however, for nearly twenty more years before it really blossomed out. Only after the substantial reduction in the size and cost of computers, parallelled by a massive increase in their power and sophistication, did it become the focus of brisk research that saw in it the basis of a soon-to-be-attained "infrastructure, national and international, that supports a network and community of knowledge linking together myriad types of information for an enormous variety of audiences" (Meyrowitz 1989).

Bush was trying to solve the information overload that after the Second World War was beginning to be felt in the personal records of the individual scholar. Half a century later, when the information overload had become a major barrier in the overall process of scholarly communication, Bush's anticipation of the crucial role of multimedia and associative organization of information was proved right by the synergy of the unimpeded global computer communication supported by the Internet and the free flowing, knowledge extending, associative

hypermedia of the World-Wide Web.

Proposed in 1989 by T. Berners-Lee and R. Cailliau and implemented in 1990, the WWW was the only Internet-based information delivery system using hypertext. It was designed to serve the information and collaboration needs of large international research teams working at CERN and other laboratories in the world, and it did not gain outside popularity at once. Even the hypertext research community, which had come to the realization that the development of standard linking protocols and a universal system, rather than separate applications, was essential to the survival of hypermedia (Meyrowitz 1989), did not immediately hail it as the answer they had been looking for. Furthermore, despite the fact that the WWW client software (the Line-Mode Browser) was released by CERN for free use in 1992, it was Gopher, which was simpler to install, that spread more rapidly in that year (Cailliau 1995).

Work at the National Centre for Supercomputing Applications (NCSA) in the USA reversed this trend in 1993. McCool and Andreessen wrote a server application (httpd) which was compatible with the wide-spread UNIX machines and the client application Mosaic which was easy to install and supported in-line graphics, as well as WAIS, Gopher, FTP and Telnet access. From then on rapid and consistent progress was made in developing software for various platform, interface and functionality needs (Morgan 1994).

Thus the WWW made accessing all the services and resources of the Internet a simple matter, as it did also publishing thereon, and thus immediately attracted a wide representation of both the specialists and the general public in the world to it. The sudden rise in the World-Wide Web's popularity has acted as another

reinforcing factor behind the Internet's accelerated spread since 1994. It is the Web traffic, Gillies (1995) notes, that now drives the Internet's growth.

In 1994 an international industry organization, the World Wide Web Consortium (W3C), was founded to promote the technological standardization and advancement of the Web. It is jointly hosted by the Massachusetts Institute of Technology Laboratory for Computer Science in the United States and the French Institut National de Recherche en Informatique et en Automatique [INRIA] in Europe. As of September 1996, the Keio University Shonan Fujisawa Campus, representing Japanese and Korean interests has also joined in hosting the W3C to "provide valuable, resident expertise and strong connections with Pacific Rim industry, both critical to the international development of the Web" (Vezza 1996). The director of the Consortium, the inventor of the WWW T. Berners-Lee (1996), describes the public services that it has to offer:

- *A repository of information about the World Wide Web for developers and users, especially specifications about the Web;*
- *A reference code implementation to embody and promote standards*
- *Various prototype and sample applications to demonstrate use of new technology*

All software produced by or officially contributed to the W3C is available free for general public use, commercial or otherwise.

Since April 1995 the WWW has become the most popular service on the global

information superhighway (see **Fig. 1- 3,** which also uses data from Rutkowski 1996).

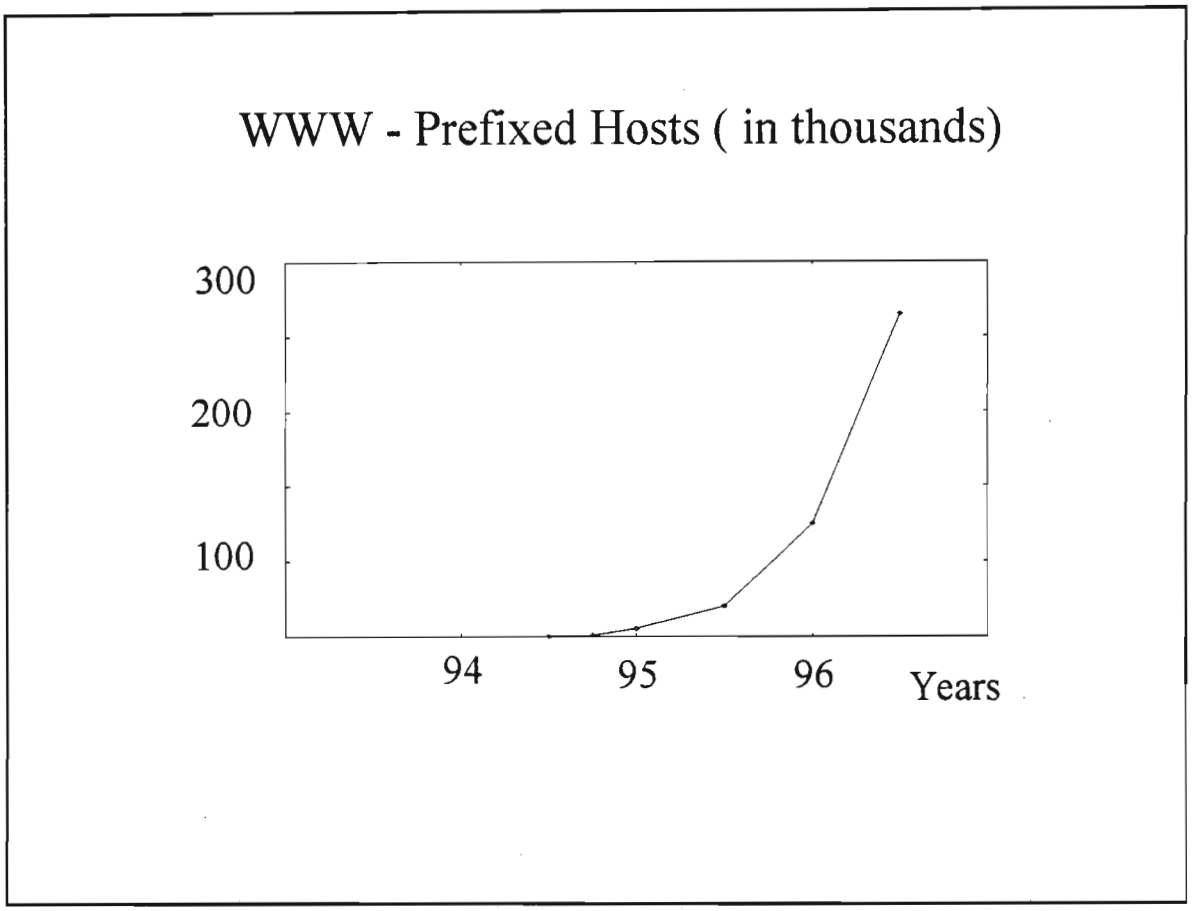


Fig. 1-3 Growth of WWW computers (hosts) connected to the Internet 1993 - 1996

The WWW has continued growing at the same breath-taking pace as more and more professions are finding it useful:

- Educationists are tapping into its potential to promote collaborative and distance learning

- Businessmen are reaching for broader markets

- The medical profession is sharing expertise through distant diagnosing and surgery

- Publishers and broadcasters are trying to conquer the dangerous rival of this new all-engulfing distribution system before it has conquered them.

The historical background of the Internet and the WWW, laid so far, will facilitate the understanding of the salient features and key concepts of the World-Wide Web with which non-computer specialists need to come to grips in order to play a successful role as information providers and intermediaries in the networked electronic age.

1.4 A Summary of the Salient Features and Key Concepts of the World-Wide Web

Librarians do not need as profound a technical knowledge as computer specialists to achieve their goals in an electronic information environment. After developing a general understanding, they need to master only certain technological features of the information system to an extent that would enable them to effectively operate it, as well as to assess the technology effects on the user's interactions with the system. However, in the case of the WWW, librarians have to be able to create their own documents in order to provide guidance and links to Internet resources of interest to their patrons. Therefore they need to learn the language of the WWW - HTML, and other authoring skills. In the summary, following now, the salient features of the WWW, which are considered necessary for a librarian's general understanding of the system, will be described and analysed while clarifying the key concepts involved in the

process. Authoring issues will be examined in Chapter 3.

The World-Wide Web is an open, boundless, seamless, dynamic hypermedia system for the creation, storage, navigation, retrieval and sharing of intricately interwoven global electronic information resources and communication interactions in the technological and social context of world-wide computer networked communication.

1.4.1 An Open, Boundless, Seamless, Dynamic Information System

Hornung's (1997) Glossary of the Web Dictionary of Cybernetics and Systems defines **information system** as "a system of functions concerning the acquisition and transfer of information, the carriers of which can be biological, personal, social or technical units. An information system is dedicated to a certain kind of information (topic), even if this may be a very broad one. It has always the purpose of providing information to a user or a group of users. In most cases a storage device is part of an information system".

Keenan (1996) points to two aspects which characterize an information system as an **open** one: the fact that it allows a variety of different computers and terminals to work freely together on the one hand and that it provides publicly available access on the other.

Kelly (1995) lists the following features of the WWW as indicative of its open information system nature:

- *Clients available on multiple platforms.*

- *Servers available on multiple platforms.*
- *Public domain clients and servers available.*
- *Specifications (HTML, HTTP) freely available.*
- *Specifications not controlled by single company/organization.*
- *Involvement in specification developments open to anyone.*

As an open system the WWW connects to and integrates material from various information systems - both hypertext/hypermedia and non-hypertext (Balasubramanyan 1995). The WWW does this in a seamless way that makes moving around its never-ending multi-user and multi-media continuum feel like "surfing" and facilitates the creation and edition of documents, as well as the creation and traversal of links at any point. Devised originally as an information delivery tool, the WWW now also stands for "the whole range of information resources available through computer networks" (ILC 1994-96), often referred to as "cyberspace". This is reflected in Hughes' (1994) claim that the Web has fulfilled Nelson's vision of creating a "docuverse" of the collective knowledge electronically available on the Earth.

The WWW has equipped users on computer networks with a simple and consistent means to access that "docuverse" by introducing a common interface capable of overcoming the incompatibilities of the existing platforms and tools on the Internet which had caused "waste of time, frustration and obsolete answers

to simple data lookup" (Berners-Lee & Cailliau 1990).

As a client-server application, the WWW is made up by a large number of computers (called hosts) which run either a server (a programme which distributes documents) or a client (a programme which orders and receives documents and facilitates their reading) or both (Hughes 1994). All the clients and servers are interconnected via the Internet and can communicate with each other irrespective of the differences between the protocols they use. Morgan (1994) distinguishes the following steps in a cycle of a typical client/server interaction:

- *The user runs client software to create a query.*
- *The client connects to the server.*
- *The client sends the query to the server.*
- *The server analyzes the query.*
- *The server computes the results of the query.*
- *The server sends the results to the client.*
- *The client presents the results to the user.*

The WWW allows for an extremely flexible and dynamic organization of information on the Internet. The property of scaling - the ability of information to

"smoothly reshape to represent the new state of knowledge" without any major or centralized changes is a principal factor for the Web's widespread popularity (Berners-Lee *et al.* 1994).

1.4.2 Hypermedia - a Non-Constraining Representation of Information

The medium of paper presupposes a two-dimensional structure of text following some traditionally meaningful or useful linear arrangement or hierarchical order. Texts in modern computer systems are also organized in a linear, sequential fashion. They are stored as files in various directories and subdirectories in much the same way as real documents in separate folders can be arranged in divisions of boxes or drawers that gather together similar material in broad-term categories. Any further relationships, however, between the texts in the individual files within the same or with other categories, remain hidden.

With the evolution of computers the new information representation technology of hypertext made it possible for such relationships to be displayed through machine-supported links allowing more complex, inter-related, non-linear organization of the text units.

> *In essence, a hypertext system allows authors or groups of authors to link information together, create paths through a body of related material, annotate existing texts, and create notes that direct readers to either bibliographic data or the body of the referenced text* (Yankelovich *et al.* 1988).

In this way, Yankelovich *et al. (ibid.)* emphasize, the analysis and further

extension of a given document can be studied without the distracting interruption of library bibliographic and other searches and without losing the original context.

> *Explicit connections - links - allow readers to travel from one document to another, effectively automating the process of following references in an encyclopaedia. In addition, hypertext systems that support multiple users allow researchers, professors and students to communicate and collaborate with one another within the context of a body of scholarly material (ibid.).*

An information system of this type which also includes media other than text, such as graphics, pictures, video, sound, animation, is referred to as hypermedia. The two terms tend to be interchangeable.

Conklin (1987) and Begoray (1990) identify the following essential characteristics of hypertext/hypermedia:

- A network representation of information
- Units of information (called "nodes", and also "documents"- the preferred term in WWW documentation) connected with various cross-referencing links allowing readers to choose from many different structures of the same information
- Machine-supported access to the information

Links between WWW documents are designated by anchors - numbered,

underlined or highlighted text or special symbols, called buttons, icons, or image maps. Activating these anchors enables hypertext jumping between documents irrespective of their format (plain text, hypertext, computer software, search result lists, images, video) or location. This new capability is provided by the Hypertext Transfer Protocol (HTTP) which is used by native WWW servers. Another big advantage of HTTP is its ability for format negotiation through gateways: the client sends a list of the codes it understands and the server acts accordingly (Berners-Lee 1994b).

1.4.3 Creation, Storage, Navigation and Retrieval

The WWW has no central authority or entry point. There are no restrictions for anyone with an access to a WWW server, no matter where it is, to publish information on it for everybody to read. The language of the WWW, which is associated with the transfer protocol and determines how documents are displayed, is known as HTML (Hypertext Markup Language). Based on the Standard Generalized Markup Language (SGML), it provides a set of tags for the structural layout of documents (titles, headings, lists) and for embedding the sources and targets of hypertext links (Furner-Hines 1995).

Automatic conversion of documents from other formats, like WordPerfect or Microsoft Word, into HTML is possible, but decisions on what to link do not lend themselves to automation easily , so at least for the present it will be up to authors and other "trail blazers" to establish most of the "useful trails" of association through the common record (Bush 1945) and to create the links manually (Krol 1994: 322).

To write an HTML document one does not need, in principle, anything more than a text editor. Nevertheless, software developers keep on producing multi-featured specialized HTML editors to facilitate the WWW authoring process. This is necessitated by the continuous upgrading of the HTML standard itself. Some of the latest features added to HTML are animated graphics, created by the specialized programming language Java, frames, forms and tables.

The Internet is a decentralized computer network, which means that "there is no one computer or group of computers to which every other computer is linked" (McKeown and Watson c1996: 4). In its complicated, multi-related and dynamic structure, the World-Wide Web has likewise no central node, no boundaries or privileged directions. It is "a boundless information world in which all items have a reference by which they can be retrieved"(Berners-Lee *et al.* 1994). Currently this role is played by the Universal Resource Locators (URLs), which since 1991 have also been used as the standard way to cite Internet resources (Chamberlain and Mitchell 1996). The network community is now working on developing schemes of naming resources. The Uniform Resource Name (URN) "identifies a resource or unit of information independent of its location. URNs are globally unique, persistent and accessible over the network" (URN implementors 1996).

A URL is a string of symbols. It consists of four parts: the first one starting from the left shows what protocol the Internet resource addressed by the URL uses: "http", "gopher", "wais", "ftp", "telnet", "file" (for local files), and "mailto" (for electronic mail) (Morgan 1994). This is followed by the name of the host where the resource is stored. Then comes the path leading to the directory in which the resource resides. And finally comes the file name. Sometimes this last item can be absent from a URL in which case the server offers a default document. When

a user knows the URL of an Internet resource, he can enter it in his browser, as Web clients which "allow users to move across or within screens, windows, records and databases" (Marchionini 1995: 100) are referred to, for an immediate retrieval of the resource.

The WWW provides the means of organizing the information distributed among the millions of its own and other computers (hosts) connected directly or through gateways to the Internet by linking them as a common set of hypertext documents (Krol 1994: 289). The user can navigate from country to country, from service to service, and from one type of server to another with the same ease as turning the pages of a book thanks to the WWW's architecture of a retrieval system for distributed information which "allows information stored in any number of host computers, or servers, to be accessed by any number of client computers connected to the servers across a wide-area network such as the Internet" (Furner-Hines 1995).

The choice of a starting point for navigation in the WWW is very important for the achievement of the user's objectives. Special documents, known as home pages, are accessible on servers to welcome and guide visitors to the contents of the information stored on their sites or to other, non-local interesting links. The home page for a WWW server, or for a special collection on a server, has the look and the purpose of a magazine's table of contents (Chamberlain and Mitchell 1996).

All the WWW browsers, e.g. Lynx, WinWeb, Mosaic, Netscape, Internet Explorer, provide the ability to select a particular home page and to go back to it from anywhere on the Internet in one move. It is also possible to go back one

step at a time through previously visited documents (or pages) or look at a list of the visited pages and go straight to any of them. Since WWW pages have no limits to their length, long ones can have links labeled as "top" (to take the user quickly to the first screen) and "bottom" (to take the user quickly to the last screen) of the current page. Collection of nodes on related topics, distributed as one, (composite nodes) can show the hierarchical or chronological relationships between their units by links such as "next" and "previous".

In addition to providing access to documents that include multimedia, the WWW graphical browsers (the most popular ones being Netscape and Mosaic) enable the user to save permanently menus with selected destinations for further use. These are called "hotlists" in Mosaic and "bookmarks" in Netscape.

Navigating by following links in the vast and disparate information environment of the World-Wide Web is not always the most effective method of information seeking. For this reason a large number of automated search tools, or search engines, have been installed on various servers to allow keyword or string searching as in traditional information retrieval mechanisms from familiar electronic document systems, e.g. online library catalogues, CD-ROM databases. The result of searching an indexed space on the WWW is "a special kind of node generated 'on-the-fly' called a virtual node" (Relihan 1994). Searching for a string within a page is available as a menu option in Mosaic and Netscape .

1.4.4 The Technological and Social Context of Global Computer Networked Communication

The World-Wide Web was designed as a "collaboratory", a system allowing

simultaneous multi-user access to "electronic networks so that investigators around the globe can witness the results of an experiment as it progresses" (Stix 1994). This forms an excellent basis for the further development of democracy in the information creation and dissemination field, underpinned by two fundamental Web principles:

- "Universal readership" (Berners-Lee 1994b), emphasising universal accessibility, both technological and human, to computerised information

- Universal, rather than individual, authorship, "in which everyone involved in an area can contribute to the electronic representation of the group knowledge" (Berners-Lee 1994a).

Unfortunately, the initial effect of the development of the Internet was to increase the gap between the information poor and information rich individuals and nations. The spread of the WWW from the research community to the general public has not favoured collaborative writing either because of the lack of the necessary technological knowledge and skills on the part of the general user.

Thus both principles - of universal readership and authorship, seem to be still more of a goal than a reality because of financial, practical and training constraints even in the most advanced countries. What could their prospects be in a problem-ridden environment, such as the one inherited from apartheid in South Africa, then? A brief discussion of this issue is due.

1.4.4.1 The Social Relevance of the Internet in South Africa

The transformation process encompassing the whole complex of socio-economic changes in South Africa today has overall development in a democratic context as its top priority. Under the impetus of the current world communications and information technology revolution, it is a widely-recognized fact that without the proper utilization of information as a resource for development (Boon 1992), no country has a chance of keeping abreast of progress in the global information society. This society is described by Shillinglaw (1988) as "crucially dependent on theoretical knowledge and access to information for its growth and development and its decision-making in social, political and economic fields".

Transformation requires a revolutionary approach to classical needs. The problem of identifying and meeting the information needs of the society at large and its various sectors has long been here, but as part of the transformation process it should be tackled from a totally new perspective. The report of the Preparatory Mission of the National Information Management Project - South Africa (Harfoush and Wild 1994) emphasizes the need for a "more intensive use of information by all those participating in the development process, and particularly by the government which is managing the process, and the communities which are driving it" as a major guarantee of the democratic constitutional right of access to information and government transparency.

A number of potential stumbling blocks along this road have been identified by the mission:

- Misunderstanding of **information** as equivalent to **information**

technology; existing information systems geared to the needs of administration rather than the planning and decision-making required for development

- Absence of awareness on the part of the general population of information as a tool for community-driven development and empowerment

- No consistent assessment of the different needs of different groups of users with a view to designing and allocating resources for systems, services and related training programmes that will "encourage training in the use of information and its application to problem-solving at all levels and not simply training in how to use information access tools"

The economic and social demands in the highly interactive world of the "global village" put a premium on information and knowledge, and especially on science, as the sources of higher productivity. This is reflected in the reference to present-day society as "the knowledge society". It is also described as "a learning society", which strives not only after a more pervasive spread of knowledge but of a more profound one.

A new educational philosophy has been brought to life by the powerful processes of globalization. It seeks to create the opportunity for learning by doing, rather than a mere delivery of instruction. This learner-centred philosophy of education underlies the framework for transformation envisaged by the National Commission for Higher Education (NCHE) in South Africa (NCHE 1996). A Working Group on Libraries and Information Technology (WGLIT)

was set up to analyze for the NCHE "the strengths, weaknesses and the future role of libraries and information technology in the development of the tertiary education sector" (WGLIT 1996).

Addressing the **Conference on Information Technology in Tertiary Education** (Cape Town, April 1996) the then Minister of Arts, Culture, Science and Technology, Dr B. S. Ngubane (1996) emphasized the importance of information technology for successful higher education and economic and social development. He then called on tertiary institutions to reach out to schools in networking and to help make every child computer familiar. What motivates such an appeal?

1.4.4.2 Information Technology and the Philosophy of Education

Comparing the impact that advances in technology have had on the development of science on the one hand and education on the other, Soloway (1994) states that while technology has driven major changes in science, it has failed to do so in education. Narrowing down to TV and radio, the author attributes their lack of effect on education not to any deficiencies in the technology itself but rather to the way that it has been used.

Soloway and Pryor's (1996) dismissal of electronic technologies as having had "little - if any impact on education" can only be accepted with certain reservations. While it may be true that Thomas Edison's 1922 prediction, quoted by them, that the motion picture will supplant the use of textbooks, has not been fulfilled, electronic technology has caused major changes in the informal general education of the international public. There is an all-too-evident world-wide

trend of the younger generation abandoning reading for pleasure in favour of TV and video viewing, and computer games, thus developing a different type of literacy and set of skills from the older one. The appeal of popular educational documentaries to all generations has been indisputably confirmed by TV stations' experience, as has all adults' need for news been evidenced by the world-wide support for the 24-hour newscasts from CNN and the BBC.

McLuhan (1964: 332) tells of an experiment in which four randomized groups of university students were given the same information at the same time on the same topic via a different channel for each group - radio, TV, lecture and printed text. The information was delivered on the first three channels by the same speaker without any deviations from the text or the use of a blackboard. Each group's session lasted for half an hour after which the same quiz was filled in by all. The results showed that the TV and radio groups performed better than the lecture and text ones, with the TV group standing well above the radio group. McLuhan attributes this effect to the high level of participation on the part of the viewer required by the engaging medium of TV.

Why has technology not been effective in the classroom, then? Norman and Spohrer (1996) find the answer in the old content-centred approach to teaching. Soloway (1994) points to the abstract character of lecturing: "TV and radio were used to support a bankrupt educational philosophy" - to deliver instruction rather than create the opportunity for learning by doing. Lectures as the dominant instructional strategy are overused and overrated.

The new learner-centred philosophy of education focuses on the needs, skills and interests of the learner. The following factors are particularly relevant to the

effectiveness of instruction:

- Motivation of the students
- Application of the learned material
- Student assessment

Motivation is ranked first of all the factors contributing to the students' success or failure. Computer-based instruction engages the student because of the diversity and intensity of interaction and the multi-sensory impact of multimedia. Problem-solving , which is a natural activity in the computer environment, is also highly-motivating (Norman and Spohrer 1996).

One reason why the mass media have been successful in informal learning is because of their ability to combine instruction with fun. In contrast, the classroom atmosphere, with its captive audience and competitive evaluation pressures, emphasizes the artificiality of the environment and detracts from the pleasure of the discovery process, thus suppressing natural curiosity as a motive for learning.

Schank and Kass (1996) argue that application of the material in the process of learning is essential to its retrieval in later life. For learning to be effective, goals must be set "that will motivate students to access information; provide an authentic context in which to situate the knowledge students access; and confront students with specific challenges that require them to analyze the information they access, and put it to use".

Student assessment should also be related to a goal-based, problem-solving approach since "traditional tests measure declarative knowledge: learned recitations and applications to small problems [and] do not necessarily address depth of understanding or the skills the students have acquired" (Norman and Spohrer 1996).

The new technologies, and especially the Internet and the World-Wide Web, make the satisfaction of the above-mentioned global societal needs through the application of the new education philosophy possible. The concept of "user-friendliness" in human-computer interaction is gaining a new focus: from the creation of user-centred interfaces (which are less time-consuming, easier to learn, and reduce the user's cognitive effort) to learner-centred ones which "support learning while doing tasks". Not only can students be given authentic tasks, but these are "appropriately scaffolded" - i.e. they enable the learner "to start doing the task with his or her current understanding, but then challenge and channel the learner to develop the next level of understanding and performance" (Soloway and Pryor 1996).

1.4.4.3 The Role of the Internet and the WWW in Education

A number of problems of traditional instruction can be solved via the use of the Internet. Among these is the application of the principles upheld by the new philosophy of education, namely:

- *From individual to group learning*

- *From competition to co-operation*

- *Accountability to the group*

- *Active involvement of all learners* (Miller 1996)

Ellsworth (c1994: 5-7) describes the Internet as "a powerful releaser of emotion, motivation and engagement for students". It offers the most current information and immediate contacts all over the world. It teaches students "to ask better questions, to make better arguments, and to present themselves more positively over the Net"and helps them improve their understanding of math and science. It encourages educational independence and intellectual autonomy and does not discriminate.

Educational uses of the Internet, highlighted by Dr Miller (1996), include:

- Local and *foreign e-mail discussions*

- Electronic forums of interest groups composed of people all over the world

- Establishment of personal contacts and peer-to-peer exchange of ideas, information, evaluation through e-mail correspondence

- Collaborative projects with international research, educational and student communities, group interaction at a distance

- *Teletrips*

- *Global videoconferencing*
- *International news magazines*
- International scientific e-journals
- *Netsurfing for research materials* all over the world
- *Distance learning*
- Tapping into *professional and government* expertise
- Accessing the world's library bibliographic and full-text resources

In this environment the instructor will be able to assume the role of a councilor and assistant to empowered, active, engaged learners.

All this is possible at a minimal cost. Computer simulated lab experiments, for example, are a way of reducing real science lab expenses. Furthermore, a simulation such as the **Virtual Frog Dissection Kit** (http://george.lbl.gov/ITG.hm.pg.docs/dissect/info.html) can save the lives and the expense of the slaughtering of three million frogs a year only in the USA.

The World-Wide Web's special contribution to education is the advantage of learning through free association in the multi-sensory environment of hypermedia. The interrelatedness of elements of information in semantic networks composed

of nodes of associated attributes, which are linked by ordered, labeled relationships, represents the model of human knowledge structures that modern cognitive psychology draws on in defining learning as "a reorganization of knowledge structures" (Jonassen 1988). Hypertext can be seen as the physical environment most amenable to the application of this model.

Among the advantages of learning through hypertext Jonassen (1988) emphasizes the learner's having control of the text's structure enabling him to tailor the learning process to his own needs of personal relevance and curiosity fulfilment, interest and experience level, information requirements and task demands.

Even textbooks can become exciting on the World-Wide Web. Brown (1996) tells of students collaborating in writing Internet materials as part of their course. Certain publishers supplement their textbooks with additional and more up-to-date information, while others debate issues raised in a book on the Web.

The World-Wide Web has numerous other technological strengths that have important social implications. Some of these will be looked at below.

1.4.4.4 The Technological Strengths of the World-Wide Web

Of the four Internet information retrieval tools which came into existence in the beginning of the 1990's - Archie, WAIS, Gopher and the World-Wide Web, the latter two were the more user-friendly and enabling ones. As mentioned in **1.3.2** Gopher took an early lead over the WWW, but this proved to be short-lived. Chamberlain and Mitchell (1996) describe Gopher as being superseded by the

World-Wide Web:

> *It's hard to imagine that gopher, while still so young and obviously effective, is already being overshadowed, overwhelmed and, in most cases, absorbed into the newest and strongest Internet navigation tool yet - the World-Wide Web; a move which is changing the face of some Internet sites ... Be prepared to encounter lots of empty gopher holes in your explorations, and always check for dates on available material! Because of the difficulty of maintaining two or more servers (Gopher and Web), some hosts are electing to close down or at least cease to update regularly their gopher servers and are moving much of the information previously stored there to their WWW Home Pages.*

The reason behind this lies with three major areas of technological advancement in which the WWW has improved upon the other information delivery systems on the Internet:

- The WWW has incorporated practically everything that these disparate systems have to offer in its integrated, uniform service

- The WWW has achieved a remarkable simplicity of operation and ease of access to information which do not require sophisticated knowledge about computers on the part of users

- Other media such as graphics, animation, video and sound can easily be manipulated as independent nodes or inserted in text which makes the WWW a truly multimedia invention

A review of several authors' opinions of the implications of these advantages of the WWW as compared to Gopher will be offered. This is not an attempt to evaluate the effectiveness of the two systems, since "effectiveness must be assessed from an explicit viewpoint", that of a particular constituency, and at a particular level - organizational or individual (Grover *et al.* 1996). The aim of the comparison is to pinpoint the areas of technological advance achieved by the WWW.

Archie and WAIS are not included in the comparison because they are closer to the category of index searching tools such as Veronica for Gopher, and the numerous search engines for the WWW, e.g. AltaVista, Excite, Lycos, WWW Worm, Web-Crawler, InfoSeek Guide, Open Text Index, Inktomi, AliWeb and others. Instead of locating specific information as the above, Gopher is the only Internet tool that offers browsing through Internet resources somewhat similar to browsing on the WWW. Since browsing is at the centre of this study, a comparison between Gopher and the WWW, rather than between Archie, WAIS, Gopher and the WWW, seems to be the more relevant one.

1.4.4.4.1 The World-Wide Web Versus Gopher

The World-Wide Web exhibits distinct technological improvements on Gopher in the following areas: access, ease of use, organization, presentation, interactivity, scope, variety and cost.

Access

Both the WWW and Gopher are client-server applications that allow the user to

find and access data from various sources on the Internet, by running for him the relevant programme - email, ftp, telnet, Archie, WAIS. However, the WWW adds its own material to the Internet, while Gopher does not - "there are no specially formatted 'Gopher resources'" (Krol 1994: 235). Gopher space is thus accessible to the WWW clients which can read ASCII (the text format used by Gopher), while Gopher clients cannot read HTML. WWW users who have slow connections and computers unable to cope with graphics can still access the Internet data with the text browser Lynx.

Ease of Use

The Web enables the user to travel the whole of cyberspace by the same press of a few keys or mouse clicks, never seeing or feeling all the different commands and transactions that go behind the computer screen. No matter what service is being accessed on the Web, all the user can do is follow links in the same way or fill in forms to perform a search. Gopher is also easy to operate. The user browses and selects items from menus or performs string searches. However, the Gopher interface can change from resource to resource.

Organization

Gopher organizes its data through hierarchical menus, while the Web uses hypertext. Mosaic and Netscape support both these and other models. While there are no restrictions to hypertext linking, Gopher's menus can only point to other menus, individual resources and servers - no links can be made between concepts and ideas inside files or from the files of one server to another.

Every Gopher server is like a separate library with its unique collection and its resources organised in categories and linked to a root subject tree - the menus. The top-level menu is the home server menu that comes with the client. This normally leads to other menus taking the user to directories, then - to sub-directories and finally to end-documents in files (Krol 1994: 235-240).

The order followed in the menus is chosen by the managers of the host sites, usually from the ones used in traditional librarianship - by subject, alphabetically, geographically (Chamberlain and Mitchell 1996). No standard subject lists are used, either, with all the familiar consequences of scattering and inaccessibility of information because of synonymity, changes and discrepancies in terminology. This effect is particularly pronounced because, unlike the WWW anchors, Gopher menu listings have no surrounding context to clarify their meaning (Krol 1994: 292).

Presentation

Gopher is predominantly text-oriented. Although some of its clients can access pictures and sound, they can do that only separately from text. The WWW's ability to insert (embed) all the other media in text, so that they can be perceived as one entity, gathers together the functions of full-fledged publishing and instant world-wide dissemination. This is opening new vistas before commercial advertising and the entertainment industry and is changing the WWW user demographics.

On the other hand, the hypertext model presents a much more readable and useful view of USENET news and mail list archives by turning their inherent cross-

references into links. Such is the flexibility of the Web that a new system like virtual reality, a three-dimensional environment in which computer-generated simulations make the user experience it as if it were an actual one (Huston-Somerville and Kreitz 1995, Shreeves 1994), will also be delivered over it (Ford 1995: 2).

Interactivity

Gopher is a read-only tool (Krol 1994: 234). In contrast to it, the World-Wide Web uses the same tools for publishing as for reading, thus bridging the gap between authoring and reading. Authors are far less dependent on intermediaries and have unlimited access to self-publishing and navigating to sites of self-publishers (Samuelson 1995).

Scope

Unlike Gopher, which is a strictly public environment, the subject matter that is considered publishable on the Web is no longer of an impersonal, scientific nature only. The added opportunity for a holistic representation of the human personality, gratifying the deep-seated human need for self-expression, is yet another factor promoting the expansion of the WWW and its transformation "from an abstract, chaotic, information web into ... a social hypertext" (Erickson 1996). In this way the Web encourages less formal and more direct contacts between authors and readers.

Variety

Since the incorporation of the human element in the WWW opens the door to general interest and recreation as well, entertainment and business are steadily gaining the upper hand over science and computing sharing, thus taking the system forward from its academic origins to serve the whole of mankind. "Recent development of the Web has been driven by commerce and entertainment, rather than scientific publishing", notes Carts-Powell (1995).

Cost

Gopher and WWW client software for any operating environment can be obtained free from the Internet (Krol 1994: 236; Furner-Hines 1995). While both systems share the "low costs associated with desktop delivery" (Collins 1996) of information, the greater versatility, leading to the much wider application and reach, of the WWW makes it the more cost-effective of the two in the long run. It is becoming less and less affordable for businesses to ignore the economies of scale and competition and the resulting reductions in costs offered by the unrivalled scope and speed of the global information network (Ricart 1994) and the advertising power of the WWW homepage. How the World-Wide Web can make publishing more cost-effective will be discussed in Chapter 2.

From the above it becomes clear that while Gopher will still find application among organizations and individuals looking for a simpler, but effective and less resource-consuming system, the WWW will be preferred by those who need to make use of its much more versatile functionality.

1.5 Conclusion

The Internet has provided the hardware and software capabilities that allow for a communication model in which all the advantages of the various types of the old media are combined:

- The immediacy and speed of face-to-face communication
- The transportability of the message across space and time of mediated communication
- The large-scale storage and mass dissemination of printing coupled with the multi-sensory powerful effect of electronic media

These features are further enhanced in the far more enabling and empowering environment of equal involvement of all parties in the communication process in the two-way human and information interactivity which is the hallmark of the Internet.

The Internet's power to provide effective all-round modern communication has been augmented by the rise of the World-Wide Web. The latter has incorporated and improved through the flexibility of hypertext all the preceding services on the Internet for the generation and control of the communication and information traffic in cyberspace.

The World-Wide Web's strengths , as discussed above, have earned its acceptance as "the friendly face of the Internet" (Ford 1995: xvi) and testify to

this system's potential for making its mark on the future. This will not be achieved, though, without the efforts of competent intermediaries to manage and improve the new information environment, as well as train the millions of non-specialists in its effective use. Library and Information Science needs to develop research into WWW-related issues and help the profession identify its new role in the global information world.

2

The World-Wide Web and Its Effects on Libraries and Library and Information Science

2.0 Introduction

The state of WWW-related research is indicative of the unprecedented pace and far-reaching impact of developments in the global electronic information revolution. Unless electronically generated, which dramatically speeds up the publisher's editing process, print publications dealing with the WWW up to 1994 inclusive usually reflect the period before the 1993 breakthrough release of Mosaic and other versatile browsers. Such publications have to be used very carefully, since the facts and conclusions in them have for the most part become outdated. Even today the constant generation of WWW innovations and applications places a strong emphasis on keeping abreast of the latest progress made. Therefore, despite the traditionally greater authoritativeness of established print publications, the up-to-date electronic sources reporting WWW-related research are steadily gaining in importance.

Young as it is, the World-Wide Web has firmly established itself as the universal

information provision system of choice on the Internet, which is one of a number of compelling reasons why it should be subjected to an extensive and multidisciplinary research. So far research and development efforts have focused on advancing the technology of the WWW which has matured significantly as a result. The major need for research now is to seek ways to improve the system's usage which requires the cross-fertilization mix of contributions from diverse fields, including computer science, information science, library science, cognitive science, human-computer interaction, sociology, communication, and education.

Since the dawn of civilization, librarians have studied the needs of users and the ways they interact with various information sources and systems. They have used their knowledge to devise schemes for organising information and improving access to it. The infinite bulk, variety and dynamism of the global electronic information universe, as well as the radical changes introduced by the new technology, challenge the traditional ways of bibliographic description, reference, retrieval, authority and control. Yet, the principles behind these are no less valid today. There are also many aspects of the navigational techniques through the network of nodes and links on the Web which are waiting to be investigated from a LIS user-centred point of view.

The development of the Internet and the World-Wide Web has put pressure on libraries, especially in the advanced countries on the planet, to embrace the new technology and transform themselves accordingly. In the context of globalization the developing countries realize the need to follow suit. Inevitably, libraries have started exerting their own influence on the global information system and this trend can only be expected to grow.

2.1 The World-Wide Web and Libraries

As any new medium with a wide-ranging impact, the Internet in general, and the WWW in particular, provoke strong reactions. Popular opinion is divided between attempts to dismiss them as "a momentary freak of fashion" and overenthusiastic claims that they are the panacea for all information provision problems and will eliminate the need for any intermediaries such as publishers, libraries and librarians.

The steady exponential growth (see **Fig. 1-1** and **Fig. 1-3**) and all-pervasiveness of the Internet and the WWW argue against their being a momentary fad. On the opposite extreme, a long list of problems in managing the networked digital world, that have proved a hard nut to crack by either traditional or novel approaches, ensures the need for information provision mediation in telecomputing for many years to come. What is more, proper user education, which has always made the difference between the effectiveness or failure of any computer system (Crawford 1995), will have a crucial role in the smooth transition to this revolutionary environment.

As key institutions of information storage and dissemination, related to the education and publishing sectors, libraries have had a historic part to play in the spread of knowledge in society. Under the impact of the global electronic information revolution, fundamental changes will have to be introduced in the theory and practice of Library and Information Science (LIS) if librarians are to continue playing a successful role as information managers and mediators.

Libraries have had to harness the growing power of the information

superhighway in transporting huge amounts of data right to the desktop of the user at work, at school, and at home. They have endeavoured to do so by evolving from the mere metaphor of the virtual library to building their own sites on the WWW and then to creating digital libraries. These developments will be the subject of the following two subheadings.

2.1.1 Virtual Libraries and Traditional Library Sites on the WWW

Keys (1995) distinguishes three stages in the development of technological revolutions:

- The first stage is characterized by the application of the new technology for improving old practices
- The second stage is the period of changing old practices
- The third stage sees the introduction of new practices

From the very launch of the WWW, its creator T. Berners-Lee has been aware of the need for library methods of organization to help users find their way amidst the enormous mass of information on the Internet. That is why as early as 1991 he started the World-Wide Web Virtual Library at CERN.

The distributed interlinked storage of universally accessible digital resources of all types and in all media on the WWW is the perfect foundation for the idea of "virtual libraries". These take the form of subject lists (or subject catalogues) linked to resources residing anywhere on the Internet. The virtual libraries are

useful as starting points for browsing when people have only a general idea of what information they are looking for (Slot 1996).

In the case of the World-Wide Web Virtual Library every subject on the list is handled by a different site (Secret 1996). In an effort to achieve high scholarly standards subject experts were asked to develop annotated lists of sites in their fields. Unfortunately this led to a lack of consistency in the authors' approaches, the completeness of the information and its maintenance. In general, human-created subject lists prove to be very labour-consuming and difficult to update in the rapidly expanding and changing Web universe (Tillman 1996).

In their turn, traditional libraries have also rushed to establish their own sites on the WWW. According to Furner-Hines (1995) the production of home pages was the most common active use of the Web in European libraries in 1995. These pages usually offer local information - the whole range from the library's location and opening hours to information about collections and staff. They also provide links to the library catalogue, sometimes to full-text documents and forms for comments, questions and criticisms from the users. External links to other libraries' OPACs or sites of interest may also be provided.

Libraries are deemed by Keys (1995) to have reached the second stage of technological change. Peters (1995) confirms this by pointing to the improvement to communication and information delivery brought about by the Internet through "a variety of networked resource publishing, organizing and discovery tools and systems ... and the transition of the information industry from a 'just-in-case' way of doing business to a 'just-in-time' one". In other words, the library profession has moved on from adapting traditional libraries and

librarianship to the new technology of the WWW and has started changing librarianship in accordance with the demands of digital libraries. This process is still in its early days. The issues facing digital librarians, to be discussed throughout the remainder of this chapter, provoke heated debates.

2.1.2 Digital Libraries

Schement (1996) points out that "the information society is evolving away from the traditional notion of community" and that online technology "has created a primary industry of information access and distribution that is bypassing the library". Therefore traditional libraries have to transform their place and role in accordance with the forces currently reshaping the information society, and so do librarians in order to "establish their profession at the very centre of the new social fabric".

Huston-Somerville and Wilt (1995) note that because of the information explosion the conventional information storage and retrieval processes have become too slow and costly, and have been challenged by the new ways of information delivery which have ignited "a paradigm shift - from ownership to access of resources". In the networked electronic world context, the library no longer has to be bound to a physical building or to any location, nor to its own collection or the number of copies of a resource in its holdings. Not only do librarians have unlimited resources at their disposal but these can be simultaneously accessed by as many users as needed and delivered at any time and anywhere without mediation.

The vision of this emerging new type of library has come to be referred to as a

"digital library". The preference of this term over a "virtual library" seems to indicate a move away from copying the structure and the functions of a traditional library to the creation of a different entity. Similarly, "digital" has come to replace "electronic" to emphasize the shift from text to multimedia.

According to the Digital Libraries Initiative (DLI) glossary (1996) "digital libraries basically store materials in electronic format and manipulate large collections of those materials effectively". The DLI, for instance, comprises "six research projects developing new technologies for digital libraries -- storehouses of information available through the Internet ... The projects' focus is to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks - all in user-friendly ways".

There is no generally accepted vision of what digital libraries are going to be like - they are still more or less in an experimental phase. Samuelson (1995) does not deem it likely that they will be centralized, but that they would rather develop as an Internet-like network of specialized collections. Hastings and Tennant (1996) believe that "massive digital libraries will be built through the cooperative activities of numerous institutions".

Tennant (1996a) describes this process as manifested in the birth of the first Digital Library SUNSITE which was dedicated at the University of California Berkeley Library on 31 January 1996. SUNSITEs are created by Sun Microsystems, Inc., positioning some of its most powerful computers at various locations around the world to foster software, information, and technology exchange. Thus the infrastructure was provided for the UC Berkeley Library to

"usher in the digital age for libraries" by marrying "technology to scholarship, the past to the future".

The Berkeley Digital Library SUNSITE includes all the components of a classical library:

- A collection of materials selected for their contribution to the library's purpose and their usefulness to the clientele

- Value-added services provided by trained professionals: organization of the collection to ensure easy access to the material, assistance to users

Furthermore, in its role as a catalyst for the global development of digital libraries, museums, and archives, the Berkeley Digital Library SUNSITE is working in accordance with the following objectives:

- *Gather and publish information about digital library projects*

- *Gather and provide access to digital content*

- *Provide a platform for digital research and development*

- *Promote discussions on topics related to digital libraries, museums, and archives*

- *Provide current awareness services*

These objectives are implemented with the co-operation of other libraries, professional associations, universities, and commercial companies working in the field: e.g. Sun Microsystems, Inc., providing continued support through innovative technology, Electronic Book Technologies (EBT) donating SGML software, University of California Press publishing humanities journals in SGML on the Web, the California-State Library-funded InFoPeople Project for training library staff in using the Internet to both access and publish information, Association of Research Libraries contributing to the collection of documents related to digital libraries.

There is a SUNSITE in South Africa too. It is at the University of the Witwatersrand and its URL is http://sunsite.wits.ac.za/. A brief review will follow now of the issues concerning the development of digital libraries.

2.2 Digital Libraries Research and Development

Friedlander (1996) states that "the central assumption of digital libraries ... is the existence of collections of digital information linked by communications networks that enable access by individual researchers anywhere and at any time". Levy and Marshall (1995) challenge the idea that certain characteristics, supposedly inherent to traditional libraries, will automatically be transferred to digital libraries. They believe that:

- Digital library collections will not contain only documents of unchanging permanent nature, but gray literature and ephemera as well.

- Digital libraries will not consist of digital material only but will also

contain and integrate non-digital material.

- Digital libraries will not be used by individuals working alone but they will offer support for communication and collaboration in the users' information-seeking activities.

The implications for research and development that Levy and Marshall foresee from the above are that while the Internet and the Web technologies provide a useful library infrastructure, their content does not constitute a collection, i.e. "a selection of items organized for a particular clientele". Questions about the relevance and nature of collection development, cataloguing, and other library value-added services on the Internet and the WWW have led to a number of experimental initiatives based on traditional or radical approaches - the former relying on established library practices (e.g. the OCLC Internet Cataloguing Project) and human or human-mediated indexing, and the latter looking for new machine-based services (e.g. Lycos, AltaVista, Infoseek and other search engines).

The WWW-induced changes in the practices and information provision principles upheld by libraries have called for a transformation in the role of the library profession as will be revealed in the coming pages.

2.3 The Role of the LIS Profession in the WWW Environment

For the past few years a substantial number of LIS publications, as well as the most prolific discussions on the special library Web managers' mail list, Web4Lib, have centred around the impending and implemented changes in the

traditional libraries' external and internal environments and the role of librarians in the global electronic information age. These changes affect all the areas of classical librarianship:

- Access provision
- Information management
- Intermediary functions
- User education

A review will now be made of some of the key issues facing the library and information science profession in the performance of their digital librarians' functions.

2.3.1 Internet Access Provision

LIS scientists and workers have a number of key problems to resolve in adopting the new WWW technology, matching it to the users' needs and providing public access to Internet resources.

Van Brakel (1994) points to the general lack of scientific publications in teaching hypertext as "an indication that universities and other tertiary institutions have not yet embarked on extensive hypertext training". Even in the USA most of the library employees have had to give way to computer science specialists or grapple with the new and rapidly developing environment of the World-Wide

Web through self-study (Keys 1995), the reason being insufficient opportunities for training (Hastings and Tennant 1996). This is quite a daunting task, especially where a new medium such as the Internet is concerned, since the difficulty of use of an information system is not only due to technology but to other factors such as social problems and financial constraints, the lack of generally accepted standards, ignorance, dogma, and the resistance to change which to a lesser or greater degree is part of human nature.

The complexity of these deterrents has crystallized into three distinct problems on the Internet which have unanimously been identified as a priority that needs to be dealt with before the "information superhighway" can become a useful conduit and "dramatically improve efficiency and quality of life" (Shearin 1994). These problems are:

- Security
- Copyright
- User-friendliness

Security and copyright will be discussed here only in the light of their relation to the provision of public access to information.

2.3.1.1 Security, Copyright and Public Access to Internet Resources

As a result of its unrestricted communication and publication power and the potentially unlimited accessibility of its resources, the Internet is infinitely more

difficult than print publishing to channel and control by governments, organizations and individuals in both the propagation of undesirable or classified information and the infringement on intellectual or financial property, all of which has crucial positive and negative implications.

On the positive side is the enhancement on a global scale of the democratic right to freedom of expression encompassing freedom of speech and the right to receive information, upheld by many library organizations, the American Library Association (ALA) included. Nevertheless, the ALA Council (1996) notes, "many people, for reasons of technology, infrastructure, or socio-economic status do not have access to electronic information". Even if libraries should lose their role as places of information storage, digital librarians will have an even greater contribution to make in empowering the information (and financial) have-nots. They will assist the public, both adults and minors, by "selecting, producing, providing access to, identifying, retrieving, organizing, providing instruction in the use of, and preserving recorded expression regardless of the format or technology".

On the negative side of the Internet's unruliness is the threat to the livelihood of authors and publishers and to confidentiality and privacy, the proliferation of electronic forgeries and thefts by unscrupulous net crackers (Wallich 1994), and the danger of children being freely exposed to pornography on the Internt. These are serious problems, yet in determining their policies on issues such as whether to allow full anonymous Web access and whether identification of users should be required in the name of security, librarians should take care that no concerns are addressed at the expense of users' access to information. "Users have the right to be free of unreasonable limitations or conditions set by libraries,

librarians, system administrators, vendors, network service providers, or others. Contracts, agreements, and licenses entered into by libraries on behalf of their users should not violate this right" (ALA Council 1996).

In view of as major a change as the commercialisation of the Internet, an entity that inherently cannot have a centralised body to govern its evolution, Kahn (1994) is rightfully concerned about its future viability, if left to market forces alone to determine its growth, standardisation, security, intellectual and financial property and privacy protection, and equitable distribution and competition.

Equal access to information has long been the raison d'etre for libraries espousing the democratic ideal. Since the WWW is expected to increase its role in offering education, health services and the means to participate in community affairs, librarians need to pro-actively assume "a leadership role in defining and distributing universal service" (Lifer 1996). A balance has to be found between two conflicting factors that have always presented a difficult problem before universal service - the financial viability of the system and the cost to the user. In the context of a progressively prevailing market-economy libraries are among the few institutions that could be relied upon to look for ways to resolve fairly the public versus commercial provision dilemma.

Measures to ensure security and privacy, copyright law to safeguard authors' and publishers' financial interests, and filtering or censorship for the protection of children have always been balanced by librarians' efforts to enhance public access to information - a role that will grow even more in importance in digital libraries. Digital librarians will have to study the pros and cons of the various ways of generating revenue for the compensation of authors and value-added

services by libraries and publishers, e.g. the pay-per-use, the fixed fee subscription or other systems and match them to the needs of their patrons. Then they will have to see to it that under these systems "public access is available on fair, reasonable, and nondiscriminatory terms" (Samuelson 1995).

Another stumbling block to universal access in addition to the financial one - the lack of user-friendliness in an information system, will be discussed now.

2.3.1.2 User-Friendliness

Government policies, commercialization and privatization have brought the Internet within reach of millions of prospective users who know little about computers and electronic networked environments. While eager to benefit fully from the resources and opportunities provided by the Internet, these users are not prepared to invest more than a minimal amount of time and effort in mastering it. The public expects their traditional intermediaries - librarians, to assist in making the environment more user-friendly and bringing the universal library's treasures home to everybody.

"User-friendliness" is a concept which includes both sides of the human-computer interaction process. The WWW presents a simple, integrated and easy to operate interface to the end-user, the non-expert seeker of information. Yet, it lacks in friendliness as far as human factors, such as "the user's understanding of a system as well as his feelings and behaviour towards it" (Geyser 1992b) are concerned.

Feelings play an important part in motivating human behaviour in any situation.

Positive feelings are particularly needed if one needs to develop an understanding of a complex and radically new environment. However, radical changes are usually perceived as threatening by most people. Certain manifestations of the effects of users' feelings on their understanding of the WWW will be touched upon now.

2.3.1.2.1 The User's Understanding of and Feelings Towards the WWW System

Both experience and research in information seeking in online environments preceding the World-Wide Web have shown that ease of operation does not remove a whole host of other communication barriers arising from a novice's lack of knowledge about the content, structure, conventions, context, terminology and presentation of a new computer application. Despite its seamless look, the WWW poses a huge cognitive challenge in its capacity as an innovative synthesis of heterogenous services, media and meanings to both intermediaries and end-users.

Libraries and librarians have experienced the most profound effect by the emergence of the information superhighway which has opened undreamt of vistas before the profession. Yet, the very magnitude of the revolutionary changes has caused this new technology to be widely perceived as a threat to the accepted values and aspirations, and even to the very existence of information workers.

For example, some, who tend to equate the librarians' profession with books alone, have blamed the emphasis on technology for having compromised traditional scholarly values and promoted information over knowledge and the

computer over people (Wisner 1994). Others have decried the dehumanizing of knowledge and the triumph of "small-scale thinking over large-scale thinking" reflected in "cyber-babble" (Quinn 1994). Worries have been voiced about the "loss of context, about the impossibility of attaining the relationship with a text that books can provide" (Shreeves 1994).

Such reactions are to a great extent a case of "resistance to change" (Fine 1994) that librarians entering the electronic world will have to be prepared to overcome, first in themselves, and then in their patrons.

Negative feelings can also be provoked by the overselling of technology, prevailing in recent years, "which raises user expectations even faster than the performance and value of systems can increase" (Kantor 1996). An example of this can be found in the commercial hype surrounding the development of new WWW browsers. Microsoft's (1997) announcement of the impending Platform Preview test release of Internet Explorer 4.0, for instance, goes as far as to state: "Now users can have the information they care about delivered straight to their desktop when they want it, the way they want it".

Experts and experienced users know that they have to take claims like that with a grain of salt. They look for well-defined evaluation criteria, developed by third, unbiased parties, to help them form their own opinion or they can rely on first-hand impressions in comparing the way different products satisfy their needs, e.g.:

If you know how to use Netscape Navigator or any other browser, you also know how to use Internet Explorer. Almost everything is similar. (Mathiassen

c1996-97)

Inexperienced users, however, would expect the "push" and "pull" technology of Internet Explorer to live up to its promise of delivering Web content in their familiar "point-and-click" TV fashion. They are bound to experience tremendous disappointment when faced with the reality. Even librarians find it difficult to keep up with the fast-moving advances in information technology and their broader implications for the future of information provision. To avoid disappointment, leading to rejection, Shreeves (1994) advises them to resist "self-defeating" claims that electronic dissemination of information will immediately solve all problems, "from the high cost of scientific serials (to be discussed in section **2.3.3.4**) to the deterioration of brittle books".

Despite their brittleness, conventional paper publications will continue to exist for quite a long time, because of a large number of reasons. For one, they will continue to be far more accessible to a much wider spectrum of the population due to the inevitable financial, infrastructure and training limitations typical of the spread of a new medium. People's habits will also have a similar effect: editing on paper rather than on screen is preferred by many (Smith, Weiss and Ferguson 1987) as is reading a book in bed over reading a computer screen.

Great hopes are entertained about the Internet revolutionizing education, but, as can be seen from the impact of electronic technologies on education so far (described in section **1.4.4.2**), there are no guarantees that this will happen automatically, if at all.

Furthermore, irrespective of what major breakthroughs may be achieved in the

technological aspects of information provision on the Internet, the "challenge of information overload on cyberspace" (Berghel 1997) will continue to grow. One underlying cause, as Berghel points out, has to do with the inherent quality of the Internet of being now, and remaining forever, "credibility- and value-neutral". The ease with which anybody can publish on the Internet and the WWW's "dual role of being both a private and public information and communication medium" furthers the need for a massive evaluative effort and channelling of relevant, quality information to the right audiences, which is the purpose of information management in digital libraries.

2.3.2 Information Management in Digital Libraries

The management of digitised information is becoming the most important part of the global communication system, which is steadily advancing on the arena occupied by the classical means of information storage and dissemination and is set to dominate the world throughout the next century and beyond. According to Prytherch (1995) "information management is an imprecise term for the various activities that contribute to the effective production, coordination, storage, retrieval and dissemination of information, in whatever format, and from internal or external sources, leading to the more efficient functioning of the organization". Although typically used in the corporate, private sector context, the term is also "extended to include the more technologically-based activities of research libraries". As such it is deemed in this study to be well-suited for the WWW digitised information environment which serves the purposes of a research collaboratory and a meeting ground for public and private communication. "Information management" includes the narrower term of "information handling", denoting "the storing, processing and retrieval of

information from acquisition to user".

With its new-technology implications and the infinite bulk and variety of a global electronic information treasury, the WWW presents an enormous challenge to Library and Information Science researchers, particularly in the areas of bibliographic control, information retrieval and reference services. Moreover, there are many navigational aspects, as well as human-computer and author-reader communication interactions on the Web which need investigation. Some of the issues involved in the information handling activities of storing, processing and retrieval of information will be discussed now.

2.3.2.1 Identifying, Evaluating and Selecting Internet Resources

Libraries emerge in answer to the information needs of the communities they serve. In view of the modern trend of globalization of the scholarly and other communities, traditional local centres, such as for example universities, can no longer be the only factor defining a library's nature and membership. Thus the identification, evaluation and selection of world-wide distributed digitised resources for integration in digital libraries can only be useful and meaningful if the librarians first identify their target audiences, anticipate those audiences' needs and gather their information accordingly (Morgan 1994).

One way of defining a community is by common research interest. The US National Centre for Supercomputing Applications' (NCSA 1996) Digital Libraries and Information Systems Research Group WWW page leads to the following description of "community systems":

A "community system" is a digital library that encodes all the knowledge of a community and a software environment for interacting with that knowledge across the network. The knowledge includes formal materials, such as literature and databases, and informal materials, such as newsletters and results. The environment enables users to browse existing materials and analyze selections, and to share new materials and forge links.

It follows from the above that digital librarians are there to serve an interactive unity of a geographically dispersed community and its knowledge, formal and informal, through a common and enabling technology.

A well-defined audience profile is a prerequisite to the development of effective criteria on what resources to select and in what way and how long to store them. These decisions are at the heart of collection development and archiving policies that digital libraries will need to formulate no less than traditional ones.

2.3.2.2 Collection Development and Archiving, Catering for Special User Needs

Collection development which involves selection based on the relevance and value of resources for a given field of users (Magier 1995) is a key function "in a post-scarcity information environment [where] the task of the librarian becomes at least partly one of filtering and evaluating a flood of information" (Keys 1995).

Although electronic technology may be able to solve the paper-created problem of storage space, matter and energy limits will still influence acquisition and

access. Furthermore, trying to save everything is counterproductive because it makes finding information difficult. That is why it is best to consider the archival value and life-cycle of documents "at the times of creation and storage" (Marchionini 1995: 174-175).

In their collection development policies, digital librarians will continue to uphold the principle of balanced objectivity in information provision by providing access to information presenting all points of view (ALA Council 1996). This principle is crucial to the development of science as well as of democratic societies which are based on the ability of citizens to make informed choices.

The Internet is a great equalizer - it frees the handicapped from some of their limitations and removes "gender, race, and personal appearance as factors in professional acceptance" (Keys 1995). As a multimedia provider the WWW has an unmatched potential to cater to the disabled. Digital librarians are called upon to make sure that there are enough special collections for various disabilities and that free public access to them is available. Work to this effect is already under way (Paciello 1995).

Collection development will also be affected by the preservation function of digital librarianship. Particularly noteworthy here is the extremely high volatility of the material on the Internet, greatly expanding the share of gray and ephemeral literature.

2.3.2.2.1 Gray and Ephemeral Literature

Decisions on what are quality resources and what is worth preserving cannot be

made lightly. Gray and ephemeral literature, which is far more prevalent on the Internet than in the paper print world, is not necessarily less worthy of archival retention. On the contrary, scientific full-text preprint archives are appearing on WWW servers throughout the world, providing "an incredible acceleration of the dissemination process - the original purpose of the preprint as a form of scholarly communication". Furthermore, according to preliminary statistics more electronic preprints than hardcopy-only ones are eventually accepted for publication (Huston-Somerville and Kreitz 1995).

Similarly, personal homepages should not be easily dismissed as mere vanity press. Neither does the fact that a source, like a listserv electronic discussion, is less formal in its tone, detract from its scientific value today, or its historic value tomorrow. Laura Guy (1995), a member of the Web4Lib electronic discussion, shares her concern that valuable Internet resources might be lost before the problem of digital archiving has been dealt with:

> *I have grave fears that listserver archives, newsgroup archives, electronic journals and newsletters and even gopher sites and web sites will disappear forever without a trace, thus eliminating the possibility (for whatever reason) to study them as artifacts.*

According to the current practice, specialized archives for particular areas of interest are created by contributions from their communities and organized by administrators. These archives do not cooperate with each other (Bowman *et al.* 1994) - a state of affairs which is obviously not conducive to the development of inter-disciplinary collaboration and the standards necessary to support it.

Even the most successful collection development and archiving policies will be meaningless, however, if the stored information resources are not organized in such a way that an effective and efficient access is provided to them.

2.3.2.3 Organizing the Internet

Information, especially when accumulated in large volumes, becomes very difficult to organize successfully. The WWW Virtual Library has demonstrated the pitfalls encountered in attempts to develop a comprehensive centralized subject catalogue for massive distributed and fast growing dynamic resources by distributed human cataloguers. In the past two years there has been a shift away from global human-mediated subject catalogues, such as the WWW Virtual Library, Yahoo, EINet Galaxy, in favour of smaller, more specific subject guides and automated indexing (Johnson 1995).

Prior to the World-Wide Web there were very few automated tools for searching the Internet - Archie for names of files in ftp archives, Veronica for gopher menus, and Wide-Area Information Servers (WAIS) for database indexes. With the explosion of the Web other alternatives had to be found. Surfing, or following links to all types of Internet documents through any of a number of possible paths, and creating bookmarks or hotlists soon grew out of hand (Randall 1995). Even bookmarks were so numerous that they could not be useful unless they were organized. That is how Yahoo was born - a hierarchical subject index that had its origin in the well-organized Web bookmarks of two Stanford graduate students (Tillman 1996).

A status report on "Networked Information Retrieval: Tools and Groups"

produced as a collaborative effort by the Internet Engineering Task Force, the Association for European Research Networks and the Coalition for Networked Information in August 1994 (Foster 1994) assessed networked information retrieval as being in its infancy compared with traditional information retrieval systems. Thesaurus construction and boolean searching were still limited, but rapid progress was being made in this area, as well as in the discussions on classifying and cataloguing information resources, the report noted.

The following major trends have been established in the approaches to organizing the Internet today:

- Software developers have created robots, or spiders, or wanderers, trying to index the whole or large parts of the Internet so that automated search tools, or search engines, such as AltaVista, Infoseek, Lycos, and many others (some of them listed in **1.4.4.4**), can retrieve material from it. Eventually they hope to create artificial intelligence agents that will supersede reference librarians.

- Some librarians believe in the expansion of traditional cataloguing, e.g. through the inclusion of an electronic resource access field in traditional human-mediated machine-readable catalogues, while others believe in innovative applications of the long-established principles of librarianship.

Examples of arguments used during a discussion on cataloguing the WWW by members of Web4Lib will give an idea of the diversity of viewpoints:

- *If we users can get directly at an item through a variety of methods,*

including traditional as well as emerging kinds of access points, why should we want to go through an intermediary step of accessing a catalogue entry? (McEvilly 1995)

- *Surrogate records -- catalogues, abstracts, finding records - are valuable in and of themselves: they afford the means to constrain searches and sharpen the relevance of search terms. This will be increasingly evident as the size of the Web grows, and result sets grow to unmanageable size. Without the structure of surrogate records and the added value of human-coded description, the retrieval problem will get worse much faster* (Weibel 1995).

- *These reference lists and cataloguing attempts are good bandaids for now. But they will never be current enough, accurate enough, complete enough, or easy enough to provide the research power you and your patrons will want. You must understand that you are applying 16th century, two dimensional concepts of organization to a dynamic, evolving, uncontrollable wilderness of data* (Hoff 1996).

Evidently the problem of whether to catalogue or not the WWW is open to discussion and it is not very likely that it will be resolved in the near future. So are the related problems of resource description for electronic objects and bibliographic control of Internet resources.

2.3.2.3.1 Evolving Resource Description for Electronic Objects and Establishing Bibliographic Control of Internet Resources

One of the principles of bibliographic control which helps librarians to organize information so that it can be found again is the ability "to design a citation that presents an integrated and intelligible description of a document" (Taylor 1996). Traditional cataloguing records, such as MARC, are too labour consuming and costly to produce manually for the enormous quantity of Internet documents. The new automatically generated indexes (e.g. Lycos, WebCrawler), however, do not contain sufficiently detailed resource descriptions, and often have nowhere to get them from, since quite a few resources on the Internet are described by nothing but their file names, and the latter will not necessarily reflect their content (Weibel, Godby and Miller 1995).

The Dublin Metadata Core Element Set (or Dublin Core, for short) proposes a description, halfway between an index entry and a formal cataloguing record in detail and complexity, that could be created by authors themselves and entered in a standardized form for collection in the automatic indexes (*ibid.*). Other ideas about what the bibliographic citation for electronic resources on the Internet should include and where it should be found will be touched upon below.

The OCLC's Internet Cataloguing Project, a US nationwide project of libraries and institutions of higher education, applied the conventional bibliographic control processes to Internet objects and created a searchable database of conventional USMARC format bibliographic records, including USMARC field 856 for electronic location and access information (McKiernan 1996).

Yahoo catalogues differently, as cataloguer Anne Callery (1996) explains. Traditional libraries need to integrate electronic with traditional resources, so they have to use MARC records. Yahoo catalogues only Internet resources and is not constrained by the rules of an existing collection. Full bibliographic descriptions of Internet materials are unnecessary since access to them is easy. Yahoo uses its own classification scheme which is developing together with the collection - a "bottom up" approach. Users submit sites for addition to the hierarchical subject list and suggest the category they should fall under. The Yahoo cataloguers examine the site and make the final decision on the category. The meta-data fields (which are not displayed at present) include: "title, URL, contact person, geographic location, descriptive comment, and indicators for the presence of Java and VRML".

McKiernan (1996) recommends the use of a WWW-based resource description template containing cataloguing and classification data for the identification and incorporation of significant resources within a collection.

Harris (1995) sees the catalogue record for each document in the document's header itself. "Searches could then look for specific header elements. Essentially the Web would be its own catalogue. The trick would be getting Web authors and publishers to put the full complement of header elements into all their documents".

The development of principles and systems to increase bibliographic control is fundamental to enhancing access to Internet resources.

2.3.2.3.2 Providing Enhanced Access to Digital Resources

One of the ways for librarians and information scientists to provide enhanced access to digital resources lies with testing and evaluating the performance of the constantly growing number of Internet information retrieval tools for their own purposes as expert seekers and for end-users.

Tillman (1996) divides the existing tools for finding information on the Net into the following categories:

- Review tools
- General and Specialized Guides
- Directories
- More traditional library resources
- Search Engines

An example of the review tools is Magellan, an online Internet directory, which reviews and rates thousands of sites on its listings according to the following criteria:

- Completeness of content presented in the resource
- Organization of the resource

- Currency of the information presented

- Ease of access to the resource

The Argus Clearinghouse is a collection of subject-specific guides to Internet resources. This more specialized type of tool is far more manageable and rich in context than any other one. Unfortunately, unless a product of this kind is a part of a big collection of guides and/or produced by a highly popular, authoritative site, it will be difficult to bring to the notice of the wider community.

Yahoo and the WWW Virtual Library are referred to here as directories (another instance of the fluidity of terminology, typical of a new and fast developing environment, such as the WWW). In Callery's (1996) opinion, the organization of material in a hierarchical subject index, such as Yahoo's, brings about a higher relevancy in search results since the grouping together of similar subjects in classes and subclasses provides enough context to distinguish between different meanings of a word. On the other hand, subject lists allow for browsing, so the user does not have to find by trial and error the synonyms of a search term used in the index.

Din (1995) finds, though, that unlike librarian-generated information systems, Yahoo is one of those efforts "which can have unscalable indexing policies for special types of materials, that suffer from inconsistent treatment of materials (even when vocabulary exists), that lack subject categories for many documents (thereby increasing scattering), and which do not show much evidence of understanding relationships between subjects".

McEvilly questions the practice, adopted by Yahoo and OCLC, of "only cataloguing main pages, not sub-pages: But what is a main page? Who decides what is THE main page for a site?" (McEvilly 1995)

A more traditional library resource is, for instance, OCLC's NetFirst database of bibliographic citations "complete with summary descriptions and subject headings, describing resources including World-Wide Web pages, interest groups, library catalogues, FTP sites, Internet services, Gopher servers, electronic journals, and newsletters" (Makssour 1995).

Proud (1995) sees an advantage of the NetFirst approach in the value-added service of resource selection "according to prescribed selection criteria and the description and classification of selected resources using agreed standards such as Library of Congress subject headings and Dewey classification".

McKiernan (1996), however, rejects the alphabetical order of listing the resources within the categories in NetFirst, and of most other organized Internet collections, as too restrictive. The lack of pointers to conceptual relationships between resources forces the user to examine all listed resources within a category. McKiernan's model for selecting, organizing, presenting and accessing WWW resources, Cyberstacks (sm), applies the Library of Congress classification scheme while adopting "hypertext and the browser medium to create context and structure to provide associated meaning".

Kuhn (1995) sees the way to improving organization and information retrieval on the WWW in:

1) Better robots which can do some hierarchical indexing and some day

even mapping of terms in different languages. (There has been a project in Germany to do the multilingual mapping for a "normal" library OPAC already ...)

2) Increased indexing by the authors of WebDocs using the controlled vocabulary of their field of interest (like e.g. MeSH or Biosis concept codes and biosystematic codes). These index-terms would then be recognizable to the improved robots.

3) Increased manual cataloguing of WebDocs of greater interest by libraries.

4) An even increasing number of lists of specific resources in certain subjects.

5) Other things we don't yet know about.

Against this background of diversity, the current two major types of Internet information-finding tools - the subject indexes and the search engines, deserve a more detailed comparative examination of the way they work and their ensuing strengths and weaknesses.

2.3.2.3.2.1 Subject Indexes Versus Search Engines

A comparison between the objectives and achievements of the human-mediated subject indexes (e.g. The WWW Virtual Library, Yahoo, EINet Galaxy, The Whole Internet Catalog, The Internet Public Library) and the search engines (e.g. AltaVista, InfoSeek, Lycos, OpenText) shows that so far neither of the current two major types of Internet information-finding tools has proved to be satisfactory enough or the more effective one.

The chief value of a subject index is in its links. Creators try to add value by either including as many as possible of the Internet materials relevant to the various categories, or linking to carefully selected resources. Yahoo has chosen the quantitative approach, while the Internet Public Library (IPL) prefers the qualitative approach (Slot 1996).

Search engines work with databases where documents are indexed on the basis of words extracted from the title or text. Indexes make their choices according to different criteria - e.g. Lycos uses the top 100 words, while AltaVista and Infoseek are full-text indexes (Lager 1996).

Retrieval is dependent on a good match between the criteria entered by the user and those used to describe the document "as it was extracted and indexed into the database" (Slot 1996) which in many cases presupposes extensive knowledge both in the domain of interest and about the search engine itself to enable the user to properly identify and encode his exact need.

Search engines compare users' queries with the indexes and return the found matches. Where every word, or the most used words, are indexed, the searcher can expect a higher recall - "the percentage of total relevant documents retrieved from all documents". Recall and precision - "the percentage of documents retrieved that the searcher is actually interested in" - are improved by the incorporation of various techniques in the different search engines: e.g. thesauri, Boolean operators, relevancy feedback, probabilistic logic, fuzzy logic, Bayesian networks, concept based searching (Lager 1996).

Unfortunately, the implications of these techniques differ and are not easy to

grasp, so very few of the information professionals, let alone of the users, can take advantage of the advanced searching features and effectively refine their queries. The situation is further aggravated by the constant upgrade to the search engines' interfaces which does not give a chance to users to get really familiar with them (Stanley 1996).

Because they are much faster than humans in processing data, the search engines can build many times larger databases than the human indexers. What is more, the whole catalogues can be rebuilt regularly at short intervals - e.g. Lycos does that every Saturday in as little as 7 hours (O'Balle 1995). This keeps the search engines databases up-to-date in both new acquisitions and the maintenance of links.

On the whole, the difficulty of search refinement is the weakest point of the search engines, while scaling brings about the worst problems faced by human-mediated indexing. Yahoo is a perfect example of how its comprehensive scope and quantitative approach have caused the levels of abstraction in the subject hierarchy to grow out of proportion. The fourteen general subject headings are divided into 50 sub-headings on the first page and the user must go through three or more index pages before reaching a page of links (Slot 1996). For this reason Yahoo has had to reinforce its data processing power by combining its subject index with the Open Text search engine (Callery 1996).

Another factor influencing access to WWW resources is related to the use of their physical storage address as a means of locating and accessing them. The stormy growth of the Web infrastructure, the move of equipment and individuals within and between organizations leading to the transfer and reorganization of

materials, leads to an enormous number of changed and dead links within original document sites and other sites that had linked to them. This is one of the major problems requiring Web resource maintenance.

2.3.2.4 Web Resource Maintenance

One needs only to visit the old CERN WWW project pages at their new W3C site to realize the magnitude of the problem of managing the storage migration and links integrity of electronic documents. Web managers are still few in number and although automated link checkers and reporting utilities (e.g. Webwatch, MOMspider) have appeared, maintenance issues will only begin to be resolved, according to Guthery (1995), with the implementation of the URN standard.

Web resource management concludes the list of activities classified under the heading of information management in this study. Although all of these activities are aimed at improving the user's access to relevant WWW information, they involve librarians in the direct handling of information only. Librarians' direct interactions with the users are perceived as part of their intermediary and user education functions. The intermediary function is centred around finding quality relevant information for the user. This includes assisting the user in recognizing his information needs and monitoring user satisfaction.

2.3.3 Digital Librarians' Intermediary Functions

In Martell's (1994) opinion "the substitution of access for ownership [will be] a paltry paradigm shift in comparison [to] ... a change in which a focus on people

is substituted for a focus on collections or information". Bearing in mind that the Internet is no less about live contacts with people than with records of information, the LIS profession must focus "not on technology but on the people for whom the technology is developed".

A user-centred approach to the librarian's intermediary functions presupposes a good grasp of his users' information needs together with the ways and means of negotiating their perceptions of those needs, as well as selecting and improving those aspects of technology applications that affect the meeting of those needs. To do so librarians must develop their understanding of the issues involved in and skills required for doing precision searches for Internet resources, facilitating the users' WWW navigation, evaluating users' feedback, publishing and disseminating information on the Web.

2.3.3.1 Doing Precision Searches for Internet Resources

Subject indexes take too long to browse and search engines are by no means user-friendly tools, so when a high recall and precision of results are required, searches will have to be done by expert intermediaries. Taylor and Clemson (1996) and Morville (1996) point to the following prerequisites to an effective automated search for Internet resources:

- Being aware of the search engine's database size and contents
- Understanding the search engine's underlying index strategy
- Knowing the query language, the principles of on-line searching and the

information-seeking process

- Knowing how to eliminate language ambiguities, enhance context and refine queries

The larger the search engine's database size and the more general its contents, the more diluted context becomes, thereby increasing ambiguity (Rosenfeld 1996). Search engines have no way of exercising quality control so they retrieve any materials they find. Ambiguity and lack of quality control grow ever worse with scaling. Fortunately, librarians have a long tradition and a solid theoretical and practical background in dealing with ambiguity and quality control in the print and on-line worlds.

2.3.3.2 Facilitating the Users' WWW Navigation

In view of the complexities and shortcomings of the search engines, Rosenfeld (1996) calls on Webmasters to pay particular attention to the navigation features of their sites as the most effective way of improving the Web's services to the user. He emphasises the importance of a clear architecture designed for the library's particular audience, the provision of a consistent and obvious way to navigate the site, the effective labelling of content, visual orientation aids, such as maps, and the "weeding" of outdated and useless material.

2.3.3.3 User Feedback Evaluation

Getting to know their patrons so that they can identify their target community as a whole as well as the various user profiles, and measure user satisfaction is a

vital goal for any service. On the Internet, where freedom of choice and interactivity are the norm, users will not stay with a library that does not allow them to play an active part in it. Librarians will have to create more direct and friendlier opportunities for their users' involvement in the library's policies besides testing their opinion through survey forms.

A valuable means of monitoring readership of a Web site or digital library is to use statistics based on the records of browser requests in the server log file. These help with identifying the site's or library's audience as well as with collection development since they show which documents are requested and which are not (Ford 1995).

Yet a simple counting of the times a given WWW document has been visited is insufficient for an accurate assessment of its usefulness and more detailed information will have to be provided by future log-file recording and analysing programmes, as well as by usability research. In the third chapter of this thesis an attempt will be made to throw some light on what this additional information could be.

2.3.3.4 Publishing and Disseminating Information on the Web

The availability to the reader of the tools for publishing, the ease of distribution, as well as the quality control and enhanced accessibility of information on the Internet brought about by collection development, are all conducive to the amalgamation of the functions of publishing and digital libraries. In the same way, "Internet access in classrooms blurs distinctions between teaching and learning" and holds the promise of turning digital libraries into digital schools,

reminiscent of the Alexandrian library of the third century BC (Marchionini and Maurer 1995). Several trends have a stimulating effect on this process:

- The increasing specialization of scientists in narrower fields is taxing severely and undermining universities and their supporting libraries as centres of a local community of scholars in favour of globally distributed research communities, "held together by telephone, fax, Internet and conferences" (Denning 1996).

- The inadequacy of the existing system of scientific publishing which has become too slow and expensive and is giving way to much faster, more effective, universally accessible and cheaper channels for scholarly communication and collaboration. This is parallelled by years of underfunding of libraries and universities creating a need for more durable and compact storage and the reduction in cost of the constantly growing information resources, as well as the sharing of expensive equipment (Heath 1995).

- The discrediting of the old content-oriented philosophy of education and the adoption of the learner-centred approach (Norman and Spohrer 1996).

Physical and geographical proximity constraints have caused traditional libraries to restrict their resources and services in accordance with the type of learning they support: school and university libraries for formal learning, public libraries for informal learning, and special libraries for professional learning. Unlike them, digital libraries have the potential of integrating the support of all three types of learning (Marchionini and Maurer 1995).

Despite their differentiated approach, traditional libraries can no longer support research because of the diversity of scientists' specialization which requires more and more of the rapidly proliferating diverse resources, storage space and varied storage media, and intermediaries to manage the resources and provide information services. At the same time, libraries and higher education institutions have had to cope with decreasing budgets over the past decade.

In its turn, the traditional system for scientific publication has become a bottleneck for scholarly communication because of the prohibitive costs of producing scientific journals which have very few subscribers to recover them from (Stix 1994). In an effort to find a solution for the unavoidable steady cut-backs on journal titles libraries have tried to share resources. On the other hand, a number of electronic journals,. that are offered free or charge only a minimal subscription, have appeared. The problem of decaying paper, made of acidic wood pulp, which papermakers have been using since the mid-19th century, has given prominence to space-saving and durable digitization as the better medium for preservation. Against this background, libraries and publishers alike have come under pressure "to find new ways that enable researchers to get and to pay for only those journal articles that are needed" (Browning 1996).

Another crippling weakness of paper-bound libraries and publishing is that they cannot compete with the speed, richness and versatility of the Internet's "community systems", described in **2.3.2.1**, in meeting the needs for communication and collaboration of geographically dispersed scholarly groups. The flourishing of electronic scientific full-text preprint archives, mentioned in **2.3.2.2.1**, and the complete works posted by professional authors and researchers on WWW servers, are indicative of the challenge to the print tradition of

scientific publication (Denning and Rous 1995).

The example of Paul Ginsparg, a physicist at the Los Alamos National Laboratory, using his computer in his spare time to provide a daily wire abstracting and full-text preprint article delivery service to researchers from more than ten disciplines in the physical sciences and mathematics in over sixty countries points to a qualitative change in the way scientific research is done. Rather than reporting the history of a completed research project scholars are now benefiting from peer comments while their work is in progress which helps them achieve much greater heights (Stix 1994).

If they cannot use an electronic preprint service, authors who need to publish in established print journals rather than in electronic ones, to gain prestige, promotion and tenure, are slowed down in their progress because of having to wait the 12-36 months from submission to publication before sharing their results with the research community (Denning and Rous 1995). Some of the most authoritative researchers, who have reached the top in their careers, no longer publish in any other way but electronically. Judging from these developments, there is little doubt that once the refereeing process, which is the current means of quality control, is firmly established for electronic publications, the latter will become the generally preferred medium.

Ginsparg's e-print archives server which for a long time was installed on a modest PC in his office replaces not only publishers but hundreds of libraries as well, as far as current research is concerned. When copyright issues, as discussed in **2.3.1.1**, are resolved, back scientific journal volumes can be archived in digital libraries.

Once the copyright obstacles are removed, most of the remaining contemporary documents which are first created in a digitized form anyway, can immediately be added to digital library collections. With the rise of multimedia and the World-Wide Web, quite a few electronic works have been produced that do not lend themselves to publication in any other medium. The same is true of the live communication genres indigenous to the Internet, e.g. discussion forums, newsgroups, videoconferences. Thus, it is certain that the already vast stores of materials on the Internet will multiply enormously, offering "information resources and tools that have traditionally been physically and conceptually inaccessible" (Marchionini and Maurer 1995).

The readership of a scientific journal is not of a homogenous nature. For the ACM (Association for Computing Machinery) journals, for instance, only 20% are experts in the journal's particular specialization, the rest are experts from other subdisciplines or practitioners. The latter need a different type of publication, referred to as "Track 2", that would help them keep up with the results of research in an accessible form (Denning and Rous 1995). Repackaging of this kind is easy to do on the Internet and involves practically no cost to prepare and distribute. This is one of the manifestations of "the interdisciplinary fluidity achievable with electronic communications [which] may have its greatest impact on science by stimulating cross-disciplinary discovery" (Calder 1996).

As a research society publisher, the ACM has taken pro-active steps to face the reality and transform itself accordingly. The major decision it has taken is to turn its journals into "streams flowing into the society's database" identifiable as "database categories" in the society's digital library. This will ensure instant publication, eliminating the need both to wait for the whole issue to be compiled

and also to impose page limits on authors' contributions. Furthermore the society has realized that publishers need "to provide well structured knowledge through digital libraries and easy-to-use tools" in order to survive (Denning and Rous 1995). Thus the logic of technology is dictating the amalgamation of publishing and digital libraries to meet the needs of scholarly research and communication (Browning 1996).

No less than for their intermediary functions, discussed above, librarians need a user-centred approach in the performance of all their tasks related to the user education function, which seeks to teach users how to effectively interact with the WWW themselves.

2.3.4 User Education

The growing importance of information technology has made librarians face considerable competition in "organizing, and helping users find, information" (McEvilly 1995). They have realized that despite their unique skills and long-standing experience in user education they might lose the users to other intermediaries if they do not keep up with technological advances. Hyman (1996) notes the signs of this change in attitude:

> *... there seemed to be three kinds of librarians: those that embraced the Internet even if only in their minds, those that hoped it was a passing fancy, and those who strongly opposed it. It has been interesting to watch the change over the last year or so. Everyone realizes that it is not going to pass by. There are even discussions now about 'Internet competencies' for teachers and librarians ... And the talk has definitely switched from*

single station public access to LAN, WAN, MAN, intranets and the rest of the alphabet soup.

Schement (1996) looks to user education as one of the major components of a 21st century strategy to be adopted by librarians. The focus of this strategy should be on teaching users:

- *to seek, find and process information*
- *to navigate through networks or locate information*
- *to think critically ... to prioritize and judge the quality of information they receive in torrents*

A brief discussion of some of the problems in achieving these goals will follow.

2.3.4.1 Seeking Information with the WWW

Hypertext, especially when applied to large databases as it is on the WWW, can be very difficult to navigate and can make the user feel disoriented. In the initial collaboratory framework of the WWW each new community, being both user and provider, had inside knowledge about where to look for its repository of information. The e-mail, news and discussion groups also helped out. Today a novice with no particular research specialisation, or a librarian who needs to find his way throughout all areas, feels quite overwhelmed by the WWW's unlimited choices.

Erickson (1996) regards the proliferation of personal Web pages, blending professional and personal communication, as the cause behind the transformation of the WWW from just another distributed information server system, such as WAIS and Gopher, into a new social phenomenon. The holistic approach in personal Web pages to professional expertise as part of an all-round, human personality with its hobbies, pets, politics and colleagues, has shifted the focus of users' searching habits from a technology-driven, "dehumanising" information retrieval to a computer network-based invisible college:

> *Rather than composing queries for search engines or going to likely places to browse, something that many ordinary users find foreign and daunting, people can instead pose the question: Who would know? Or who would know someone who would know?". The added bonus of the Web is that the source of the "personal communication" need not be directly involved in it.*

Finding information about persons is not simple on the Internet and librarians can offer great assistance in this, as well as in other areas related to good communication in the global information system by creating awareness of the services offered by the Internet and the digital library.

2.3.4.1.1 Creating Awareness of the Services Offered by the Internet and the Digital Library

Users need to be made aware of what is available on the Internet to be able to decide which services are best suited to their needs. The unique contribution of the Internet is that there is a ready live community for everybody to join and get

an easy introduction to everything they need to know. Recommending newsgroups and mail lists that are relevant to the needs of the users, teaching them the rules of "netiquette", and helping with access instruction should be the first step to be taken by librarians in user education.

Once they know what to look for, users need to be taught how to get to it, how to make use of the navigation clues and moves available to them on the WWW.

2.3.4.2 Navigating the Web

Basic navigation on the WWW is easy, but as was mentioned in **2.2.4.1** it will not take one very far. For example, Netscape's "Go" menu uses normal language titles of documents, but the Mosaic History list shows URLs. Users will not be able to keep track of their movements and jump back to any of the documents already visited if they do not know what the various parts of the URLs denote and are unable to extract location and semantic clues from them.

URLs are the only sources of document metadata available to the user. Understanding them, for example, helps when the all-too-familiar "Error 404 - document not found" message appears after trying to activate a URL from the location box or by clicking an anchor. That means that the document has changed its address or disappeared. Clues to its whereabouts or to another useful resource might still be found, though, in the directory or host levels if the user knows how to abbreviate the URL.

Introducing the user to the structure and purpose of navigational pages, such as contents pages and FAQs (Frequently asked questions), as well as how to create,

save and organize hotlists and bookmarks will also improve navigation.

Depending on their needs and preferences, users will also need a more or less detailed guidance on how to make the most of the WWW search engines.

2.3.4.3 Making the Most of the WWW Search Engines

Perkins (1995) identifies a very serious obstacle to the end-users' effective use of the WWW search engines without special training in the fact that it "requires a fairly sophisticated knowledge, not only of the subject being researched, but of language itself". As Scoville (1996) puts it, "No matter how big the database, or how sophisticated the indexing, a search engine is only as good as the query you give it".

In the course of a study Nahl (1996) has observed the following problems experienced by novices with search engines:

- Inability to formulate a query based on conceptual analysis - search terms too broad, narrow, numerous, extracted directly from the chosen topic statement

- Lack of understanding of options to control the precision of searches and of lists of links leading to various engines

- Failure to explore any links in the retrieved set to check whether the material is relevant to their topic

- Failure to see the search results on the screen

The last problem is caused by "screen-blindness", Nahl suggests. However, in certain cases this effect is due to the fact that the search results come at the bottom of an exact copy of the search page, with the form just filled in by the user - one needs to scroll down to see them. This is very confusing to a novice who is given no clue that the search has actually been performed and thinks that he has stayed where he was. Schneider's (1995) experience shows that "newbies" need to be given explicit instructions to scroll down pages.

Even experienced users have a hard time learning how to use every new engine. Exhaustive help information is not readily available, the search page layout is not standardised and the attempts at creative originality and catchy advertising are worse than distracting.

Unless one reads enough LIS evaluative articles about Alta Vista, for example, it will be impossible to decipher the purpose of the various parts of its logo. The latter represents the picture of a mountain divided into 5 sections with the following inscriptions:

1) Alta Vista Search OnSiteKnowledge
2) Advanced
3) Simple
4) Private eXtension products
5) Help

While 2) and 3) could safely be presumed to mean advanced and simple search, the meaning of 1) is quite elusive. There is no other alternative to a simple and

advanced search, so what is section 1) supposed to do? Then, if there are three options of searching to connect to, what is the purpose of the search form already offered on the main page? As it turns out, this actually is the Simple Search mode which is not the best option to encourage users to choose because a Simple Search is bound to produce hundreds of thousands of unranked and unordered results that will be impossible to wade through.

The technical lingo of the Advanced Search Help makes section 5) rather unhelpful too, all the more that it refers the user to the Simple Search Help for what the authors consider to be basic rules. This is not even done through hypertext links, but the user is expected to switch back and forth and read whole documents to get clarification on one point. All this is a proof that good computer programming, which Alta Vista definitely has, does not automatically result in an environment conducive to good author-reader communication.

Therefore, novice and casual users alike should be warned not to rush into using the search engines. Even with the best of them the searches that could be done without extensive training will more often than not produce unsatisfactory and even frustrating results. Using keywords and Boolean operators, for example, generally does not lead to end-documents containing the required combination, but whole sites with the keywords dispersed between the documents.

Users have to be made to realize that search results produced by the WWW search engines cannot be accepted at face value due to certain weaknesses, identified by Taylor and Clemson (1996):

- *Results are unpredictable.*

- *Results can be quite misleading - the same search can retrieve no hits by one engine and many hits by another.*

- *Search engines do not readily disclose the contents of their databases nor do they provide a description of the criteria used to include a document in their files.*

- *Vocabulary is not controlled, and punctuation and capitalization rules are not standardized.*

- *Names of persons are particularly difficult to find.*

The unpredictability and misleading potential of the search-engines-obtained results were confirmed by experiences in searches done for this research project. Thus, for instance, a search for "lurking" produced on Yahoo a long list of sites, while AltaVista came up with a glossary definition. A search for "communication model" produced no hits on Yahoo, but 833 668 hits on AltaVista.

The remaining of the above-mentioned weaknesses also need to be urgently pointed out to users so as to prevent immediate disappointment and rejection of the WWW as a whole. Examples illustrating the search engines' lack of transparency are:

1) *AltaVista does not provide any information in their 'help' files about how often the database is updated with new material* (Stanley 1996).

2) When search engines proudly announce their impressive sizes, they do not mention what measure of the number of documents indexed in their

databases they use:

Some services index the complete text of a document, some only selected portions. Other databases count a document as indexed simply because another document contains its URL -- on the assumption that descriptive text accompanies such a hyperlink! Although each method represents a count of "indexed" documents, only the first is the best measure of a service (Slot 1996).

The lack of controlled vocabulary results in the need for synonyms as well as for more specific and less specific terms to be searched for separately (Kuhn 1995). Simple, easy-to-follow, and up-to-date instructions for the punctuation and capitalization rules of the various engines must be provided for the users and they must be warned that the traditionally popular searching by author does not work so well on the WWW.

Users need to be made aware of the need to question the authoritativeness and lack of bias of Internet resources, since formal and informal, public and personal, library and commercial products exist side by side in cyberspace.

2.3.4.4 Evaluation of Internet Resources

In addition to the problems posed by the uneven quality and authority of its resources and the weaknesses of its information-finding tools, the Internet environment is also plagued by certain pitfalls typical of commercialism. Unlike the objectivity-oriented approach of public information provision, commercialism thrives on advertising which overplays the advantages of its products and can even be misleading.

Digital librarians must explain to their users that, among others, most of the search engines and directories are commercial ventures, "and not necessarily designed/organized by librarians" (Gooch 1996). Some of the consequences of this situation have been highlighted in an article reporting Yahoo's practice of "selling words" to companies - in this case the Iron Mountain Global Information Systems (IMGIS) purchased the exclusive rights to 55 words related to politics (e.g. "elections", "Republican", "Democratic") in return for which result lists for searches containing these words include a banner ad saying, "Top Ten Political Sites".

> *What is a bit sleazy is that this "top 10" listing is not based on any qualitative or quantitative measure. It is not the 10 most popular political sites with cybersurfers. Nor is it an expert's judgment of the 10 best political sites. It is, in actuality, a listing of 10 political sites that are willing to pay IMGIS to be included on the list. IMGIS charges $3,000 a month for the display space at the top of the Top 10 page; spots further down go for less. And on Yahoo! -- the most popular search engine -- the initial banner that directs a consumer to the Top 10 site is not even labelled as an advertisement. The bottom-line: IMGIS is buying up key words in order to push Internet traffic to its clients' sites* (Corn 1996).

Schneider (1996) underlines that user education reaches only a fraction of the WWW communities' members and the majority remain unaware that they are using commercial services' tools which are not concerned with the users' best interests. "The display of the 'Top10 Political Sites' banner is particularly misleading. A significant fraction of new users will not understand that these are merely ads and not results of their search" (Hogle 1996). Therefore Schneider

advocates that the public providers take a lead in the creation and maintenance of information-finding services.

2.4 Conclusion

Libraries have entered the second stage of technological change under the impact of the global electronic revolution. Rather than being obliterated or reduced in significance, they have become a focus of the efforts to address the numerous problems, generated by or associated with communication on the WWW. This has led to the emergence of a new type of service - the digital library.

At the forefront of technological advancement there is universal acknowledgement now that, although the creation and maintenance of WWW resources and tools involves certain new skills, "librarians already possess the more critical skills necessary to make these Internet services truly useful ..." (Morgan 1994).

There is a need for Library and Information Science research to launch an intensive investigation into all the traditional areas of librarianship as transformed by the Internet, namely: Internet access provision, digitised information management, digital library professionals' intermediary functions and user education.

The librarians' theoretical background, long-standing democratic traditions, professional ethics and experience as intermediaries between information provision and the users enable them to research the manifold challenges faced by the user in his information-seeking interactions in cyberspace. This research

will equip digital librarians to adopt a comprehensive and constructive approach in dealing with the obstacles to an ever-wider Internet access. Among these obstacles are issues related to pornography, censorship, piracy, copyright and user-unfriendliness, as well as with the difficulties in resource evaluation and selection, collection development, archiving, organizing, establishing bibliographic control, and maintaining the integrity of the Internet. By adding their perspective to the search for solutions to these problems librarians will help enhance the users' access to Internet resources through better information-seeking tools, expert mediation and user education.

3

The Relationship between Authoring and Information Seeking on the World-Wide Web: an Experimental Study

3.0 Introduction

The nature of the Internet and the WWW media is conducive to the convergence of all the elements in the information creation and dissemination chain: authoring, publishing, access provision, reading and learning. The librarians' traditional part in this chain has been to add value by selecting, storing and organizing quality resources that are relevant to their target audience, so as to enhance access. What is unique about WWW collection development, though, is that being relieved of the concerns arising from the acquisition and organization related to the physical storage of resources, it emphasizes the creative contribution of information managers and intermediaries (e.g. Web masters, digital librarians, educational courseware compilers) to the writer-reader communicative acts.

An experimental study of the information seeking process was carried out on a collection of interlinked hypertext documents constituting a PC-based model of the World-Wide Web. The method used was that of computer simulations. The objective was to investigate the correlation of different WWW authoring concepts and the success rate in browsing sessions performed by novice, casual

and experienced Internet end-users. The analysis of the results from a number of browsing sessions on the model led to some specific recommendations for comprehension-oriented WWW authoring.

Before the actual experiments are presented, it is necessary to clarify some theoretical aspects that are of vital importance to understanding the results from the simulations.

3.1 Document Linking and Structuring - an Important Part of WWW Authoring

Even if they are not directly involved in the creation of content on the WWW, intermediaries and even readers are active participants in the authoring process. That is so because, unlike books on the library shelves, documents on the WWW are not fixed, static, discrete units whose meaning stays exactly the same irrespective of how they are arranged. The WWW documents' environment, and hence - their immediate background information, is subject to constant changes by the opportunities created, according to Conklin (1987), by hypertext:

- for modularizing information and making it possible for *the same text segment* [to] *be referenced from several places*

- for tracing references *forward to their referent, or backward to their reference*

- for *creating new references: users can grow their own networks, or simply annotate someone else's document with a comment (without*

changing the referenced document)

- for segmenting and organizing the same *text in different ways depending on differing viewpoints*

- for imposing both hierarchical and nonhierarchical organizations; *even multiple hierarchies can organize the same material*

- for customizing documents: *text segments can be threaded together in many ways, allowing the same document to serve multiple functions*

A WWW document is created within the context of its author's purpose. Since it can be part of different organization schemes, however, its role and structural position within the same set of documents can change according to the nature of the different underlying substructures. When another author or reader makes a link to this same document, it invariably becomes incorporated into a new unit of information that reflects a different viewpoint and imposes a different structure. How text is segmented and linked, anchors chosen and overall structures created distinctly affects the meaning conveyed by documents, as well as the ease and effectiveness of following paths through, or browsing them, and is therefore a crucial part of the WWW authoring process.

It is this aspect of WWW authoring which is of the greatest significance to information intermediaries, educators, and end-users whose primary need is to structure disparate bits of content created by other authors into coherent entities. From a psycholinguistic point of view, a document is regarded as coherent "if a reader can construct a mental model from it that corresponds to facts and

relations in a possible world" (Thüring, Hannemann and Haake 1995).

Since no visible boundaries exist between the various structural layers of information on the WWW, in order to achieve coherence in a collection of documents special attention must be paid to assisting the information seeker by providing the necessary context for a complementary macro and micro retrieval of information. Of prime importance here is the understanding of the "two methods for explicitly linking two points in hypertext - by reference and by organization" and the two basic types of documents that go with them - primary (the ones carrying information content) and navigational ones (providing overall structure and entry points to the networks of primary documents) (Conklin 1987). Structural (or organizational) links differ from associative (or referential) relationships in that "the former connect information pieces within the same entity instance, while the latter interconnect different entity instances belonging, in most cases, to different entity classes" (Isakowitz, Stohr and Balasubramanian 1995).

The awareness of the relationships between the elementary, composite and organizational units of hypertext is particularly significant in browsing, which is "the typical means for accessing information in multimedia documents for both readers and authors" (Foss 1989). The next section will discuss in greater detail some aspects of this activity.

3.2 Browsing - a Useful Approach to Information Seeking in Hypertext Environments

In view of the problems in using the search engines, described in Chapter 2, there are a number of reasons that make browsing an important method of information

seeking on the WWW:

- It can be used by anybody with very little training.

- It is the way to search the subject indexes and guides.

- As elaborated upon in sections **2.3.2.3.2.1**, **2.3.3.1** and **2.3.4.3**, extensive knowledge of each search engine's indexing and database characteristics, as well as great skill are needed for users to be able to formulate and refine their searches. End-users and even inexperienced intermediaries usually cannot go much farther beyond simple searches with the WWW engines. These produce results that often contain just entry points to sites and directories that need to be browsed.

- Browsing is the major road to serendipitous discovery.

- It gives an idea of the database's content and organization, of the concepts related to the search area, and enables the searcher to redefine his original goals if necessary (Foss 1989).

- *It requires a smaller cognitive load* (straining of mental capacities) than the conceptual analysis and query formulation, necessary for the search engines, does (Marchionini 1995: 103)

On the other hand, as an informal and opportunistic approach to information seeking, browsing is heavily dependent on the information environment, characterized in electronic systems by "the blurring of boundaries between

document collections or databases and discrete documents or records". This situation is further "exacerbated in networked environments in which documents from different databases appear in the same window or the same document is represented in different databases" (Marchionini 1995: 100-101) and becomes a cause for disorientation and cognitive load in its own right.

3.2.1 Cognitive Load and Disorientation - the Main Obstacles to Effective WWW Browsing

Conklin (1987) defines **cognitive overhead** (synonymous with cognitive load, overload) as "the additional effort and concentration necessary to maintain several tasks or trails at one time". Together with **disorientation**, "the tendency to lose one's sense of location and direction in a nonlinear document", cognitive overhead constitutes the main obstacle to effective browsing of hypertext, especially in large and interwoven databases as those on the WWW.

Like all text, hypertext is written in order to be read and understood. For the latter to happen the author's intentions and meanings encoded in the text must be, in the first place, correctly conveyed and then interpreted by the reader. Comprehension of linguistic descriptions leads to the construction of mental representations or models, based on the meanings from the text and general knowledge (Eysenck and Keane 1995: 415-416). The mental effort with which this is done is a measure of the readability of a document (Thüring, Hannemann and Haake 1995).

Writing techniques which enhance the readability of a document in sequential reading apply to hypertext too. However, it is necessary to discover what novel

aspects of the WWW environment might induce digital librarians, who are building their collections, and educationists, who are producing WWW courseware for various subjects, to create communication and navigation barriers causing cognitive overload and disorientation and impeding the progress of the information seeker.

Versatile graphical browsers, such as Mosaic, Netscape and Internet Explorer, have greatly improved navigation on the WWW. Nevertheless, even experienced users are still overwhelmed by the infinity and disparity of cyberspace resources. Cognitive overhead and the sense of disorientation are chiefly the result of the kaleidoscopic content nature, the freeing organization of hypertext and the varied mix of presentation standards and styles on the WWW. In addition to that, various authors' intentions, purposes and meanings have to be interpreted simultaneously or in succession and according to changing perspectives as readers follow different paths of interlinked nodes in the multi-dimensional dynamic world of the Web. Yet, issues of human comprehension, interpretation and learning "are widely neglected and barely influence hypermedia design" (Thüring, Hannemann and Haake 1995).

Even less attention has been paid to the effects on the users' success in information seeking on the WWW of the original and intermediary authors' decisions on segmenting, linking and structuring the information in their documents. As Van House *et al.* (1996) point out, the effective and easy use of a system by its target users, referred to as usability, has been at the centre of research mainly of the human-computer interaction (HCI) community (Nielsen 1990; Catledge and Pitkow 1995, Cockburn and Jones 1996), whose ultimate goal is to improve computer interface design. In contrast to that, a LIS approach

to the problems of improving the usability of the WWW as an information seeking tool was adopted for the study reported below. It focuses on the information seekers in their interaction with the intermediary authoring products on the WWW, rather than on interface design.

3.3 Background of the Study of Authoring for Comprehension through Simulations on a WWW Model

The experimental part of this research set out to explore the extent to which the length and modularity of hypertext documents on the one hand, and the configuration of internal and external hypertext links on the other, influence the outcome of browsing for information on the WWW. The first aspect of this research problem - the effect of the length and modularity of hypertext documents on browsing outcomes, centres on the role of WWW individual document characteristics in the author-reader communication process. The second aspect - the effect of the configuration of internal and external hypertext links on browsing outcomes, highlights the role of the authors' document structuring and organization patterns in building collections in the same process. And the third aspect is the interaction between the above two aspects and the browsing strategies, employed by the information seekers.

WWW digital librarians have accumulated both direct and indirect impressions of the factors which have a negative effect on the Web's usability by contributing to the increase of cognitive overhead and disorientation in information seeking in the real environment. However, to study these factors in a systematic way and under equal conditions at all instances and for all searchers, a researcher needs a method of enquiry that would allow most of the variables of this complex

information system to be controlled.

Catledge and Pitkow (1995) emphasize the growing need for a sophisticated understanding of the WWW audience, going deeper than demographics and certain behavioural high level trends in the system's use. They focus on actual user behaviour, as determined from client-side log file analysis, to study WWW navigation strategies and interface usage.

All user interface level events can be captured in client-side log files. The method allows for an easy accumulation and processing of huge amounts of data, which is a prerequisite to the generation of statistically conclusive results. Problems arise, however, with user behaviour which cannot be monitored by computers. Catledge and Pitkow (1995), for example, found it necessary to determine session durations artificially, since users were found to be leaving their browsers running for long periods of time without interacting with them.

Other factors, influencing human actions, decisions and choices, will also remain uncaptured, since the variables affecting search behaviour in electronic systems are manifold: "individual searcher characteristics or performance, system characteristics or performance, task characteristics, and setting" (Lin, Liebscher and Marchionini 1991).

Quantifying variables, related to human cognition, learning and search behaviour patterns is an enormous challenge. Furthermore, according to Lin, Liebscher and Marchionini (1991) representations of electronic searches require:

- the definition of *a state space that accounts for the search process, i.e.*

each move by the searcher results in a transformation from one system state to another

- *collecting data (e.g. keystroke logs, verbal protocols, etc.)*
- *mapping data to search states*
- *displaying sequences of search states*

From the user's point of view, states represent typically conceptual situations that require a decision or action. Because of the lack of standards as to the nature or coding of states for electronic search systems, "state spaces must be handcrafted for each system and each set of research questions".

Various aspects of complex information systems, of which information seeking is an integral part, have been studied, emphasizing first the input processes and later the output processes, users and interactions (Saracevic *et al.* 1988). Information science currently views information seeking as "a problem-solving activity that depends on communication acts" and emphasizes "the needs, characteristics, and actions of information seekers" rather than focusing exclusively on information systems (Marchionini 1995: 28-29). It has come to rely more and more on notions and approaches borrowed from cognitive science. "The key to the future of information systems and searching processes" is seen by Saracevic *et al.* (1988) "not in increased sophistication of technology, but in increased understanding of human involvement with information".

A major characteristic of information seeking in today's systems, and on the

WWW in particular, is its interactive nature. Therefore a study of information seeking on the WWW needs to address "the processes and patterns of interaction rather than only the products of such interactions" (Lin, Liebscher, and Marchionini 1991). For a new information system, such as the WWW though, "understanding the interrelationships between the parts of the information system" (Oswitch 1983) is a prerequisite to investigating the users' interactions with it. A simulation model capturing the essential elements of the WWW for the purposes of the problem under study and serving as a testbed for the information seeking experiment is the best answer to these requirements.

3.3.1 Modelling and Simulations

Reed (1976) defines **model** as "the representation of the system - the dynamic analogue, its rules and relationships", and **simulation** as "the exercise of the model under various specific policies or conditions". Simulations, and especially those using computers, have become a very powerful research tool in both natural and social science. By studying the response of a well-defined model to different inputs and variation of parameters they enable the researcher to gain insight into the behaviour of complex systems.

Simulation in library and information science exhibits the main features inherent in this tool:

- A well-defined model of the processes under investigation must exist as a precondition for the simulation. This model must be sufficiently simple to implement while still revealing the most important aspects of the interactions observed in the real-life processes.

- The model must be operational which means that it must possess clearly defined input and output parameters so that it will be relatively easy to drive it, as well as to monitor its performance.

The factors of information seeking, described by Marchionini (1995: 32), include *the information seeker,* whose information need leads to the verbalization or performance of the search *task*, the *search system*, which is the source of information and the rules for access, the *domain*, representing the fields of knowledge, the search *outcomes* - "the feedback from the system", and the all-encompassing *setting*. While these factors are consistent with the various models of communication, the absence of the author as the source of information is notable.

The reason why authors were not among the factors highlighted by studies of information seeking in electronic environments is that "all of these studies were performed on closed, single-author systems" (Catledge and Pitkow 1995). Since the WWW is "an open, collaborative and exceedingly dynamic hypermedia system", the effect of Web authoring, with its diversity in perspectives and styles, on all stages of the communication process rises dramatically and can no longer be regarded as a constant. This is particularly true in browsing which, as noted in **3.2**, is heavily dependent on the information environment

In order to investigate the WWW-mediated communication process in its systemic entirety of input, process, output and feedback, this thesis modified Saracevic *et al.*'s (1988) general model of information seeking and retrieving, which is automated searching oriented, and focused on the following classes of

variables, involved in browsing on the WWW: authors, questions, browsers, browsing sessions, Web documents, target documents reached.

The major reason why a model is preferable to the real environment for the study of authoring effects on information seeking on the WWW is that there is no way to restrict or channel the user's moves in accordance with a set goal through the infinity of the real cyberspce. Furthermore, from the user's point of view the distributed storage of information and the client-server model are an extraneous factor interfering through unpredictable document loading times and delays, as well as dead links and server failures, with the information seeking process.

Therefore, for the study of the WWW author-information seeker interactions, a simplified model comprising a well-developed set of locally-stored HTML documents with disactivated external links was created and given the name of **MicroWeb**.

3.3.2 The Pre-History of the MicroWeb Project

Since the writer had set out to investigate how authoring decisions will affect users' information seeking on the WWW it was deemed appropriate that she should go through a genuine authoring process of constructing a Web space. As is the case with all creative processes, the decision on where to start was not an easy one, especially in view of the fact that the task required capturing the emergence of a new information world.

In 1994, when this research project was started, the WWW was just out of its infancy and still waiting to embark on its way to becoming in April 1995 the most

popular service on the global information superhighway (see **Fig. 1-2**, Hobbes 1996). Very little information about it was obtainable locally. Finding exhaustive information about WWW research on the Internet was also difficult because of the huge problems posed to organization and bibliographic control in this vast, decentralized, dynamic entity and the imperfections of human and computer mediated efforts at facilitating information retrieval. Only late in November 1994, when the first experimental WWW server was installed on the Pietermaritzburg campus of the University of Natal, did the writer get her first chance to try her hand at WWW authoring (Horton and Ilcheva 1995).

It was not long before it was realized, however, that similarly to all other research, the Internet provides the opportunity of speeding up the grasp of its own development by enabling a researcher to join a newly formed community of experts and get a closer look into the process of the making of a new science field. Finding an appropriate Internet discussion group and analysing the content of its members' contributions over a considerable period of time seemed to offer the fastest way of gaining a comprehensive view of the theoretical and practical issues facing the WWW at its early stage of development . The choice fell on Web4Lib, the first electronic library Web managers' discussion forum which was created in May 1994 (Tennant 1996b).

3.3.2.1 Web4Lib as an Authoritative Source of Comprehensive and Up-To-Date Information on the Development of the WWW

Using Magellan's criteria for rating listserv discussion groups, which measure depth of content, signal-to-noise ratio, and amount of traffic (Tillman 1996) it will

be seen that Web4Lib is a highly authoritative resource:

- Depth of content

The purpose for which Web4Lib was set up is to discuss *the creation and management of library-based World-Wide Web servers and clients. Particularly appropriate issues for discussion include, but are not limited to:*

1. Web resource selection and information mounting in relation to existing acquisition and collection development procedures
2. Cataloguing and metadata issues regarding Web information
3. In-house patron access to Web servers (e.g., Netscape on patron-accessible computers)
4. Training staff or users to use the Web or to create Web resources (SUNSITE Manager 1996)

From the above, one can conclude that Web4Lib's topics do not deal with everyday practical and technical concerns alone but with fundamental issues as well. These are discussed in depth and from different perspectives. The most active participants in the Web4Lib discussions are among the world's leading library Web managing professionals who have authored reports at scientific workshops, conferences and symposia.

The depth of content of the Web4Lib discussion is reflected in the creation of The Library Web Manager's Reference Centre at the University of California Berkeley Digital Library SUNSITE "first to provide easy access

to some of the most informative Web4Lib postings, then later to also point to key resources for library Web managers" (Tennant 1996a).

- Signal-to-noise ratio

  This measures how much of the discussion conforms to the group's purpose and ultimately - the usefulness of the discussion to its members. Although sometimes the participants get carried away and produce threads of over sixty messages, their contributions are usually to the point, very informative and helpful which is proved by the growing membership of the forum.

- Amount of traffic

  The Web4Lib membership includes individuals from more than 40 countries and is growing at ever faster rates - a gain of approximately 200 subscribers in less than two weeks was recently experienced to bring the total to 3175 (Tennant 1996c). The group is very active, producing an average of 10 messages every day.

In addition to being authoritative and prolific, Web4Lib makes a very exciting reading inviting rereading all the time. The author's e-mail box was soon overflowing with hundreds of messages that she felt the urge to keep for later reference. But the growing bulk of unorganized material was becoming too difficult to use in addition to obscuring mail from other sources. It was time to have a look at the Web4Lib archives.

3.3.2.2 The Web4Lib Archives

The author discovered addresses for three sites maintaining archives of the Web4Lib list. The one at the US Medical College of Wisconsin Libraries claimed to be in hypertext and was immediately selected for investigation. The first impressions were quite disappointing. Automated hypertext archiving was found to be very far from the real idea and potential of hypertext for the following reasons:

- It sorts messages by thread, date, subject, author in the form of bulky index lists which take a long time to load.

- The messages by subject list follows an alphabetical arrangement of the authors' subject lines - browsing it involves going back and forth between the index and the messages and is also very time consuming and frustrating because quite often one has to wait for the whole list to reload again.

- The subject lines of the messages, as formulated by their authors, do not always reflect clearly or meaningfully the content of the messages. Sometimes they get changed by others continuing the discussion thread, and sometimes they may even be missing. Auxiliary and unhelpful words in the beginning of the subject line, such as the indefinite article and "another" for example, influence the position of the subject line on the index. Ultimately, the lack of controlled vocabulary and the alphabetical ordering lead to arbitrary scattering of related information.

- On returning to the index there is no indication of which was the last

message selected and the user has to rely on his memory for orientation.

This last frustration might have been the fault of the text browser, Lynx, that the author was using at the time. The text browser proved even less capable of representing the frames arrangement of information on the authoritative Web4Lib archive maintained by the hosts of the discussion group at the University of California Berkeley Digital Library SUNSITE (http://sunsite.berkeley.edu/Web4Lib/archive.html). The only choice left was to start creating one's own "memex".

From there on it was but one step to the decision to use the most substantial, informative and interesting threads from the discussion as one of the sources of material for the construction of the simulation model. In the course of sixteen months - from March 1995 to July 1996 - 372 Web4Lib messages were very carefully selected, organized and stored in 23 directories and were ready to be transformed into a coherent hypertext whole.

3.4. The Construction of MicroWeb

Even though it was conceived as a research tool, from the very outset **MicroWeb** was planned to follow the principles governing the creation of a real Web space. In order to produce any coherent and useful publication, a WWW one included, an author needs to have:

- A constructive purpose making an original contribution to meeting a social need

- A clearly envisaged target audience

- A relevant and effectively conveyed content

According to a cognitive model of the communication process writing and reading are reverse processes of encoding into sequences (which have to be decoded in reading) initially loose networks of information pieces, organized in hierarchies (Rada 1989). Thus the process of writing starts from networks and goes through hierarchy to sequence, and that of reading starts from sequence and goes through hierarchy to network, as shown below:

NETWORK ⇆ HIERARCHY ⇆ SEQUENCE

As for navigational design, which is indispensable in any hypertext system, there are two approaches to choose from: bottom-up, starting from the primary building blocks (entities) and then devising "the more general access structures" and the

inverse "top-down" one (Isakowitz, Stohr and Balasubramanian 1995).

A major feature of the WWW that needed to be preserved in **MicroWeb** was the former's multi-author diversity of content, structure and presentation. Therefore a bottom-up design approach was adopted to enable the incorporation of other authors' structures in the design of the model. In sections **3.4.1** through **3.4.6** the steps of this design will be described.

3.4.1 The Design Principles Followed in Preparing MicroWeb's Individual Documents

As a first step towards achieving the experimental objective of this research, namely to establish the extent to which the length and modularity of hypertext documents on the one hand, and the configuration of internal and external hypertext links on the other, influence the outcome of browsing for information on the WWW, three types of individual documents - original, customized and borrowed, were prepared as building units for the model. These types required different approaches and degrees of treatment in their design.

Original documents

A few original documents were created to meet the targeted audience's needs, as perceived by the researcher, for information specially presented and structured to motivate and guide the reader in using **MicroWeb**. Several important considerations, specified by Isakowitz, Stohr and Balasubramanian (1995), need to be borne in mind in designing original or customized hypertext documents,

namely:

- *dividing an entity into slices*
- *choosing one slice to be the head of the entity*
- *interconnecting the various slices*
- *labeling the links*

The length of elementary units and modules in **MicroWeb**'s original documents was determined by the requirement that each document must make a meaningful unit of retrieval of information, as well as by the position of the document in the overall structure (Berners-Lee 1993). Most of the original documents were clustered around the homepage in an initial position and consequently had to be kept short to allow novices to get used to moving within and across documents. "Back to homepage" links were provided at their ends to support the widespread use of the "spoke and hub" browsing strategy noted by Catledge and Pitkow (1995). A **MicroWeb** graphic logo was added to indicate the documents' authoring and thematic unity, thus providing global context, which is one of the great challenges to the WWW author/designer (Kahn 1995).

Customized documents

The Web4Lib discussion messages represent the bulk of borrowed elementary units in the model because of the value of debates in collecting and ordering "the best available evidence on each topic" (Lowe 1985). They are, by origin, ready-

made slices of entities already united in threads during the discussion. However, due to problems with the subject line as the sole means of labelling organizational links (discussed in section **3.3.2.2**), the latter indicator of thematic unity needs to be double-checked in following the discussion live, and even more so in archival retrieval.

Therefore, the Web4Lib messages were subjected to extensive adaptation to achieve a structural complexity reflective of the rich semantic relationships typical of a discussion, which are not catered for by automatic archive mailers, as well as consistent presentation. Discussion messages also lend themselves to a higher-level, issue-based-information-system type of classification of units, as described by Conklin (1987), under the categories of issues, positions, arguments, assessments.

Some examples of the semantic relationships, which formed the basis of grouping linked Web4Lib messages into major and branching paths, are: chronological, cross-reference, question and answer, elaboration, objection, redirection, clarification, reaffirmation, new perspective. Navigational aids, such as typed links (Thüring, Hannemann and Haake 1995), to reflect the above-mentioned issue-based-information-system type of classification categories, might be considered in an environment like this. However, that has not been practised on the WWW so far.

For the sake of presentation consistency, e-mail data in the postings' heads, which are quite often longer than their bodies and extremely distracting to novice and casual users, were cut down to two items (date and author's name) alone. A Web4Lib graphic logo was added to recognize the discussion group's intellectual

property of the documents, as well as to serve as a contextual and navigational signpost, alerting the reader that he is in the discussion environment.

Borrowed documents

In contrast to the customized discussion messages, which the researcher threaded together in multiple structures, long and complex WWW documents, including composite units and even whole sites, were borrowed in full, together with their originally imposed structure, in order to incorporate the real authoring diversity feature of the Web. Both the local and global structural links and overview navigational pages of the borrowed composite nodes and sites were preserved, with certain associative links disactivated internally and others added in accordance with the model's real-life purpose and target audience as set out below.

3.4.2 MicroWeb's Real-Life Purpose and Target Audience

In order to model the author-reader communication on the real Web and thus serve its purpose as a research tool, **MicroWeb** had to have a real-life purpose. Consequently, it was planned to also serve as a hands-on tool for self or peer training in the use of the WWW and the Internet of librarians who are new to the environment. There are a number of reasons why the various tutorials of this type on the Internet itself cannot be relied upon to make a difference in the South African libraries.

While the Internet and the WWW have been introduced at some of the tertiary institutions (Van Brakel 1994; Horton and Ilcheva 1995) and libraries in South

Africa, the overwhelming majority of the librarians in this country have no or only a limited knowledge about them (Kaniki 1996). For librarians to start on the road to providing digital library services, helping educationists to integrate the Internet into the learning process and training end-users, they need to quickly get a solid theoretical background as well as practical skills in the use of the new technology. It will take many years before this happens on any significant scale if librarians do not resort to efficient self and peer training.

The following factors, identified by Kaniki (1996), stand in the way of this process:

- Insufficient equipment

- Lack of time and motivation

A considerable number of the academic librarians in Kwa Zulu Natal, for instance, have to share workstations. They are not given any motivation or extra time to devote to the Internet. They are practically unaware of the potential of the Internet as a learning tool for professional development. According to the answers in Kaniki's (*ibid.*) survey they rank the difficulties they face in using the Internet in the following order:

- *Slowness of the system*

- *Frequency of downtime of the system*

- *The time it takes to access and search for relevant information (too time*

consuming)

- *Technical difficulties in accessing the Internet*
- *Lack of clarity about the pathways to follow to access relevant resources*
- *Too much "unnecessary" stuff*

The first four problems are easily taken care of by a PC loadable model combined with a stand-alone WWW browser. The last two are solved by making sure that **MicroWeb** is made up of well organized quality material that is relevant to the needs of the target audience.

3.4.3 The Content of MicroWeb's Documents

MicroWeb is targeted at people who might be feeling somewhat threatened by what it stands for. The author needed to do her best then to attract the users to the new medium by selecting information that would be both instructive and exciting, challenging but not too difficult to understand and apply. It was realized that the users would be made to feel at home with the WWW much sooner if a bridge was provided between the new technology and the traditional principles and concerns of librarianship.

To be a true representation of the WWW, **MicroWeb** had to include both the formal and informal styles and the varying look of different types of Internet

documents, such as:

- Homepage
- Contents page indexes - **MicroWeb**'s contents page and the BACK 2 SCHOOL Internet tutorial contents page
- FAQs - The Library Web Manager's Reference Center homepage, mentioned in **3.3.2.1**, and the WWW FAQ page by T. Boutell
- Electronic discussion messages as received straight from the Web4Lib list or retrieved from the Web4Lib archive - an example of the latter format is the Search Engines' Reference List linked to The Library Web Manager's Reference Center homepage mentioned above
- Internet tutorials
- Internet and WWW glossaries and history lists
- Research articles and presentations

The homepage is like the first sentence of a book - it must capture the attention and set the stage for what is to come. The **MicroWeb**'s homepage welcomes the user, relaxing and encouraging the novice to enjoy the novelty of hypertext, and points to the contents page. An optional tutorial, created by the author, is available to teach users with no WWW experience basic graphical browser navigation skills. Key concepts and terminology are covered at progressive

lengths and depths by the inclusion of several Internet tutorials (distributed for free on the Web), designed for novice end users and librarians. Together with the glossaries, the tutorials can be used as reference resources to help read the Web4Lib discussion - in the author's opinion new concepts and terms are best learned in context, especially in an informal conversational setting.

To guard against any copyright or other objections, a message was posted to the Web4Lib subscribers and manager explaining the purpose of **MicroWeb** and asking their permission to use some of the discussion material for the construction of the model. The request met with enthusiastic approval (see **Appendix A**). The next step before proceeding with the construction was to devise a method of incorporating the experimental research problem in the structure of **MicroWeb**.

3.4.4 The Incorporation of the Experimental Research Problem in the Grouping of MicroWeb's Individual Documents

In addition to **MicroWeb**'s instructional purpose, the experimental research problem of establishing the extent to which the length and modularity of hypertext documents on the one hand, and the configuration of internal and external hypertext links on the other, influence the outcome of browsing for information on the WWW, had a particularly pronounced effect on the choice and grouping of the model's individual documents.

The first aspect of the experimental research problem, under study in this thesis, concerns the length and complexity of WWW documents. It is generally assumed that long and complex WWW documents create cognitive overload and cause disorientation in the Web space, which impedes the navigation process. It was

decided to test the validity of this assumption, often to be encountered in style guides on WWW authoring, urging authors to make their pages short and modular.

There are some very strong arguments in favour of the recommended WWW authoring style, as Lynch (1996) and Tilton (1996) point out:

- The need for logical decomposition into units of information allowing branching from the linear structure

- Preservation of local context and entry points to linked pages which tend to be lost in scrolling

- Reduction of loading times for individual pages

- Separation of supplementary information from the main flow

However, extreme segmentation may lead to fragmentation, "a lack of interpretive context" (Thüring, Hannemann and Haake 1995), and an increase in distraction and cognitive overload because of having to make more choices and wait for more links between related pieces of information to perform. It does not seem to have been scientifically proved that recommendations, such as Lynch's (1996) of "two or three screens" at the most, represent the maximum desirable or optimum length for individual WWW documents. For this reason it was decided to design **MicroWeb** in a way that will allow to test how the lengthiness or shortness, and complexity or simplicity of WWW individual documents will affect the success of information seeking through browsing.

Since many trails (or paths) of links can lead to the same document in hypertext, it is possible to make some of these paths go through sets of longer and complex pages and others - through sets of shorter and modular or elementary pages. Then the researcher can monitor the information seeker's progress through the various paths and compare the time and effort taken in reaching the destination via the different routes.

The answers to FAQs about the WWW are a source of short and elementary pages. They try to answer one question, and they cannot be interconnected heavily because they deal mostly with specific technical aspects of the Web which, in the **MicroWeb** context, cannot be easily grouped together in larger classes.

In contrast to that, the Web4Lib messages are a source of short, but modular units. They can be linked together in networks in accordance with all sorts of relationships inherent in any discussion. They can also become part of hierarchical structures due to the conceptual and theoretical relationships of the common subject. Furthermore, the emphasis of discussions is not so much on concepts and names, which are the more common targets of cross-references, but on ideas which is another criterion of grouping modular units together. As a result, the potential for branching in all directions within and across the original threads of the discussion is very high, indeed.

Tutorials, research articles and presentations, whose complexity allows rich cross-referencing and multiple structures, were suitable for linking into paths of larger documents leading to the destination.

The collected material was reviewed and its usefulness and interest assessed in the framework of the collection. Less significant documents and directories were discarded, and new documents retrieved or created to fill identified gaps. Finally 330 HTML files and hundreds of image files, totalling just over 3 MB and divided between fourteen directories, were interlinked in a tightly woven Web space which needed an organized overview of its contents to provide entry points for the reader.

3.4.5 MicroWeb's Top Level Organization of Material

According to the cognitive model of writing/reading, mentioned in **3.3.3**, hierarchies play a central part in conveying meaning or making sense of received information in communication. Hierarchical information is implemented by structural links, which are often concentrated in organizational nodes, such as contents pages (Conklin 1987).

MicroWeb's home contents page provides a hierarchical organization of the material of the collection around four thematic sections. The first one explains the dual purpose of the model as an instructional and research tool; the second one contains reference material helping with the understanding of WWW terms, concepts and applications; the third one presents an overview of the role of librarians in the information technology age, and the fourth one focuses on various more traditional library approaches and initiatives on the Internet. The themes are presented as headings, accompanied by bulleted lists of subheadings (see **Appendix B**).

3.4.6 MicroWeb's Overall Research Design Structure

For the purposes of the information seeking tasks and simulations three target documents were positioned in the model as a destination to multiple paths going through sets of short-elementary, short-modular or long and complex documents:

- The answers to the final examination of the BACK 2 SCHOOL Internet tutorial, to be referred to as Target 1 (http://zorba.phy.unp.ac.za/micro.web/TUT/answers.htm) (or see **Appendix C**)

- A conference paper on Yahoo, to be referred to as Target 2 (http://zorba.phy.unp.ac.za/micro.web/WEB/YAHOO.HTM) (or see **Appendix D**)

- A discussion of the differences between the WWW search engines and subject indexes, to be referred to as Target 3 (http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/answers.html) (or see **Appendix E**)

All the paths converging on each of these three targets had to go through the navigational pages, that they were originally linked to, before reaching their destination. The navigational pages (which are basically lists of organizational links) were as follows:

- For Target 1 - the BACK 2 SCHOOL contents page, to be referred to as **Nav. 1** (http://zorba.phy.unp.ac.za/micro.web/TUT/fall.htm)

(or see **Appendix F**)

- For Target 2 - The Library Web Manager's Reference Center homepage, to be referred to as **Nav. 2** (http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/FAQ.HTM) (or see **Appendix G**) and the Search Engines' Reference List, to be referred to as **Nav. 2a** (http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/0103.htm) (or see **Appendix H**)

- For Target 3 - the same as for Target 2

In a somewhat simplified way this structure is illustrated on **Fig. 3-1**. The different sizes of the ovals and rectangles drawn between **MicroWeb**'s contents page and the navigational and target documents symbolize the variety and the complexity of the hundreds of documents comprising the **MicroWeb** space. The actual structure of the links between the documents is, of course, too complicated to be shown in detail. Therefore, only an idea of the connections leading to and from the search destinations is given. Note that only links from Nav. 1 and Nav. 2a lead to the target documents.

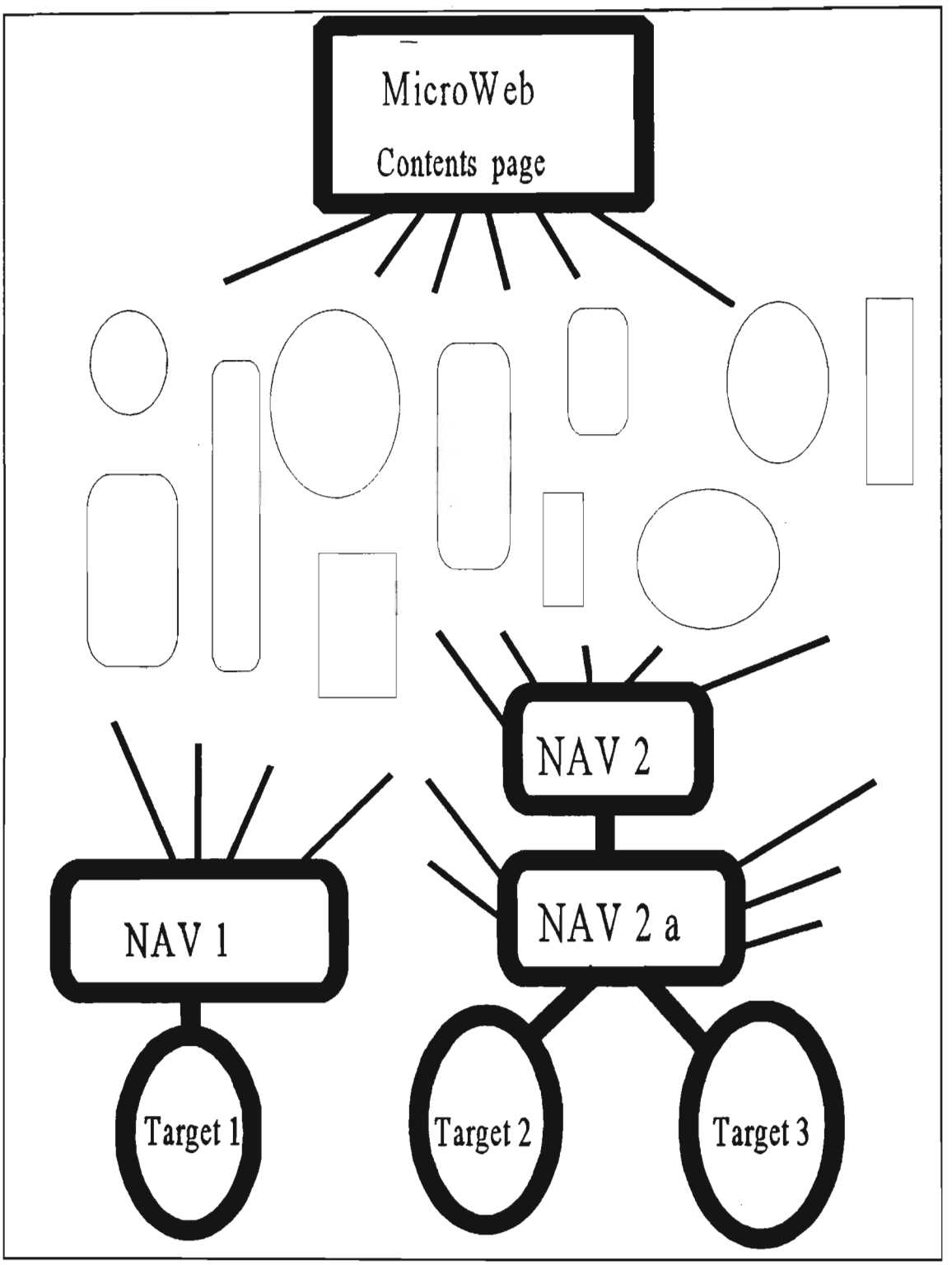


Fig. 3-1 The overall research design structure of the model.

3.5 The Simulations Strategy

For the execution of the simulations the model needs to be stored on a hard drive and read by a WWW graphical browser (Mosaic or Netscape) in a standalone mode. The author used NCSA Mosaic.

A considerable effort was made to choose a programme for automatic log-file creation which could store all the relevant information for a long browsing session on the model. There are many tools which monitor and even analyse WWW statistics but it soon became clear that none of them is suitable for this particular research project. Here a much more detailed and diverse information on a browsing session was needed to clarify the intents and interpretations of the seekers and their resulting searching strategies behind their navigational moves. That is why, like in other studies of search behaviour in electronic environments, observer notes had to be kept manually (Lin, Liebscher and Marchionini 1991).

As Marchionini (1995: 100-102) points out, unlike automated searching, browsing integrates the information-seeking subprocesses of problem definition, query formulation, execution, examination and iteration. The reasons for this are:

- *The parallel engagement of physical, perceptual and cognitive processes*
- *The close coupling between the information seeker and the organizations and representations provided by the information environment*
- *The open-ended information problems that lend themselves to browsing*

Browsing reflects the behaviour of people in using certain strategies in information seeking on two levels: across documents and within documents. Across-document browsing involves navigation through records in order to find items relevant to the user's information need for closer inspection. Within-document browsing aims at grasping the essence of a document, using the skimming strategy, or locating a specific term or piece of information, using the scanning strategy.

The second aspect of the experimental research problem, investigated in this study, which centres on the configuration of internal and external hypertext links, has been addressed by an attempt to compare across-document to within-document browsing patterns and to reveal any possible effects of the latter on the information seeking outcomes.

To do this it is not enough to gather just keystroke data in accordance with the User Interface Design Environment (UIDE) guidelines, used by Catledge and Pitkow (1995). Moreover, these authors have proved that selection of anchors and backtracking account for nearly 93% of all the UIDE user interface level moves (also including "open URL", "Hotlist - Go to", "Forward", "Open Local File", "Home Document", "Window History"). Therefore it was necessary to study these two major navigational moves in greater detail, gathering data on: the nature (structural or referential) and the position within documents of links followed, the purpose of backtracking - for a fresh start, for internal path re-exploration or branching. Other data collected included the identity of documents visited, time spent in a document, what the subject was actually doing once in a given document, how the subject read the document, what the subject's behavioural and emotional patterns were, what happened when they hit the

navigational pages and targets.

All these are believed to be of importance for the final results of the author-seeker interaction. But before the browsing simulations could start at all, it was necessary to prepare the setting and find the right subjects, who could cope with what was required of them to do.

3.5.1 Selection and Preparation of Subjects for the Experiment

Logistical constraints did not allow the experiment to be conducted in a public information system setting, such as a library or a university department. This was one of the reasons which made it difficult to find volunteers to take part in it. Another difficulty had to do with the lack of experience of the Internet among end-users and even LIS specialists and with people feeling intimidated by the new environment.

Yet another problem, which ultimately reduces the number of participating subjects, seems to affect all research into information seeking. This is the challenge of quantifying variables, related to human cognition, learning and search behaviour patterns and their representation, as explained in **3.3**. The effect is compounded by the effort and time-consuming procedures of gathering, processing and analysing the data manually.

In addition to those embracing the "discount engineering usability method", which is based on observations of "two users or so" (Nielsen 1990), all authoritative user-centred experimental investigations of information seeking in electronic environments, reported in the literature and studied in preparation for

this research, have made use of a restricted number of subjects. Here are a few examples:

Marchionini *et al.* (1993) - 8 subjects
Cockburn and Jones (1996) - 11 subjects
Cove and Walsh (1988) - 13 subjects
Liebscher and Marchionini (1988) - 26 subjects
Borgman, Case and Meadow (1989) - 28 subjects
Marchionini (1989) - 52 subjects

Even "the largest study to date of online searching by expert intermediaries" (Marchionini 1995), conducted by Saracevic *et al.* (1988) used 79 subjects altogether, with 9 subjects conducting a search each on the same question.

Fourteen subjects, all university graduates from various fields, took part in the **MicroWeb** experiment. All of them had varying degrees of computer literacy and experience in on-line searching. Four had had no exposure to the WWW or the Internet, five were occasional users and five made daily use of the Internet. Since this thesis represents an exploratory research in a new, and largely unstudied, area, it is believed that this group is sufficiently representative for the purposes of the study.

The four novice and four of the casual users performed a search each on various simple questions related to one target document. The five casual and the five experienced users conducted a search each on the same set of questions connected to another target document. Two experienced and one occasional user conducted a search each on the same set of questions connected to a third target

document.

Despite assurances that the outcome of their searching would not affect the success of the research, subjects felt nervous prior to the experiment. They were told that they had total control over how to search and when to stop, yet they perceived the experiment as a challenge and were obviously bracing themselves to succeed.

For this reason before doing anything else the researcher gave the subject a few minutes to go through a simple tutorial, which was specially created by her to teach novice users how to navigate in **MicroWeb**. Fifteen minutes with the tutorial enabled even subjects who had never used a graphical interface before to cope with the search. A look at the tutorial gave the experienced users a chance to relax and get used to the model.

In addition to being convinced that the experiment required only a few basic skills, all subjects got the idea, purposefully implied through the tutorial, and verbally reinforced by the researcher, that they were restricted in their navigation options to following links and going back a step at a time. They were not shown any of the possible ways of jumping (such as through the *History* list, the *Home* menu option or button, *Hotlist*, opening *Local File*) to avoid confusing and disorienting the less experienced ones. The *Forward* navigation operation was also excluded as it is one of the particularly confusing of the WWW browsers' features requiring user-interface adjustment. No mention was made of the *Find* option on the *Edit* menu to prevent its distracting subjects from browsing.

3.5.2 The Information Seeking Tasks and Sessions

The subjects were given questions, representing their simulated information need, to answer by choosing an entry point in the **MicroWeb**'s contents page and then browsing through the opening paths. They were not told that there were target documents which contained all the information they needed to answer their questions. In this way they were allowed to decide on whether or not the outcomes of their searching were satisfactory. The questions were carefully chosen or formulated so that there would not be an immediate match on the contents page. By engaging the subjects in an informal chat, the researcher made sure that they could not answer these questions from their own knowledge.

The researcher asked the subjects to search for the needed information as they would do in real life if they had a resource like **MicroWeb** at home. It was up to them to decide how to search, whether they were satisfied with the answers found, and when to stop. The researcher was there to help with any unforseen crises - e.g. users rarely wait until all the graphic images have been loaded by Mosaic and by trying to work with the document they interrupt the loading process. Otherwise the researcher tried to be as unobtrusive as possible.

The researcher observed and recorded all the data necessary to reproduce the whole session if desired. The duration of each session was measured and verbal and behavioural clues pointing to the subjects' feelings, comprehension, choice of information seeking strategies, winning tactics and problems were monitored and recorded.

3.5.2.1 The Target 1 Task Experiment

The Target 1 task simulation was used as a pilot experiment to test the functionality of the model and particularly the suitability of different classes of questions for the purposes of the research. Saracevic *et al.* (1988) classify questions into five categories, according to domain, clarity, specificity, complexity and presupposition. Of these specificity and complexity assumed a special importance in this study, since it was not certain how novice users would cope with browsing for answers to even simple questions.

On the other hand, it was difficult to foresee whether, in a specific domain focused collection, such as **MicroWeb**'s, answers to simple or general questions would be limited to the target documents, thus encouraging the subject to persist following paths to the destination. Formulating a specific, but complex question challenges the researcher to build paths to a document discussing in detail a narrow-focused topic avoiding the provision of direct explicit links to that topic. This requires the availability of a specific topic that relates to a central issue. The latter would be likely to be referred to often in many documents, thus lending itself to rich cross-referencing allowing the construction of multiple paths towards the target.

Four novice and four occasional users took part in the Target 1 task simulation. The duration of their sessions ranged between 9 and 55 minutes - an average of 24'30" per session.

The subjects were given simple questions, requiring only one answer from a multiple choice or true/false selection (see **Appendix I**). The questions, based on

the material taught in the 28 lessons of the BACK 2 SCHOOL Internet tutorial, are part of the "final exam" included as "Lesson 29" in the tutorial and its index, (Nav. 1). On opening Lesson 29 the reader can see that there is a link to the answers of the questions in this exam (Target 1).

Paths towards Target 1 were built through long and complex documents (individual or composite nodes containing Internet tutorials). The author did not provide direct links from the **MicroWeb** contents page or from any other page to Nav. 1. However, Nav. 1 was available through "Back to Index" links, provided by the original author and kept functional by the researcher, at the end of each lesson in the tutorial. Thus searchers could either find relevant information to answer their questions in one of the BACK 2 SCHOOL tutorial lessons or find ready answers in Target 1.

None of the eight subjects, performing Target 1 searches, reached the target, although two had reached Nav. 1. Three reached the relevant lesson in the BACK 2 SCHOOL tutorial and answered correctly. Three found the necessary information to answer correctly in other documents, and two could not find the right information anywhere.

These results revealed the importance of the formulation of the questions, representing the information need, for the performance of the experiment's task as required by the research design. Simple questions of the Target 1 task type are the only questions that can be tackled by novices. However, in a specialised collection, such as **MicroWeb**'s, clues to basic information are bound to exist outside the desired paths, as was suggested by the three subjects who found correct answers to their questions elsewhere.

Thus it was found that only the subjects with at least some experience of the Internet could take part in the experiment. They were given more sophisticated sets of questions to answer. The one used for the Target 2 task (see **Appendix I**) had a specific focus (on Yahoo) and proved to have enough detail that could not be gleaned from other documents not specially devoted to the subject.

The Target 3 task set of questions (see **Appendix I**) was a variation on the Target 2 one (with a more general focus - on search engines versus subject indexes). Three searches performed on it showed that it had the same effect on the outcomes as the Target 1 question did. One of the 3 subjects reached Target 3, but did not see that the needed information was there. Both this and another subject reached a document that came after Nav. 2 and Nav. 2a and contained two thirds of the information needed. Nav. 2 was reached 3 times between these two subjects, and Nav. 2a - twice by one and 4 times by the other. The third subject used information from a document before Nav. 2 and Nav. 2a to answer the questions. Therefore, only the Target 2 task could be used for testing the effect of individual document characteristics on browsing outcomes, i.e. the first aspect of the **MicroWeb** research problem.

3.5.2.2 The Target 2 Task Experiment

The nature of the topic of the Target 2 task questions set made it possible to connect a large number of the short Web4Lib messages and the WWW FAQ pages, as well as the long, complex documents into paths leading towards the target. Before reaching the target, all the paths went through Nav. 2 and Nav. 2a - the original organizational pages of the document selected as a target for this study. While some of the needed information could be found in other documents,

to answer all the subquestions the searchers had to get to Target 2.

Ten subjects performed one browsing session each for the Target 2 task. They had varying backgrounds (with three from the LIS and two from the computer science domains), as well as varying degrees of Internet knowledge and experience.

Measures were decided upon to describe the monitored variables and the data gathered in the experiment were arranged in three tables (see **Appendix J**).

Table 1 displays a detailed breakdown of the data describing the subjects' navigational moves. The category of chosen links is made up of a count of the structural links connecting entry points in the contents page to the **MicroWeb** documents, the number of sole links in documents, first links in documents and other links in documents. Backtracking consists of going back to the contents page for a fresh start and internal path backtracking for document revision or branching purposes.

Table 2 represents a detailed breakdown of the subjects' browsing behaviour data, divided into two categories: inter-document browsing behaviour (described through measures reflecting the subjects' tactics in reading **MicroWeb**'s contents page) and across-document browsing behaviour. Interdocument browsing behaviour is measured through the number of false starts - paths aborted at the first document, useful path starts - paths consisting of two or more documents, reading direction changes - representing the frequency of scrolling up and down the **MicroWeb**'s contents page and entry points repetitions. Across-document browsing behaviour is represented by the number of short and long documents

visited by a searcher and the average length of the useful paths they traversed.

Table 3 lists the durations of the browsing sessions and the subjects' success rates measured by the number of the navigational pages (Nav.2 and Nav. 2a) and target hits.

Such abundant data make it extremely difficult for the researcher to recognize patterns of correlations if the data are not represented in a suitable form. One such form is graphical representation.

3.5.2.3 Analysis of the Results and Observations from the Target 2 Task

All the graphs presented in this section reflect the data summarised in Tables 1, 2 and 3. Although the character of the data is obviously discrete, a graphical style of points connected with lines was chosen to avoid ambiguity. Most of the data reflect fractions and have a natural interval of variation between 0 and 1. To make comparisons easier, whenever the data was not in the above interval, a normalization was performed. The graphical programme used was gnuplot, a widely used free software.

The statistical methods (for definitions see **Appendix K**) used for analysing the experimental data are:

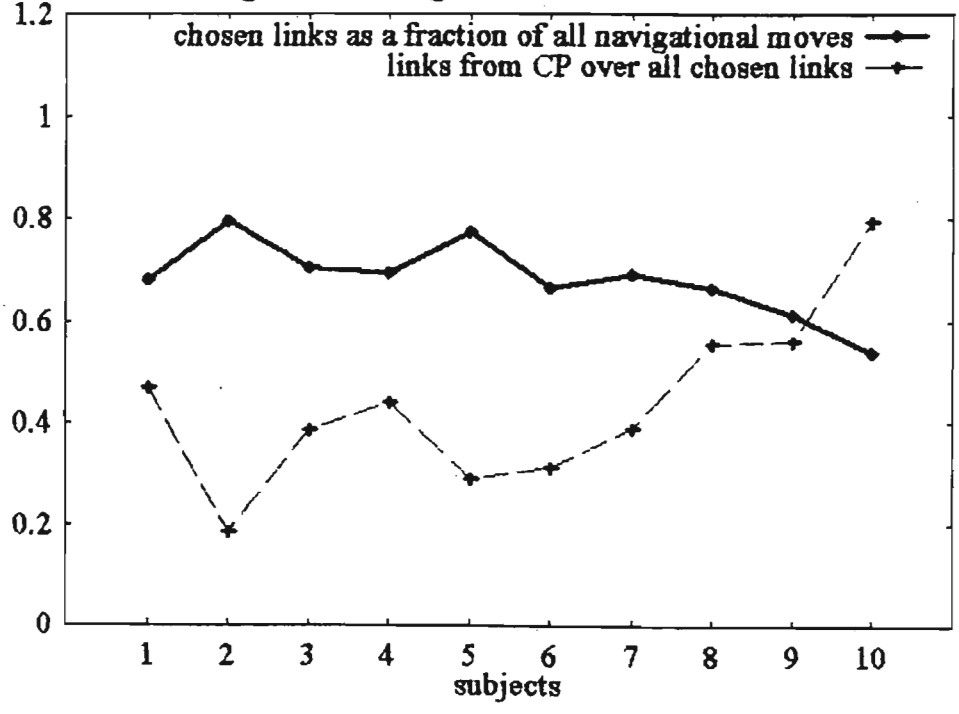
- qualitative correlation analysis
- computation of the first moments (mean and standard deviation) for some of the measured quantities

- computation of covariances and correlation coefficients

For the purpose of the statistical analysis a simple programme was written in Fortran 90. The significant results of the statistical analysis will be presented along with the relevant graphs.

Fig. 3-2 shows the use of chosen links as a fraction of all navigational moves and of links from the **MicroWeb** contents page (the structural links) as a fraction of all chosen links.

Fig. 3-2 Navigational Moves Correlations



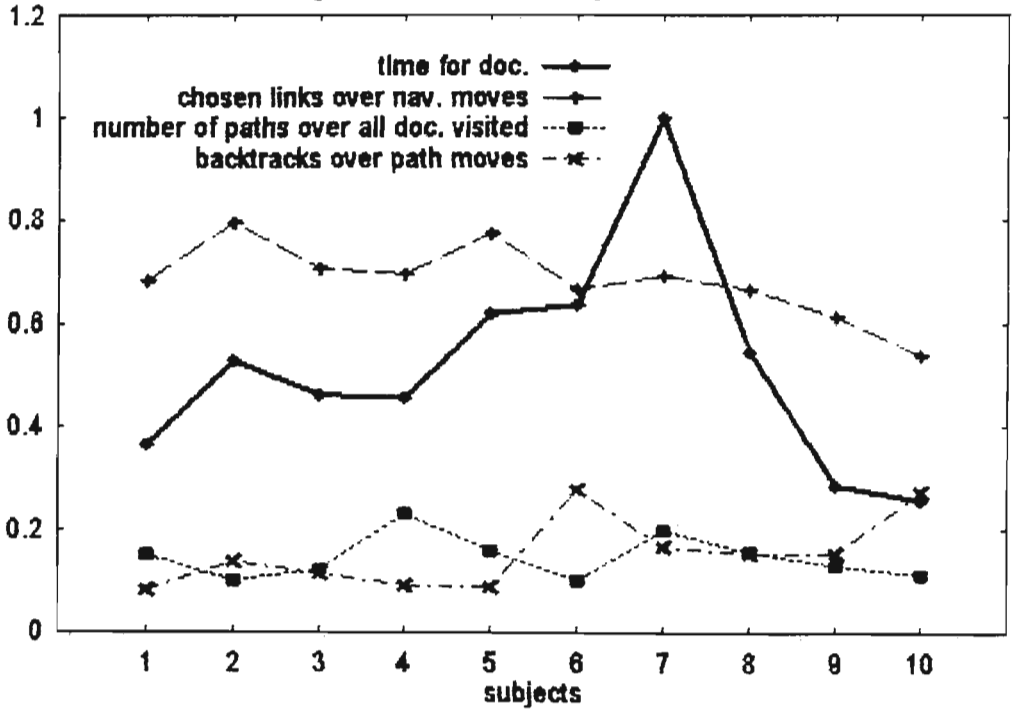
Choosing links (and consequently backtracking), seems to be a very nearly user independent quantity of mean value of about 0.7 (0.3 for backtracking) irrespective of the subjects' otherwise great differences in browsing styles. The statistical fluctuations in the conducted experiments have a variance of 0.005 which corresponds to a standard deviation of 0.07. The preference for hyperlinks as a method of traversal, established by Catledge and Pitkow (1995), is confirmed.

It must be noted that the decreased share of backtracking (in Catledge and Pitkow backtracking accounts for 41% as against 52% for hyperlinks and 7% for all other navigational moves) is also due to corrections made by this researcher. Backtracking moves which were used by the subjects to go back to the contents page or revisit a document without stopping to read any intervening documents were reduced to 1.

The comparative analysis of the two curves on **Fig. 3-2**, representing the use of structural links as a fraction of all chosen links and the use of chosen links as a fraction of all navigational moves, confirms the visible mirror-like correlation between them with a correlation coefficient of -0.92.

Fig. 3-3 and **Fig. 3-4** compare the efficiency of the searches according to various criteria: average time (in minutes) for reading a document, chosen links as a fraction of all navigational moves, number of useful paths as a fraction of all documents visited, backtracking as a fraction of useful path moves.

Fig. 3-3 Search Efficiency Correlations: I



The average-time-for-reading-a-document curve reflects the variety of reading styles of the subjects. As can be seen, the normalized time for reading a document fluctuates significantly about the average value of 0.5. On the other hand, the values for the number of useful paths as a fraction of all documents visited and for backtracking as a fraction of useful path moves are remarkably stable. But this is not all. These two quantities have practically the same mean value of 0.15 and their variances are of the same order of magnitude: 0.002 and 0.004 respectively. There is a certain negative correlation between the two curves with -0.51 correlation coefficient.

Further efficiency measures correlations are shown on **Fig. 3-4** : number of useful

paths as a fraction of all documents visited, internal backtracking as a fraction of all backtracking moves and average length of the useful paths traversed.

The quantity internal backtracking as a fraction of all backtracking moves is strongly correlated to the average useful path length: correlation coefficient of 0.87. The correlation between the number of useful paths as a fraction of all documents visited on the one hand and the internal backtracking as a fraction of all backtracking moves and average length of the useful paths traversed on the other is significantly smaller in magnitude and negative, namely -0.38 and -0.64 respectively.

Fig. 3-4 Search Efficiency Correlations: II

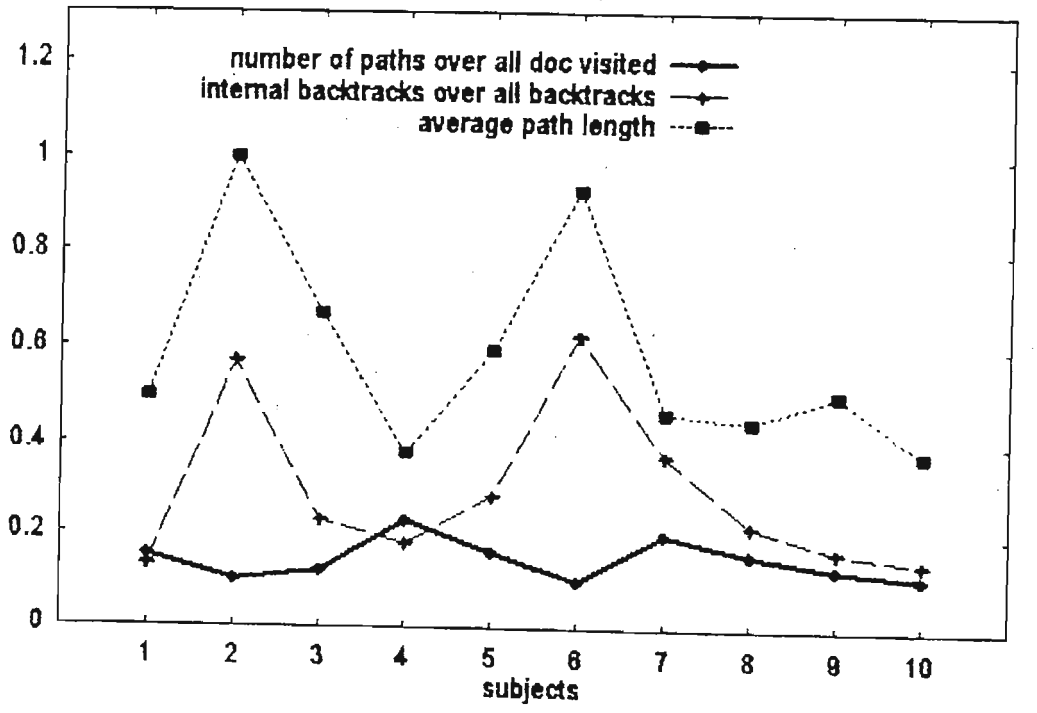


Fig. 3-5 represents the subjects' within-document reading strategies measured by the number of reading direction changes as a fraction of all contents page visits and first links selected as a fraction of other links. The analysis of the data plotted on this figure does not show any significant correlation between the two curves. The correlation coefficient is 0.23.

Fig. 3-5 Reading Strategies Correlations

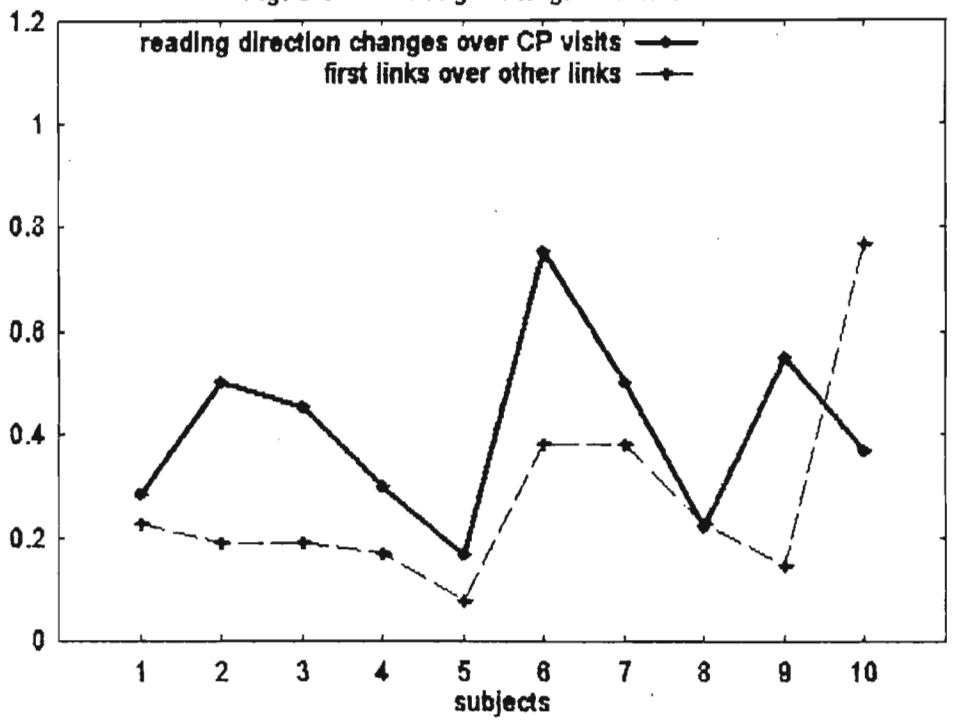
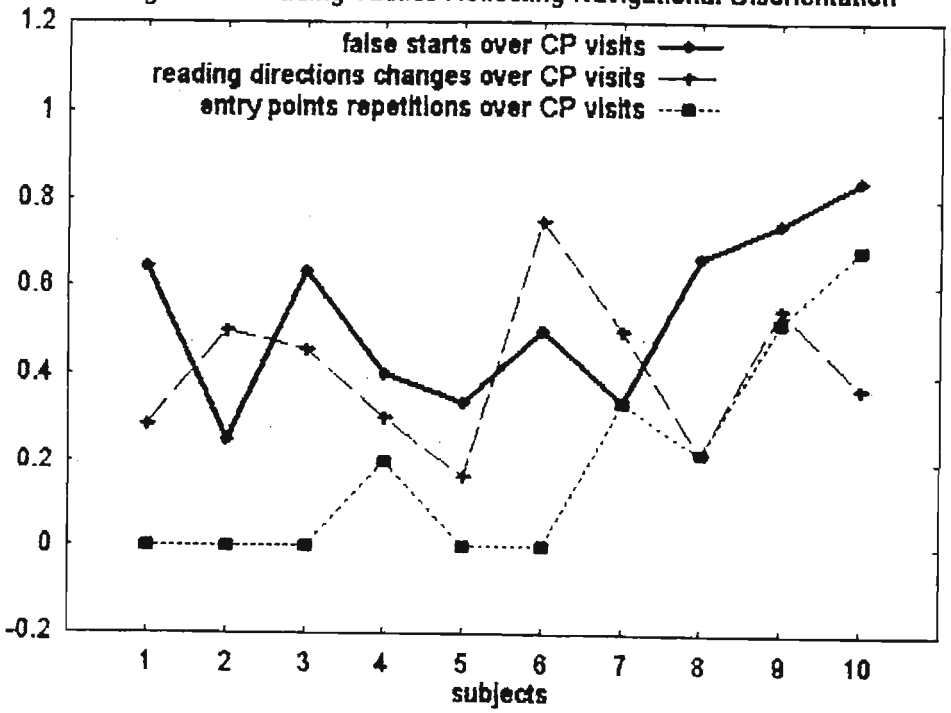


Fig. 3-6 compares the subjects' within-document reading strategies, measured by the number of reading direction changes as a fraction of all contents page visits, to disorientation, measured by the number of false starts as a fraction of all contents page visits and the number of entry points repetitions as a fraction of all contents page visits.

The correlation analysis of the data shows that there is no correlation between the number of false starts as a fraction of all contents page visits and the number of reading direction changes as a fraction of all contents page visits - the correlation coefficient is -0.03. No correlation was found between the number of reading direction changes as a fraction of all contents page visits and the number of entry points repetitions as a fraction of all contents page visits either - the correlation coefficient was 0.02. In contrast, the number of false starts as a fraction of all contents page visits and the number of entry points repetitions as a fraction of all contents page visits may seem to a certain degree to be correlated with their correlation coefficient of 0.57.

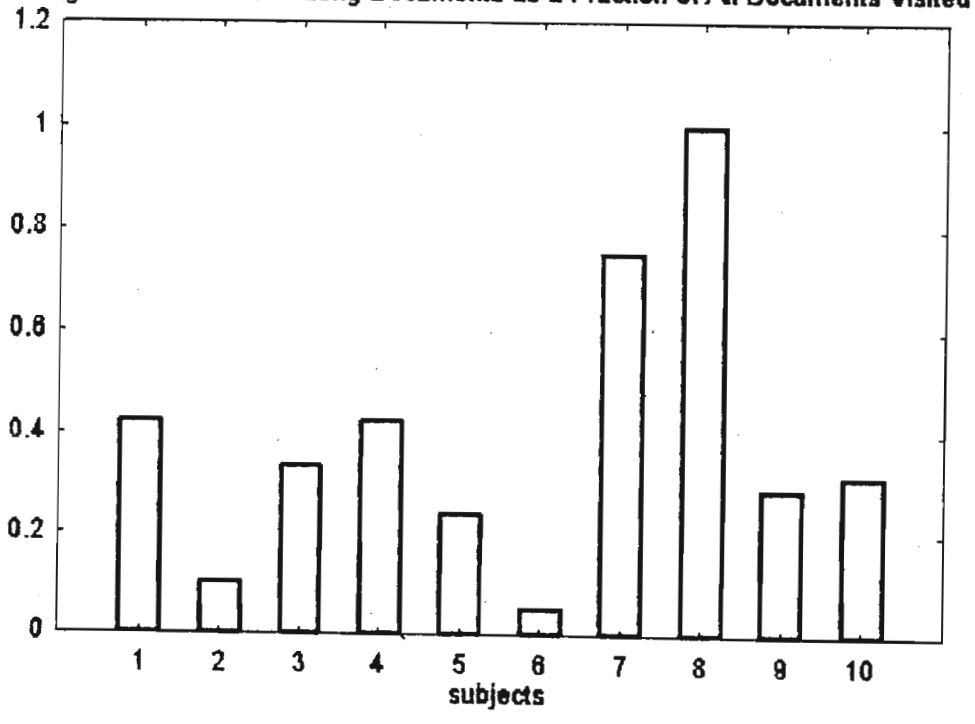
Fig. 3-6 Reading Tactics Reflecting Navigational Disorientation



Three of the ten subjects (numbers 7, 8 and 9) reached Target 2. All subjects reached Nav. 2 - six of them more than once. Nav. 2 was visited both via paths of short, simple documents and long, complex ones, 20 times altogether. Eight out of the ten subjects reached Nav. 2a - five of them once, two - twice, and one - three times.

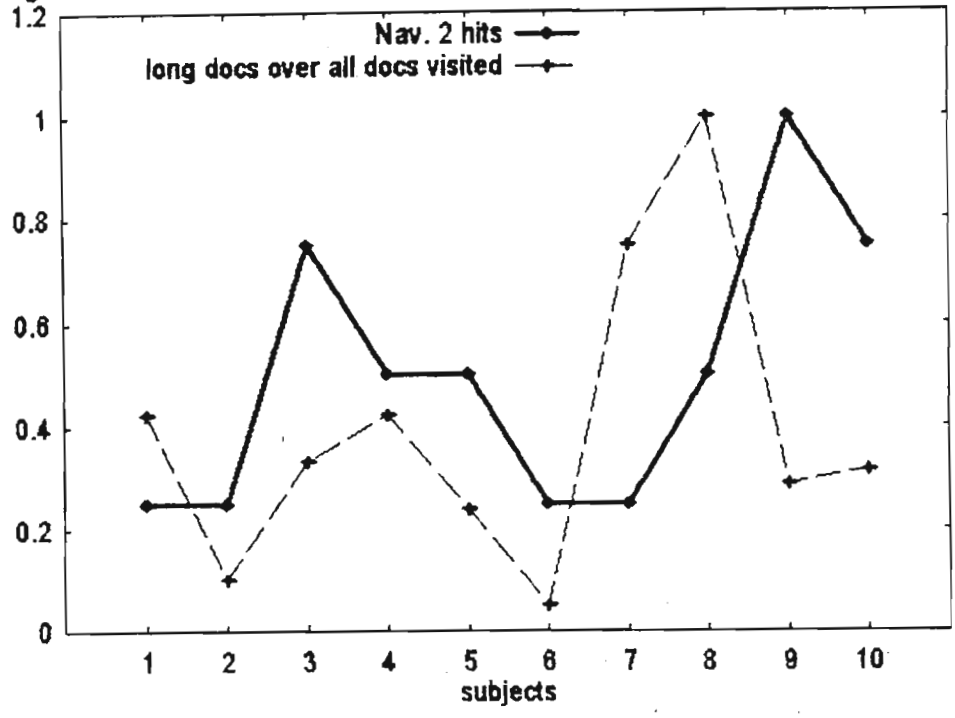
If we use the Target 2 hits as the only measure of success, we will see that of the three subjects, who reached it, number 7 went predominantly via paths of long documents, number 8 via long documents only and number 9 predominantly via short documents (**Fig. 3-7**).

Fig. 3-7 Number of Long Documents as a Fraction of All Documents Visited



If the number of Nav.2 hits is used as a measure of browsing success, then the statistical analysis of the data on the Nav. 2 hits and the number of long documents as a fraction of all documents visited, plotted on **Fig. 3-8**, clearly shows the lack of any noticeable correlation. In fact the correlation coefficient is just -0.03.

Fig. 3-8 Correlations between Retrieval Effectiveness and Length of Documents: I



3.5.2.4 Discussion of the Experimental Results

It must be emphasized again that this study is an exploratory one and its findings must be treated with caution. As Saracevic *et al.* (1988) point out, the results of an exploratory study, called for by "the meagre state of knowledge and observations on the variables involved, ... are really reflective of the

circumstances of the experiments alone". Therefore the findings, reported in this thesis, should be regarded as questions for further research, rather than generalizations.

The discussion coming below will examine the findings from all the task searches matching them to the three aspects of the experimental research problem:

- the effect of the length and modularity of hypertext documents on browsing outcomes

- the effect of the configuration of internal and external hypertext links on browsing outcomes

- the interaction between the above two aspects and the browsing strategies, employed by the information seekers

All the variables monitored - authors, questions, browsers, browsing sessions, Web documents, target documents reached, proved to have an interrelated effect on the WWW-mediated communication process. Of these, the questions, which in browsing materialize both the user's information need and the information seeking task, brought forth the strong interaction between WWW individual document characteristics and the authors' document structuring and organization patterns in building collections, (the focus of the first two aspects of the experimental research problem of this study), as a major factor influencing the success of browsing for information on the Web.

3.5.2.4.1 The Effects of Documents, Authors, and Questions on the Outcomes of the MicroWeb Browsing Sessions

When a WWW intermediary author is building a Web collection, the decisions he has to make about creating, structuring and organizing his original documents are dictated by his own purpose and perspective. He will have a clear-cut structure of the components and levels of organization required. Aggregating modular units into semantic networks through associative links and providing entry points to the latter from the top-level document of the collection (the home contents page) seems to be mostly dependent on the intermediary author's judgement as well.

The **MicroWeb** experiments, however, have given rise to a number of questions concerning the treatment of other Web authors' aggregations, such as composite nodes and sites. If an intermediary author is really intent on building a collection, with a definite purpose, meant to serve a well-defined audience, and not just making links to "interesting" or "cool" sites, the importance of how complex, multi-level documents from other sources are incorporated becomes paramount.

An overview document, such as the **MicroWeb** home contents page, cannot provide the level of specificity required for a focused browsing search as the Target 2 task one. Since most of the subjects were not supposed to know much about Yahoo (or the WWW terminology at any rate), the set of Target 2 task questions deliberately emphasised the relationship between Yahoo and the WWW search engines for which there are a number of possible entry points. The collection developed by the Library Web Manager's Reference Centre covers the WWW search engines extensively in many materials so it seemed worthwhile to

incorporate as much of this collection into the model as possible, regarding it as a ready-made complex node, complete with its homepage (Nav. 2). The Nav. 2 complex node was made available at various positions and layers in the **MicroWeb** environment: after long paths of all three types of documents - short and simple, short and modular, long and complex, after short paths of the three types, and on one occasion - directly linked to the **MicroWeb** contents page.

The large number of visits to Nav. 2 via all types and lengths of paths showed that neither of the individual document characteristics, such as the length and complexity of documents or the larger number of links between the short, modular ones, impeded the search progress through the networks of referential links, irrespective of the amount of branching they offered. It was the navigational pages that seemed to act as search barriers throughout the experiment. Nav. 2 stopped two of the subjects from progressing any further. Nav. 2a stopped or deviated the succeeding moves towards the target, so that only 3 out of the 13 Nav. 2a hits led to success.

These effects were so consistent that very early in the experiment the researcher had to start questioning whether it was any characteristics of the primary documents at all that were creating cognitive overload and causing disorientation in the Web space. Navigational pages are supposed to aid with navigation. What was it that made them act as an obstacle to browsing? Was it the density of the links that contributed to the sense of cognitive overload, especially after a long path and towards the end of a long browsing session? That did not seem to be the case, since Nav. 2a was reached (to no avail) with his second move by subject 3, as was Nav. 2 (again to no avail) with her fifth move by subject 5. Subjects 9 and 10 both chose the **MicroWeb** contents page entry point leading directly to

Nav. 2 very early in their searches, yet subject 10 did not even select the link to Nav. 2a, while subject 9 did not progress beyond Nav. 2a.

Could the reason for the obstructive effect of the navigational pages lie with the relevance to the search of the entry points provided by them then? The model's contents page is also a navigational page - how does it compare in its effect on the browsing outcomes to Nav. 2 and Nav. 2a?

A valuable insight into these problems was gained from the observations and analysis of the Target 1 task sessions. The subjects consistently chose contents page entry points that were strikingly different from what the author expected them to choose. It seemed logical to the author that the Internet tutorials and glossaries she had collected should be grouped together under a heading of the "understanding cyberspace" type, with subheadings in the FAQ style, e.g. "What is the Internet?, What is the WWW?". To the author's great amazement, six out of the eight subjects looking for answers to the questions based on the Internet tutorials chose entry points from other sections.

When this happened for the first time, the author tried to emphasize her meaning by capitalizing and underlining the headings. This made no difference. The author then thought that probably the unfamiliar word "cyberspace" and the too general wording of the subheadings were the reason for the subjects' avoidance of the reference section of the model. But what could account for the subjects' reluctance to check the glossaries for definitions of unfamiliar words?

The prevalence of words, such as "navigating", "catalogue", "searching", in the chosen subheadings revealed the subjects' need to use familiar methods of

searching for information. The subjects were not looking for reference materials to browse, but for tools of the library online catalogue type to do the searching for them, irrespective of the fact that they had been made to understand that they could do nothing but follow links and go back a step at a time in **MicroWeb**.

Conversely, in the Target 2 task sessions, six out of the ten subjects made their first choices of entry points from the motivational and reference sections, despite the availability of far more specific entry points, explicitly mentioning searching, under section 3. Only two of those checked the glossaries for references to Yahoo, but then proceeded with other entry points from sections 1 and 2. The four subjects who did use entry points from section 3, confirmed again the users' preference for words related to "catalogue" and "navigating". Only one chose an anchor using the word "searching" as his first entry point.

A discrepancy between the context intended for the contents page by the author and the context the seekers were working from is not surprising as the information seeker brings to an information environment his own context determined by his domain knowledge, information seeking experience, motivation. It is worth investigating to what an extent this clash is responsible for cognitive overload and disorientation in hypertext environments.

The experience of the discrepancy between the author's intention and the seekers' interpretation of the meanings encoded in **MicroWeb**'s contents page suggested the idea that the major problem with navigational pages is insufficient context. Disorientation and cognitive overload, due to insufficient context, differ depending on the type of navigational page - examples from this study are the home contents page - the model's top-level overview navigational document, a

secondary home contents page (The Library Web Manager's Reference Centre homepage, referred to as Nav. 2), a top complex node page, representing a collection of documents on one topic (the Search Engines' Reference List, referred to as Nav. 2a). The lesser the disparity in content in these pages, the more contextual specificity and detail are needed.

For example, the subheadings under the third heading "THE ROLE OF LIBRARIANS IN THE INFORMATION TECHNOLOGY AGE" on **MicroWeb**'s contents page do not convey the connotation of "aboutness" that the author assumed they did. An original home or FAQ page, such as Nav. 2, which contains long lists of content-disparate links selected and organised with a different intent, becomes a barrier to communication when incorporated in a new author's perspective. Even more so do numerous annotated links to materials on one topic, as are gathered in the Search Engines Reference List - Nav. 2a. These make the document too bulky to read, yet are not specific enough to guide the reader. Comparing multiple items of disparate, and even worse - similar non-specific, information, and trying to match them to a specific query is a multi-process cognitive activity with an extremely high overhead.

The Library Web Manager's Reference Centre collection, based on FAQs, is an excellent resource for the digital librarian who is familiar with its content. It might even be of some use to a proficient Internet user too. But it will not do much for a new end user, nor will the Search Engines Reference List when positioned between two end documents.

The retrieval of this type of pages may be compared to the discovery of a relevant journal title. Support is needed then for micro-retrieval - the equivalent of

relevant article titles. A complementary index to be browsed or searched might be one of the means to solve the problem if navigational pages like these are not to act as barriers on the way to specific matches to a focused information search.

On the other hand, Nav. 1 (the BACK 2 SCHOOL Internet tutorial contents page) did not show signs of becoming a barrier to communication because it was only available via a "Back to Index" page link at the end of the lessons comprising the BACK 2 SCHOOL composite node. Linked in this way top levels of composite nodes are very useful because they provide a global coherence to the components of a hypertext unit, as well as further similar or related entry points supported by sufficient context from the preceding document.

To sum up, the analysis of the results from this study seem to point to the following recommendation to intermediary authors of WWW documents: when linking to other authors composite documents and sites they should be careful to avoid crossing the lines of micro and macro retrieval which often results from indiscriminate linking to navigational pages. In practice, this means that one must not connect one end-document to another end-document through a navigational page. Navigational pages are the front-end of macro retrieval and, if incorporated in a new collection as part of a borrowed composite node, seem to suffer from lack of context. This can be remedied by providing them only via retrospective links in end-documents, as was the case with Nav. 1. Home contents pages, even for relatively modest collections, need complementary information finding tools to support focused browsing

3.5.2.4.2 The Effects of Browsers on the Outcomes of the MicroWeb Browsing Sessions

In the preceding section it was discussed how browsers affected the outcomes of their searches in a reactive role, interpreting and responding to the communication sources' meanings. This section will be devoted to their proactive role in the WWW-mediated communication process.

Except for the number of false starts as a fraction of all contents page visits and the number of entry points repetitions as a fraction of all contents page visits used as indicators of reading tactics reflecting navigational disorientation, none of the data collected on inter-document and across-document browsing behaviour showed any significant correlations. This seems to confirm Saracevic and Kantor's (1988) conclusion that the way a search is done "seems to affect the results less than many other factors associated with users, questions, searchers, and search term selections".

However, the statistical analysis of the subjects' navigational moves data suggests that the users' browsing strategies may have an effect on the efficiency of their searches. A very strong correlation was revealed between the use of links from the contents page as a fraction of all chosen links and the use of chosen links as a fraction of all navigational moves (correlation coefficient of -0.92). This shows that the less users resort to backtracking moves in general, the greater will be the number of internal links they follow, which creates conditions for more and/or longer useful paths. Since internal backtracking is strongly correlated and contributes positively to the average useful path length (correlation coefficient 0.87), it is evident that an excessive use of "hub and spoke" backtracking

(Catledge and Pitkow 1995) to the home contents page will signal fewer and/or shorter useful paths.

Saracevic and Kantor (1988) have found that more command cycles tend to produce better searches since they allow for feed-back. In the highly interactive browsing context, more and/or longer useful paths should have an even more pronounced effect. It would be interesting to investigate in the future whether dividing the subjects into groups according to the different browsing styles as described below and conducting a comparative study on their searches on the same questions will not reveal correlations between browsers' strategies and the outcomes of their searches.

From the very start of the experiment, it was noted that different subjects had different ways of studying the **MicroWeb** contents page. Some (subjects 9 and 10) would scan quickly down the list, click on what seemed to be promising entry points and quickly come back to scan for more. This group seems to be representative of the style, identified by Fidel (1984), of operationalist searchers who believe that the more they manipulate the system, the more they increase their chances for success and for getting a quick overview of the available material. These two subjects happen to be computer science specialists.

Another type of operationalist is the methodical searcher, represented by subject 1. The methodical searcher would explore every entry point on the contents page in the order that it comes from top to bottom. She would read the opening parts of documents carefully, but would tend to pursue the early links in documents without having seen its end. Together with the "overview" operationalists she would make more moves per minute and would also have more "false starts" -

entry points discarded after the first move. However, she would not show their high entry points repetition numbers or other signs of disorientation - e.g. short meaningful paths including a substantial number of revisited documents. While the "overview" operationalist approach is not necessarily ineffective (subject 9 was one of the three who reached the target document), its efficiency could be improved.

The rest of the subjects seemed to represent Fidel's (1984) conceptualist approach searchers. These would make fewer navigational moves because they would explore thoroughly the contents page, they would scroll pages up and down several times before choosing a link to activate. They would tend to select links further down documents since they needed to analyse the whole content before they made their choices. They would be reluctant to pursue links that they had tried before, even when those links were the only obvious choice. They would read around the topic for general understanding in the beginning of the session, but as it progressed they would focus more and more on specificity (in this case scanning for Yahoo only).

All the subjects needed to keep notes for one reason or another - to analyse the task questions and choose the focus of the search, to record interim and final results from the search and to remind themselves where they had already been. The latter need was felt particularly strongly by the "overview" operationalists when browsing the model's contents page - not only were their "false starts" numerous, but practically all of them were repeated several times. Trying to keep notes on paper with the computer keyboard and mouse taking up most of the desktop space is quite a problem, so the ability to make margin notes on the computer screen would be very useful for WWW browsing.

Outside the contents page, the subjects did not experience considerable difficulty in recognizing documents already visited by them. Backtracking was mostly used for returning to the model's contents page when the seeker had finished with a chosen path. Backtracking to previously visited documents was limited. From the way it was done - users did not take any time to read, but went straight to another path branching off an unexplored link, it was clear that they did not go back to find old information.

The analysis of the data and the observations related to browsers and browsing strategies point once again to the major contents page as the greatest source of cognitive overload and disorientation due to insufficient context.

3.6 Conclusion

The ease with which useful WWW material can be collected or made easily accessible through linking by millions of end users or information intermediaries can lead to a serious tangling of the Web. Trying to impose order on disparate ready-made resources can be quite overwhelming. Intermediary authors need every support to become effective communicators on the WWW. The LIS approach, adopted in this study, to the task of improving the usability of the WWW as an information seeking tool has the advantage that it reflects the end users' behaviours and needs.

A WWW computer simulation model, such as **MicroWeb**, provides a unique opportunity for a face-to-face feedback on the users' comprehension of the authors' meanings, thus helping to identify and eliminate communication noise, bottlenecks, and barriers. The results from the experiment have shown that

WWW individual document characteristics, such as length and complexity or the density of links in the short, modular documents, do not in themselves contribute in any way to cognitive overload and disorientation and have no effect on the browsing search outcomes. However, the simulations have highlighted the importance of WWW structural links and documents, as well as of their position for the success of browsing searches.

If linked directly to primary end-documents, not belonging to their set, secondary top structural documents obscure even further the boundaries of elementary and composite primary documents and turn into communication obstacles because of insufficient context. A WWW intermediary author should only provide secondary top level structural documents via retrospective links to increase the global coherence of the documents they organize.

Collection overview top pages, such as the **MicroWeb** home contents page, are a major source of cognitive overload and disorientation, because of insufficient context as well. They need interface aids and complementary finding tools to support focused browsing, especially for the information seekers using the "overview" operationalist approach.

Despite the fact that practically no correlations were found between the subjects' browsing behaviour patterns and the search outcomes, indirect evidence to the effect of the former on the efficiency of the searches indicates the need for a further comparative study focused on the performance of subjects divided into groups according to their browsing styles.

Conclusion

The starting point of the study, reported here, is the identification of the areas of the greatest impact, exerted by the new information environment of the WWW, on the role of libraries and librarians in the information creation and dissemination chain. One such area is the improvement of the WWW-mediated communication process in its systemic entirety of input, process, output, and feedback. This area has formed the basis of the aim of this thesis, namely: to reveal the importance of local and global structural coherence of documents, especially of those incorporated by intermediary authors into new collections, as a factor facilitating the effectiveness of information seeking on the WWW and enhancing access to its resources.

It is the rapidly growing part of computer-mediated communication not only in the workplace, but also in the everyday life of mankind, that has given prominence to the problem of improving the usability of computer-based media in general. Internet-based information systems, and above all the WWW, have added to the problem by their potential for universal readership and authorship. Furthermore, the open and dynamic nature of the WWW and the flexibility of structuring and organizing approaches in hypertext, have emphasized the need for support both of automated searching and human browsing on the one hand, and of authoring on the other. This support can no longer focus, as has mostly been the case so far, on the technology interface alone, but has to spread to all aspects and stages of the author-reader interactions in the WWW-mediated communication process.

The research problem, addressed by the experimental part of this study, focuses on investigating the extent to which the length and modularity of hypertext documents

on the one hand, and the configuration of internal and external hypertext links on the other, influence the outcome of browsing for information on the WWW. There are three aspects, inherent in it, which highlight the interactions between WWW individual document characteristics, authors' document structuring and organization patterns in building collections, and browsing behaviours exhibited by information seekers.

The study of information seeking on the WWW under real conditions, based on user interface level events captured in client-side log files, accumulates data from a very large number of instances. However, this is at the expense of oversimplifying the events and severely restricting the number and the complexity of the monitored parameters, which, of necessity, obscures a number of the variables affecting search behaviour in electronic systems: e.g. the context brought to the system by the information seeker and the latter's motives behind his moves. Moreover, in contrast to single-author systems, authoring decisions in the collaborative framework of the WWW have a far more pronounced effect, especially in browsing which is heavily dependent on the information environment. Authoring decisions, then, have to be taken into account when assessing browsing search outcomes on the WWW.

In its turn, assessing search outcomes in relation to the above-mentioned factors requires some channelling of the user's moves in accordance with a carefully set task, which is impossible to achieve in the unpredictable and dynamic infinity of the real cyberspace. Consequently, in order to formulate a manageable concrete research task for an experimental study of the author-information seeker interactions, the real cyberspace was replaced by a simple model, named **MicroWeb**. It captures the important features of the WWW for the purposes of the problem under study and serves as a testbed for the information seeking experiment.

While allowing the controlling or elimination of extraneous factors, the method of computer simulations is powerful enough to provide insights into the behaviour of complex systems. The **MicroWeb** simulations are designed to monitor the complex interactions of the following variables: authors, questions, browsers, browsing sessions, Web documents, target documents reached. Authoring decisions are reflected in the selection, treatment and arrangement of the various types of documents included in the model: original, customized and borrowed. Target documents, to which multiple paths via short-elementary, short-modular and long-complex documents are built, are used to measure the browsing search success rates in relation to individual document characteristics, authoring decisions, browsing sessions and browsers' behaviour. The choice of targets is crucial to the success of the experiment, as it strongly affects the questions, which must ensure that the subjects persist in their searches until they reach the targets. A set of questions on a specific topic, related to a key issue, that is bound to be referred to in most of the documents, was found to be the best suited for this purpose.

Since no programme for automatic log-file creation, storing all the relevant information for a long browsing session on the model, is available, the number of data in manual records has to be limited to the major identifiers that would enable the researcher to reproduce the session for a detailed study of the navigational moves. Thus enough time is left to record all the information, concerning human behaviour, that cannot be captured by a computer.

Hypertext browsing within the model does not differ in principle from an authentic WWW session. **MicroWeb** is much larger than the size of an average WWW site and incorporates various authors' documents, complete with their original document structuring and organizing patterns. It is, therefore, reasonable to assume that the

model exhibits the characteristic features of an interlinked system of sites. Size here should be understood in the sense not only of volume (e.g. in megabytes), but also of number of documents.

The results from the experiment, when compared to existing related research, seem to indicate that the model is sufficiently realistic. An example in point is the remarkable stability for all subjects tested of the complementary quantities of choosing hyperlinks and backtracking as methods of traversal on the WWW and the preference for hyperlinks over backtracking, established in a statistically-convincing, real Web log-file study by Catledge and Pitkow (1995).

Similarly, the quantities representing the number of useful paths and backtracking within useful paths were found to be very stable and their mean values practically identical. This could be an indication of some universal relationship which is in need of further investigation.

Another significant finding, based on the analysis of the subjects' navigational moves, shows that proactive user behaviour, manifested in a higher ratio between hyperlinks and backtracking in favour of the former, corresponds to a lower use of structural links. It would be worth investigating in a further study whether excessive use of the "hub and spoke" pattern, documented by Catledge and Pitkow (1995), is not suggestive of the effects of cognitive overload and disorientation, typically experienced in navigating hypertext environments. Excessive use of the "hub and spoke" pattern was found to signal fewer and/or shorter useful paths. In the light of Saracevic and Kantor's (1988) finding that more command cycles tend to produce better searches, the number and/or length of useful paths might prove to be of importance to the browsing search outcomes.

The Target 2 task simulations have produced some definite answers to the first two aspects of the experimental research problem. The results have shown that WWW individual document characteristics, such as length and complexity or the density of links in short, modular documents, do not in themselves contribute in any way to cognitive overload and disorientation and have no effect on the browsing search outcomes. However, these characteristics of individual documents presuppose the provision of certain structural links and documents, whose position has a major effect on the success of browsing searches.

If linked directly to external primary end-documents, top structural documents of borrowed composite nodes or sites turn into communication obstacles because of insufficient context. Such secondary top level structural documents are very useful, though, when provided via retrospective links. In that position they gain in context from the preceding document and in their turn increase the global coherence of the components they organize.

Insufficient context was also found to create problems for information seekers in a primary, front-end structural document, such as a collection's overview top page of the **MicroWeb** home contents page type. The values for the number of false starts as a fraction of all contents page visits and the number of entry points repetitions as a fraction of all contents page visits have indicated greater navigational disorientation on the part of certain subjects. Judging from this, as well as from other behavioural patterns suggestive of common features, the subjects participating in the **MicroWeb** experiment, were divided into the following main groups, according to their browsing styles:

- the operationalist searchers

- the conceptualist searchers

The operationalist and conceptualist approach styles were identified by Fidel (1984). The **MicroWeb** experiments pointed to the existence of two subgroups within the operationalist category:

- the methodical operationalists

- the "overview" operationalists

The lack of control on the subjects' browsing approaches might be a reason why none of the remaining data collected on inter-document and across-document browsing behaviour showed any significant correlations. It would be worth investigating whether this situation will change in a further comparative study focused on the performance of subjects divided into groups according to their browsing styles.

In order to alleviate the major effects of cognitive overload and disorientation, experienced in the overview top pages of WWW document collections, interface aids and complementary finding tools are needed to support focused browsing, especially for the information seekers using the "overview" operationalist approach.

The conclusions drawn from the analysis of the information seeking simulations on the model can be expected to apply for the whole World-Wide Web only if the scaling to the full size of the actual WWW does not introduce significant corrections. It is believed that scaling corrections due to the finite size of

MicroWeb will not negate the validity of the main conclusions drawn from the simulations, although the problem exists and is one of the factors determining the limitations of the research. Another limitation, caused by the lack of massive background research and logistic constraints is the modest number of simulation instances. However, these were performed on a carefully selected sample of subjects and attempted to record as many of the relevant parameters of the simulation as possible. Still, possible directions for further studies could be to investigate scaling effects from the finite size of the model and to improve the statistics.

While this was an exploratory study, suffering from certain unavoidable limitations, the research, reported in this dissertation, has attempted to make several contributions. It is believed to be the first to have examined the author-reader interactions on the WWW taking into account the whole complexity of the communication process. It has brought about the creation of a research tool for experimental studies of the LIS-oriented WWW-mediated communication process by means of simulations. The model is available on the Internet and can be used to replicate the reported experiment as well as for further studies along the directions identified above. It can also be upgraded to increase its size and complexity. The principles embedded in the structure of the model can be implemented in similar hypertext-based systems in other fields of science and can help with the formulation of tasks for the investigation of other research problems.

References

ALA Council. 1996. Access to electronic information, services, and networks: an interpretation of the Library Bill of Rights.
gopher://ala1.ala.org:70/00/alagophx/alagophxfreedom/electacc.fin

Balasubramanian,V. 1995. State of the art review on hypermedia issues and applications.
http://www.csi.uottawa.ca/~dduchier/misc/hypertext_review/index.html

Begoray, J.A. 1990. An introduction to hypermedia issues, systems and application areas. **International Journal of Man-Machine Studies**, 33: 121-147.

Benson, A.C. c1995. **The complete Internet companion for librarians**. New York: Neal-Schuman Publishers, 405 p. ISBN 1-55570-178-7.

Berghel, H. 1997. Cyberspace 2000: dealing with information overload. **Communications of the ACM**, 40(2): 19-2. ISSN 0001-0782.

Berners-Lee,T. & R.Cailliau.1990. World-Wide Web: proposal for a hypertext project.
http://www.w3.org/pub/WWW/Proposal.html

Berners-Lee, T. , R. Cailliau, A. Luotonen, H. F. Nielsen and A. Secret. 1994. The World-Wide Web, **Communications of the ACM**, 37(8): 76-82. ISSN

0001-0782.

Berners-Lee, T. 1993. Style guide for online hypertext. http://www.w3.org/pub/WWW/Provider/Style/Overview.html

Berners-Lee, T. 1994a. World-Wide Web seminar. http://www.w3.org/pub/WWW/Talks/General.html

Berners-Lee, T. 1994b. WWW concepts. http://www.w3.org/pub/WWW/Talks/General/Concepts.html

Berners-Lee, T. 1996. About the World-Wide Web Consortium (W3C). http://www.w3.org/pub/WWW/Consortium/

Boon, J. A. 1992. Information and development: towards an understanding of the relation. **South African Journal of Library and Information Science**, 60 (2): 65-74.

Borgman, C.L., D.O. Case and C.T. Meadow. 1989. The design and evaluation of a front-end user interface for energy researchers. **Journal of the American Society for Information Science**, 39(3):99-109. ISSN 0002-8231.

Borgman, C.L., D. Krieger, A.L. Gallagher and J. Bower. 1990. Children's use of an interactive science library: exploratory research. **School Library Media Quarterly**, winter 1990: 108-111.

Bowman, C. M., P. B. Danzig, U. Manber and M. F. Schwartz. 1994. Scalable Internet resource discovery: research problems and approaches. **Communications of the ACM**, 37(8): 98-107. ISSN 0001-0782.

Brown, I. 1996. From books to bits... **Ariadne**, Issue 6, 20 Nov 1996. http://www.ukoln.bath.ac.uk/ariadne/issue6/textbooks.html

Browning, J. 1996. What is the role of libraries in the information economy? **Wired**
http://www.hotwired.com/wired/1.1/features/libraries.html

Bush, V. 1945. As we may think. **Atlantic Monthly**, 176(1): 101-108.

Cailliau, R. 1995. A short history of the Web: text of a speech delivered at the launching of the European branch of the W3 Consortium,Paris, 2 November 1995.
http://www.inria.fr/Actualites/Cailliau-eng.html

Calder, N. 1996. The global research village: contrasting scenarios for the impact of information technology on science.
http://www.oecd.org/dsti/sti_new.html

Callery, A. 1996. Yahoo! Cataloging the web. Proceedings of the "Untangling the Web" conference, sponsored by the Librarians' Association of the University of California, Santa Barbara and Friends of the UCSB Library, University Center, University of California, Santa Barbara.April 26, 1996.
http://www.library.ucsb.edu/untangle/callery.htm

Carts-Powell, Y. 1995. Software products enrich Web options for scientists. **Computers in Physics, 9(4):364-368.** ISSN 0894-1866.

Catledge, L.D. and J.E. Pitkow. 1995. Characterizing browsing strategies in the World-Wide Web. **Computer Networks and ISDN Systems: proceedings of the Third International World-Wide Web Conference**, 10-14 April, Darmstadt, Germany, 27: 1065-1073.

Chamberlain, E. and M. Mitchell. 1996. BCK2SKOL: the electronic library classroom 101.
http://web.csd.sc.edu/bck2skol/fall/fall.html

Cockburn, A. and S. Jones. 1996. Which way now? Analysing and easing inadequacies in WWW navigation. **International Journal of Human-Computer Studies**, 45: 105-129.

Collins, B.R. 1996. Webwatch. **Library Journal**, 121(8): 25 - 26. ISSN 0363-0277.

Conklin, J. 1987. Hypertext: an introduction and survey. **(IEEE) Computer**, 20(9): 17-41.

Corn, D. 1996. Anatomy of a netscam: why your Internet search may not be as honest as you think. **The Washington Post**, July 7 1996; Page C05

Cove, J.F. and B.C. Walsh. 1988. Online text retrieval via browsing. **Information Processing and Management**, 24(1): 31-37. ISSN 0306-4573.

Crawford, D. 1995. Editorial pointers. **Communications of the ACM**, 38(7): 5. ISSN 0001-0782.

Denning, P.J. 1996. The university's next challenges. **Communications of the ACM**, 39(5): 27-31. ISSN 0001-0782

Denning, P.J. and B. Rous. 1995. The ACM electronic publishing plan. **Communications of the ACM**, 38(4): 97-103. ISSN 0001-0782

Din, J. 1995. Web4Lib contribution (see archive - 18 Apr 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

DLI. 1996. Digital Libraries Initiative (University of Illinois at Urbana-Champaign) glossary. http://dli.grainger.uiuc.edu/glossary.htm

Ellsworth, J.H. c1994. **Education on the Internet: a hands-on book of ideas, resources, projects, and advice**. Indianapolis: Sams Publishing. 591 p. ISBN 0-672-30595-X.

Erickson, T. 1996. The World-Wide Web as social hypertext. **Communications of the ACM**, 39(1):15-17. ISSN 0001-0782.

Ewers, A. 1994. The European Internet. **In** Smith, N. ed. **Libraries, networks and Europe: a European networking study**. pp. 5 - 28. London: **The British Library**. Research and Development Department. 91p. (Library and Information Research Report 101) ISBN 0 7123 3295 2.

Eysenck, M.W. and M.T. Keane. c1995. **Cognitive psychology: a student's handbook**. 3rd ed. Hove: Erlbaum (UK) Taylor and Francis. 542 p. ISBN 0-86377-375-3.

Fain, B.H. 1995. Web4Lib contribution (see archive - 18 April 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Fidel, R. 1984. Online searching styles: a case-study-based model of searching behaviour. **Journal of the American Society for Information Science**, 35(4):211-221. ISSN 0002-8231.

Fiderio, J. 1988. A grand vision. **Byte**, 13(10): 237-244. ISSN 0360-5280.

Fine, S. 1994. A psychologist's response. **The Journal of Academic Librarianship**, 20(3): 138-139. ISSN 0099-1333.

Ford, A. 1995. **Spinning the Web: how to provide information on the Internet**.London: International Thomson Publishing. 227 p. ISBN 1-850-32141-8.

Foss, C.L. 1989. Tools for reading and browsing hypertext. **Information Processing and Management**, 25(4):407-418. ISSN 0306-4573

Foster, J. ed. 1994. A status report on networked information retrieval: tools and groups.
http://ecsdg.lu.se/pub/standards/internet/rfc/rfc/689.txt

Friedlander, A. 1996. Net gains for digital researchers.**Ariadne**, issue 5, September 1996. ISSN 1361-3200.
http://www.ukoln.ac.uk/ariadne/issue5/digital-researchers.html

Furner-Hines, J .1995. The World-Wide Web in libraries: an overview. **Vine**, 99 (June): 3 - 14. ISSN 0305 5728.

Geyser, E. P. 1992a. Indiscriminate use of the term 'user friendly' and its shortcomings in the evaluation of information retrieval systems. **South African Journal of Library and Information Science**, 60 (2): 80 - 88. ISSN 0256-8861.

Geyser, E. P. 1992b. Human factors in the interaction process between man and the user friendly information retrieval system. **South African Journal of Library and Information Science**, 60 (2): 167 - 173. ISSN 0256-8861.
Gillies, J. 1995. Caught in the Web. CERN Courier, 35(4): 1-4. ISSN 0304-288X.

Gooch, M. 1996. Web4Lib contribution (see archive - 11 July 1996).
http://sunsite.berkeley.edu/Web4Lib/archive.html

Goodman S. E., L. I. Press, S. R. Ruth and A. M. Rutkowski.1994. The global diffusion of the Internet: patterns and problems. **Communications of the ACM,** 37(8): 27 - 31. ISSN 0001-0782.

Grover, V., S.R. Jeong and A.H. Segars. 1996. Information systems effectiveness: the construct space and patterns of application. **Information and Management,** 31: 177-191.

Guthery, S. 1995. Web4Lib contribution (see archive - 15 Apr 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Guy, L. 1995. Web4Lib contribution (see archive - 18 April 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Handley, M. and J. Crowcroft. 1994-5. The World-Wide Web - beneath the surf. UCL Press. ISBN 1-85728-435-6
Also available: http://www.cs.ucl.ac.uk/staff/jon/book/node1.html

Harfoush, N. and Wild, K. (1994) **National information management project - South Africa: report of the Preparatory Mission** Johannesburg: International Development Research Centre, 34p.

Harris, S. 1995. Web4Lib contribution (see archive - 18 April 1995) http://sunsite.berkeley.edu/Web4Lib/archive.html

Hastings, K. and R. Tennant. 1996. How to build a digital librarian. **D-Lib Magazine**, November 1996. ISSN 1082-9873.

http://www.dlib.org/dlib/november96/ucb/11hastings.html

Heath, F. 1995. Libraries, information technology and the future. **Resource Sharing and Information Networks**, 10(1/2): 1-20. ISSN 0737-7797.

Hoff, T. 1996. Web4Lib contribution (see archive - 11 July 1996). http://sunsite.berkeley.edu/Web4Lib/archive.html

Hogle, J. 1996. Web4Lib contribution (see archive - 9 July 1996) http://sunsite.berkeley.edu/Web4Lib/archive.html

Hornung. 1997. Glossary of the Web Dictionary of Cybernetics and Systems http://pespmc1.vub.ac.be/ASC/IndexASC.html

Horton, W. and S. Ilcheva. 1995. A university on the World-Wide Web: the case study of the University of Natal, South Africa. **African Journal of Library, Archives and Information Science**, 5(2): 99-108. ISSN 0795 - 4778.

Hughes, K. 1994. A guide to cyberspace. http://www.eit.com/web/www.guide/guide.01.html

Huston-Somerville, M. and P.A. Kreitz. 1995. Gray sci-tech information resources and information networks: focus on Western Europe. **Resource Sharing & Information Networks**, 10(1/2): 59-76. ISSN 0737-7797.

Huston-Somerville, M. and Wilt, C. C. 1995. Preface. **Resource Sharing & Information Networks**, 10(1/2): xiii-xiv. ISSN 0737-7797.

Hyman, L.W. 1996. Web4Lib contribution (see archive - 15 Nov 1996). http://sunsite.berkeley.edu/Web4Lib/archive.html

ILC. 1994-96. Internet Literacy Consultants Glossary of Internet Terms. http://www.matisse.net/files/glossary.html

Isakowitz, T., E.A. Stohr and P. Balasubramanian. 1995. RMM: a methodology for structured hypermeida design. **Communications of the ACM**, 38(8): 34-43. ISSN 0001-0782.

Jonassen, D. H. 1988. Designing structured hypertext and structuring access to hypertext. **Educational Technology**, November:13-16.

Johnson, J.Q. 1995. Web4Lib contribution (see archive - 9 October 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Kahn, P. 1995. Visual cues for local and global coherence in the WWW. **Communications of the ACM**, 38(8): 67-69. ISSN 0001-0782.

Kahn, R. E. 1994. Viewpoint. **Communications of the ACM** 37 (8): 15 - 19. ISSN 0001-0782.

Kaniki, A. 1996. Virtual digital libraries: whither South African academic libraries? Paper presented at the Conference on Information Technology in Tertiary Education, University of Cape Town, Cape Town, 10-12 April 1996.

Kantor, P.B. 1996. Assessing the factors leading to adoption of digital libraries, and growth in their impacts: the Goldilocks principle. 38th Allerton Institute 1996: libraries, people, and change: a research forum on digital libraries. Monticello, Illinois , October 27-29, 1996
http://edfu.lis.uiuc.edu/allerton/96/kantor.html

Keenan, S. 1996. **Concise dictionary of library and information science**. London: Bowker Saur. 214 p. ISBN 1-85739-022-9.

Kelly, B. 1995. Overview of WWW. Presentation at the "WWW - a Strategic Tool for UK Higher Education" workshop held at Loughborough University on 13/14 February 1995.
http://www.leeds.ac.uk/ucs/people/BKelly/WWW-HE95/overview/paper/overview.html

Keys, M. 1995. Beyond Gutenberg and gigabits: librarians and the emerging digital revolution.**Resource Sharing & Information Networks**, 10(1/2): 21-32. ISSN 0737-7797.

Krol, E. 1994. **The whole Internet: user's guide and catalog** . 2nd ed. Sebastopol CA: O'Reilly & Associates, 543 p. ISBN 1-56592-063-5.

Kuhn, H.C. 1995. Web4Lib contribution (see archive - 10 October 1995).
http://sunsite.berkeley.edu/Web4Lib/archive.html

Lager, M. 1996. Spinning a Web search. Proceedings of the "Untangling the Web" Conference, sponsored by the Librarians' Association of the University of California, Santa Barbara and Friends of the UCSB Library, University Center,University of California, Santa Barbara. April 26, 1996.

http://www.library.ucsb.edu/untangle/lager.html

Levy, D.M. and C.C. Marshall. 1995. Going digital: a look at assumptions underlying digital libraries. **Communications of the ACM**, 38(4): 77-84. ISSN 0001-0782.

Liebscher, P. and G. Marchionini. 1988. Browse and analytical search strategies in a full-text CD-ROM encyclopedia. **School Library Media Quarterly**, summer 1988: 223-233.

Lifer, E.S. 1996. Libraries' crucial role in the 1996 telecomm act. **Library Journal**, 121(5): 30 - 31. ISSN 0363-0277.

Lin, X., P. Liebscher and G. Marchionini. 1991. Graphical representations of electronic search patterns. **Journal of the American Society for Information Science**, 42(7):469-478. ISSN 0002-8231.

Lowe, D.G. 1985. Cooperative structuring of information: the representation of reasoning and debate. **International Journal of Man-Machine Studies**, 23: 97-111.

Lynch, P.J. 1995. Yale C/AIM WWW style manual.
http://info.med.yale.edu/caim/StyleManual_Top.HTML

Magier, D. 1995. Web4Lib contribution (see archive - 23 April 1995).
http://sunsite.berkeley.edu/Web4Lib/archive.html

Makssour, M. 1995. Web4Lib contribution (see archive - 7 April 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Marchionini, G., S. Dwiggins, A. Katz and X. Lin. 1993. Information seeking in full-text end-user-oriented search systems: the roles of domain and search expertise. **Library and Information Science Research**, 15: 35-69. ISSN 0164-0763.

Marchionini, G. 1989. Information-seeking strategies of novices using a full-text electronic encyclopedia. **Journal of the American Society for Information Science**, 40(1): 54-66. ISSN 0002-8231.

Marchionini, G. 1995. **Information seeking in electronic environments**. Cambridge: Cambridge University Press. 224 p. (Cambridge Series on Human-Computer Interaction). ISBN 0-521-44372-5.

Marchionini, G. and H. Maurer. 1995. The roles of digital libraries in teaching and learning. **Communications of the ACM**, 38(4): 67-75. ISSN 0001-0782.

Martell, C. 1994. Sometime soon: a new Renaissance. **The Journal of Academic Librarianship**, 20(3): 129-130. ISSN 0099-1333.

Mathiassen, L. c1996-97. Internet Explorer review. http://www.sharepaper.com/apps/reviews/msie.html

Matrix Information and Directory Services. 1996. Internet statistics. http://www.mids.org/

McEvilly, C.I. 1995. Web4Lib contribution (see archive - 17 October 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

McKeown, P.G. and R.T. Watson. c1996. **Metamorphosis: a guide to the World Wide Web & electronic commerce**. New York: John Wiley & Sons. p. 159. ISBN 0-471-13689-1.

McKiernan, G. 1996. CyberStacks: an alternative model for selecting, organizing, presenting and accessing of WWW resources. Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas. January 19, 1996. http://www.public.iastate.edu/~CYBERSTACKS/OCLC.htm

McLuhan, M. 1962. **The Gutenberg galaxy: the making of typographic man**. London: Routledge & Kegan Paul. p. 293

McLuhan, M. 1964. **Understanding media, the extensions of man**. Cambridge, Mass.: MIT Press. p. 381

Melody, W. 1994. Electronic networks, social relations and the changing structure of knowledge. **In** Crowley, D. and D. Mitchell eds. **Communication theory today**. pp. 254-273. Cambridge: Polity Press. 312 p. ISBN 0 7456 1046 3.

Meyrowitz, J. 1994. Medium theory. **In** Crowley, D. and D. Mitchell eds. **Communication theory today**. pp. 50-77. Cambridge: Polity Press. 312 p. ISBN 0 7456 1046 3.

Meyrowitz, N. 1989. Hypertext - does it reduce cholesterol, too? Hypertext '89 keynote address, Pittsburgh, Pa, 6 Nov. 1989. **Brown University.** Istitute for Research in Information and Scholarship: Technical Report 89-9.

Microsoft Corporation. 1997. Coming soon: Internet Explorer 4.0! http://www.microsoft.com/ie/ie40/

Miller, J. 1996. How the Internet is transforming education. Workshop of the Conference on Information Technology in Tertiary Education, University of Cape Town, Cape Town, 10 -12 April 1996.

Morgan, E.L. 1994. The World-Wide Web and Mosaic: an overview for librarians. **The Public-Access Computer Systems Review** 5(6): 5-26. Available at: http://www.lib.ncsu.edu/staff/morgan/www-and-libraries.html

Morville, P. 1996. Revenge of the Librarians. **Web Review**, 10 May 1996. http://www.webreview.com/96/05/10/webarch/index.html

Nahl, D. 1996. Web4Lib contribution (See archive - 5 June 1996) http://sunsite.berkeley.edu/Web4Lib/archive.html

NCHE. 1996. National Commission of Higher Education discussion document: a framework for transformation. Pretoria: HSRC. 163 p. http://star.hsrc.ac.za/nche/discuss/sgf_toc.html

NCSA. 1996. NCSA digital libraries and information systems: community systems research.

http://csl.ncsa.uiuc.edu/

Ngubane, B. S. 1996. Keynote Address to the Conference on Information Technology in Tertiary Education, University of Cape Town, Cape Town, 10-12 April 1996.

Nielsen, J. 1990. The art of navigating through hypertext. **Communications of the ACM**, 33(3): 296-310. ISSN 0001-0782.

Norman, D.A. and J.C. Spohrer. 1996. Learner-centered education. **Communications of the ACM**, 39(4): 24-27. ISSN 0001-0782.

O'Balle, A. 1995. Web4Lib contribution (see archive - 20 October 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Oswitch, P. 1983. Modelling information system dynamics: a perspective. **Library and Information Science Research**, 5(2): 129-155. ISSN 0164-0763.

Paciello, M. 1995. Web4Lib contribution (see archive - 27 October 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Perkins, M. 1995. Web4Lib contribution (see archive - 10 October 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Peters, P. E. 1995. Where networking is taking us: interconnectivity of systems and convergence of networks. **Resource Sharing & Information Networks**, 10(1/2): 49-58. ISSN 0737-7797.

Press, L. 1995. McLuhan meets the Net. **Communications of the ACM**, 38(6): 15-20. ISSN 0001-0782.

Proud, M. 1995. Web4Lib contribution (see archive - 21 Nov 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Prytherch, R.J. 1995. **Harrod's librarians' glossary**. 8th ed. Aldershot:Gower. 692 p. ISBN 0-566-07533-4.

Quinn, D.B. 1994. The information age: another giant step backward. **The Journal of Academic Librarianship**, 20(3): 134-135. ISSN 0099-1333.

Rada, R. 1989. Writing and reading hypertext: an overview. **Journal of the American Society for Information Science**, 40(3):164-171. ISSN 0002-8231.

Randall, N. 1995. Search Engines: Powering through the Internet. **PC Computing,** September 1995. http://www.zdnet.com/pccomp/features/internet/search/index.html

Reed, M.J.P. 1976. Computer simulation: a tool for analysis of library service. **Journal of Library Automation**, 9(2): 117-136. Issn 0022-2240.

Relihan, L. 1994. Untangling the World-Wide Web. http://itdsrv1.ul.ie/Research/WWW/utwww.html

Ricart, G. 1994. Free the Internet! **Computers in Physics**, 8(5): 508 - 510. ISSN 0894-1866.

Ricart, G. 1995. The scholarly information Web. **Computers in Physics**, 9(4): 360 - 363. ISSN 0894-1866.

Rosenfeld, L.B. 1996. Stupid agents: a vain search for the grail. **Web Review**, 11 October 1996.
http://www.webreview.com/96/10/11/arch/index.html

Rutkowski, A.M. 1996. Internet trends: a summary of the latest Internet surveys. http://www.genmagic.com/Internet/Trends/index.html

Samuelson, P. 1995. Copyright and digital libraries. **Communications of the ACM**, 38(3):15-21,110. ISSN 0001-0782.

Saracevic, T., P. Kanor, A.Y. Chamis and D. Trivison. 1988. A study of information seeking and retrieving. Background and methodology. **Journal of the American Society for Information Science**, 39(3):161-176. ISSN 0002-8231.

Saracevic, T. and P. Kanor. 1988. A study of information seeking and retrieving. Searchers, searches and overlap. **Journal of the American Society for Information Science**, 39(3):197-216. ISSN 0002-8231.

Schank, R. C. and A. Kass. 1996. A goal-based scenario for high school students. **Communications of the ACM**, 39(4): 28-29. ISSN 0001-0782.

Schement, J.R. 1996. A 21st-century strategy for librarians. **Library Journal**, 121(8): 34-36. ISSN 0363-0277.

Schneider, K.G. 1995. Web4Lib contribution (see archive - 10 May 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Schneider, K.G. 1996. Web4Lib contribution (See archive - 10 July 1996). http://sunsite.berkeley.edu/Web4Lib/archive.html

Scoville, R.1996. Find it on the Net. **PC World**, Jan. 1996. http://www.pcworld.com/reprints/lycos.htm

Secret, A. 1996. The Virtual Library. http://www.w3.org/vl/coordination.html

Segal, B.M. 1995. A short history of Internet protocols at CERN. http://www.cern.ch/pdp/ns/ben/TCPHIST.html

Shearin, R. 1994. Government and industry get together on the superhighway. **Computers in Physics**, 8 (6): 630-631. ISSN 0894-1866.

Shillinglaw, N. (1988) "South Africa as an Information Society" in Shillinlaw, N. and Thomas, W. **The Information Society**. Parklands: Ad. Donker, pp. 9-20.

Shreeves, E. 1994. Embracing the inevitable. **The Journal of Academic Librarianship**, 20(3): 136-137. ISSN 0099-1333.

Slot, M. 1996. Web matrix: what's the difference?some answers about search engines and subject catalogs. http://www.sils.umich.edu/~fprefect/matrix/answers.html

Smith, J.B., S.F. Weiss and G.J. Ferguson. 1987. A hypertext writing environment and its cognitive basis. University ofNorth Carolina technical report 87-033. 14p.

Smith, N. 1994. Introduction. **In** Smith, N. ed. **Libraries, networks and Europe: a European networking study**. pp. 3 - 4. London: **The British Library**. Research and Development Department. 91p. (Library and InformationResearch Report 101) ISBN 0 7123 3295 2.

Soloway, E. 1994. Ways of seeing. **Communications of the ACM**, 37(2): 15-20. ISSN 0001-0782.

Soloway, E. 1996. Teachers are the key. **Communications of the ACM**, 39(6): 11-14. ISSN 0001-0782.

Soloway, E. and Pryor, A. 1996. The next generation in human-computer interaction. **Communications of the ACM**, 39(4): 16-18. ISSN 0001-0782.

SOUTH AFRICA. Department of Arts, Culture, Science and Technology. 1996. **South Africa's green paper on science and technology: preparing for the 21st century**. Pretoria.108 p. ISBN 0-621-17321-5.

Stanley, T. 1996. Alta Vista vs. Lycos. **Ariadne**, issue 2, March 1996. ISSN 1361-3200.
http://ukoln.bath.ac.uk/ariadne/issue2/engines

Stix, G. 1994. The speed of write. **Scientific American**, 271(6): 72-77. ISSN 0036-8733.

SUNSITE Manager. 1996.Web4Lib archive.
http://sunsite.berkeley.edu/Web4Lib/archive.html

Taylor, A.G. and P. Clemson. 1996. Access to networked documents: Catalogs? Search Engines? Both? Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas. January 19, 1996. http://www.oclc.org/oclc/man/colloc/taylor.htm

Tennant, R. 1996a. The Berkeley Digital Library SunSITE. **D-Lib Magazine,** February 1996. ISSN 1082-9873 http://www.ukoln.ac.uk/dlib/dlib/february96/ucb/02tennant.html

Tennant, R. 1996b. Web4Lib: the library web manager's electronic discussion list. **Ariadne**, issue 5, September 1996. ISSN 1361-3200. http://www.ukoln.ac.uk/ariadne/issue5/web4lib/

Tennant, R. 1996c. Web4Lib contribution (see archive -12 November 1996). http://sunsite.berkeley.edu/Web4Lib/archive.html

Thompson, J.B. 1994. Social theory and the media. **In** Crowley, D. and D. Mitchell eds. **Communication theory today**. pp. 27-49. Cambridge: Polity Press. 312 p. ISBN 0 7456 1046 3.

Thüring, M., Hannemann, J., and J.M. Haake. 1995. Hypermedia and cognition: designing for comprehension. **Communications of the ACM**, 38(8): 57 - 66. ISSN 0001-0782.

Tillman, H.N. 1996. Evaluating quality on the Net. http://www.tiac.net/users/hope/findqual.html

Tilton, J. 1996. Composing good HTML.
http://www.cs.cmu.edu/~tilt/cgh/

URN implementors. 1996. Uniform resource names: a progress report. **D-Lib Magazine**, February 1996. ISSN 1082-9873.
http://www.ukoln.ac.uk/dlib/dlib/february96/02arms.html

Van Brakel, P.A. 1994. Teaching hypertext techniques with Mosaic and WWW. Preliminary proceedings of the First International Conference on the World-Wide Web held at CERN, Geneva (Switzerland). May 25-26-27 1994.
http://www.cern.ch/WWW94/PrelimProcs.html

Van House, N.A., M.H. Butler, V. Ogle and L. Schiff. 1996. User-centered iterative design for digital libraries: the Cypress experience. **D-Lib Magazine**, February 1996. ISSN 1082-9873
http://www.ukoln.ac.uk/dlib/dlib/february96/02vanhouse.html

Vezza, A. 1996. Keio University joins the MIT Laboratory for Computer Science and INRIA in hosting the International World Wide Web Consortium.
http://www.w3.org/pub/WWW/Keio-PR.html

Voss, D. 1994. Buiding the US superhighway. **Physics World**, 7 (9): 43 - 47. ISSN 0953-8585.

Wallich, P. 1994. Wire pirates. **Scientific American**, 270 (3): 72 - 80 ISSN 0036-8733.

Weibel, S. 1995. Web4Lib contribution (see archive - 17 October 1995). http://sunsite.berkeley.edu/Web4Lib/archive.html

Weibel, S., J. Godby and E. Miller. 1995.OCLC/NCSA metadata workshop report. Metadata Workshop, sponsored by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA), held in Dublin, Ohio, March 1-3, 1995. http://www.oclc.org:5047/oclc/research/conferences/metadata/dublin_core_report.htm

WGLIT. 1996. Policy, planning and co-operation: smart solutions for information provision - a report by the National Commission on Higher Education Working Group on Libraries and Information Technology. http://star.hsrc.ac.za/nche/wglit/final/toc.html

Wisner, W. H. 1994. Symposium: back toward the people. **The Journal of Academic Librarianship**, 20(3): 131-133. ISSN 0099-1333.

Yankelovich, N., B.J. Haan, N.K. Meyrowitz and S.M. Drucker. 1988. Intermedia: the concept and the construction of a seamless information environment. **(IEEE) Computer**, 21(1): 81-96.

Zakon, R.H. c1993-6. Hobbes' Internet timeline. http://info.isoc.org/guest/zakon/

Appendix A - Author's Correspondence with Web4Lib

Subject: Request to use web4lib messages

Author: ilcheva@unpsun1.cc.unp.ac.za at INTERNET

Date: 5/27/96 3:36 AM

Dear Web4lib subscribers,

I have been a silent avid reader of your discussion for a year now and I have come to realize its tremendous educational potential. Here in South Africa most of the librarians have had very little and teachers - practically no exposure to the WWW even at the traditionally advantaged libraries and schools. They need to be made aware of how powerful the WWW can be as a provider of democratic access to the world's information resources, communication and collaboration. Otherwise, as with any new medium, there is bound to be a lot of negativism and obstruction. Financial and other constraints might be overplayed to slow down its advent.

Therefore I feel it will be beneficial if librarians and teachers in South Africa can get an easy, hands-on, inexpensive and non-threatening introduction to the wealth of cyberspace before deciding on whether they need an Internet connection. As a research student in Information Studies I am working on a project involving the creation of a mini model of the WWW that could be used as an instructional tool to introduce librarians who are new to this environment to its nature, salient characteristics, concepts, terminology; its available and potential services and uses; its functions, operations,

navigation and retrieval strategies; its authoring capabilities.

My request is:

1. Will it be OK if I use a selection (in the region of 100) of your messages in this model? Duly acknowledged, of course.

2. If yes, could I leave your e-mail and snail-mail addresses in?

3. Could I download and make links to some of the documents recommended by you?

4. Would you object to the model being mounted on a University of Natal, Pietermaritzburg server for public access?

5. If not, would you care to have a look at and comment on it?

Your cooperation and help will be greatly appreciated.

Stoya

--

Stoya Ilcheva
University of Natal, Pietermaritzburg, Natal, South Africa
ilcheva@unpsun1.cc.unp.ac.za

* * * * * * * * *

Date: Tue, 28 May 96 08:34:15 PST

From: lbickham@stem.com

To: ilcheva@unpsun1.cc.unp.ac.za

Subject: Re: Request to use web4lib messages

Stoya,

I'm sure you'll hear from whoever manages the Web4lib listserv, but I think what you are doing sounds great -- it sounds like your project could make a real difference for people trying to decide about the value of the Web. I am a librarian in a small biotech company, but I suspect you will hear from librarians in large University libraries who will have a lot of ideas to share with you.

Good luck in your endeavor

Linda Bickham

lbickham@stem.com

SyStemix

Manager, Library Services

* * * * * * * * *

Date: Mon, 27 May 1996 14:16:16 -0400 (EDT)
From: "C. W. Tazewell" <cwt@exis.net>
To: Multiple recipients of list <web4lib@library.berkeley.edu>
cc: Stoya Ilcheva <ilcheva@unpsun1.cc.unp.ac.za>
Subject: Educational Potential of WWW

Hi,

Your message about the educational potential of the WWW is very, very good. The same problem exists everywhere!

I operate (alone) a major WWW Digital Library. It serves an important metropolitan area in Southeastern Virginia. I have not been able to get any assistance or the necessary widespread support for The Hampton Roads Central Library. I envision The Library also as an area-wide centralized K12 media center, since much of the same material can be shared by students and the public.

The general attitude about the Internet and The Library is that "maybe it will just go away." And, when something is done - setting up a home page, etc. - it is usually tokenism, just to keep up with the latest fad.

I will be glad to assist you in any way I can. You may like to check The Library: Internet 2000, Professional Librarians' Section, and the Tidewater Virginia Public Libraries Home Page. (The latter was set up by me, but it has never been accepted by the professional librarians.) One of my problems is that I am not a professional librarian, although I

have almost 75 years experience as a user and manager. I founded and was first head of a library system that now serves a million people.

Also, my South Africa page may be of interest. I try to tie in local Hampton Roads and Virginia activities and people with other parts of the world. I have done the same thing with the Scottish Section. I have relatives in South Africa - the family of former Ambassador Taswell.

I feel that The Hampton Roads Central Library is (among other things) a powerful advocate to the world for this area. But, the major purpose is to make it easy for "Jack and Jill and Little Joey" to use the overwhelming resources of the Internet.

One thing that concerns me locally is our local Ruffner Middle School Internet Project. They have 1000 students and over 6 dozen teachers. They have more than 300 computers connected to the Internet. They apparently feel that they can sit the students and teachers in front of a computer and the Internet will work like magic - somehow. My philosophy is that you don't adapt people to the Internet; you adapt the Internet to the people.

When $500 Network Computers are connected to the Internet in every home, the WWW has to be point and click with a TV remote comtrol. There are too many people in the world for them all to become Web surfers. That's for the professionals.

USENET News Groups messages are available in archived form, as an example from DejaNews. DejaNews can be accessed thru Yahoo, and direct at "www.dejanews.com ." You can just link messages without including the actual file.

Perhaps what you need is a local/community/metropolitan digital

library!

Good luck. You'll need it.

Bill.

Will print and libraries survive the Internet?

The Hampton Roads Central Library

(You don't drive to it; you point to it.)

Your *REASON* to have the Internet

Front Entrance E A S Y L I N K S - Your WWW Hotlist

http://wwwp.exis.net/~cwt/ http://www.infi.net/~cwt/easylink.html

* * * * * * * * *

Date: Tue, 04 Jun 1996

From: Elisabeth Roche <ace@Opus1.COM>

Subject: Re: Educational Potential of WWW

You should get in contact with the Maricopa College in Arizona, at:

http://www.mcli.dist.maricopa.edu/

Elisabeth Roche ace@opus1.com

serendipity RULES!

* * * * * * * * *

MICRO-WEB CONTENTS PAGE

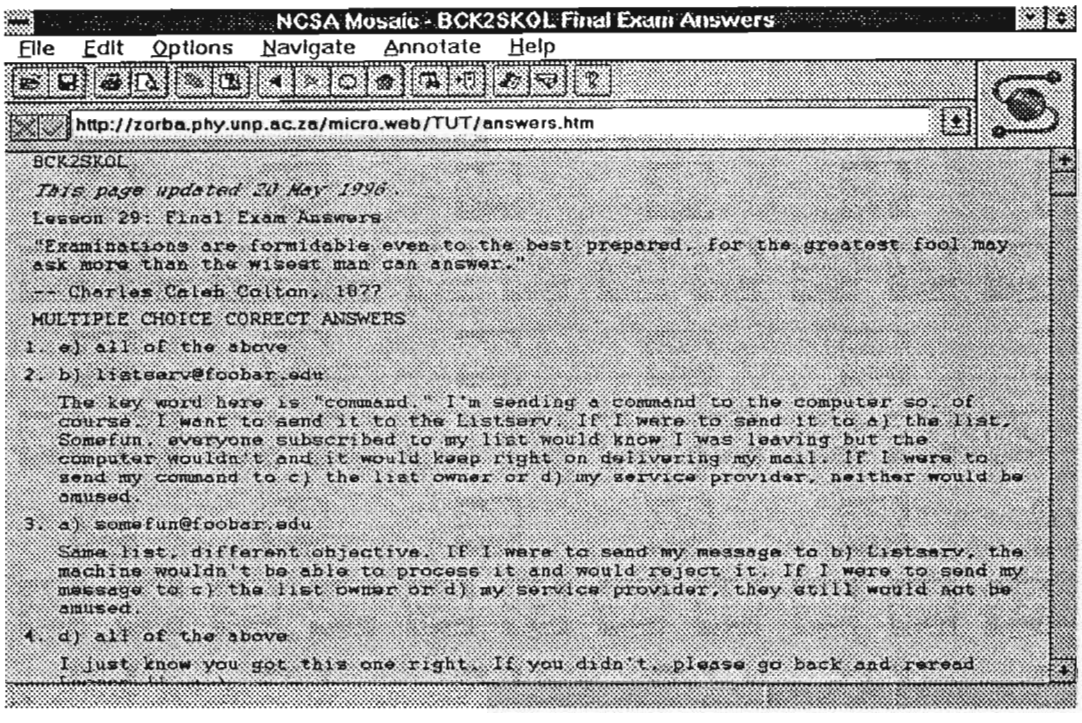
- doing precision searches for Internet resources
- archiving Internet resources
- providing for special user needs
- managing the storage migration and links integrity of electronic documents
- ***user education:***
    - creating awareness of the services offered by the Internet
    - navigating the Web
    - searching for Internet resources
    - evaluation of Internet resources
    - publishing on the Net
    - homepage design
    - homepage guidelines
    - homepage maintenance

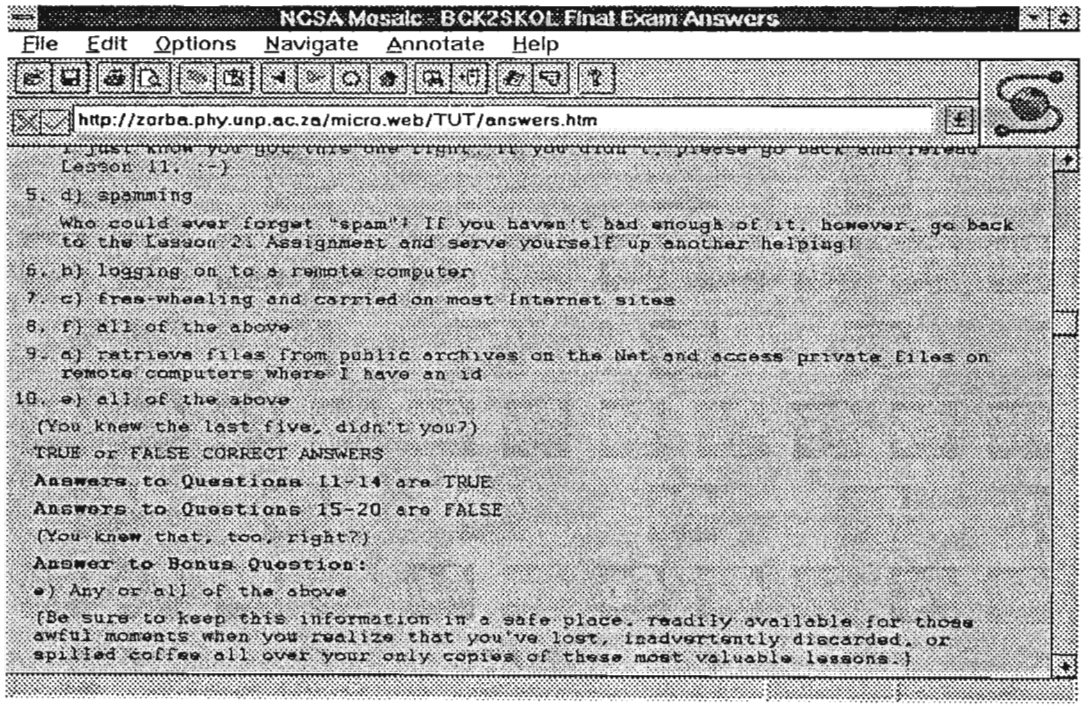
4. LIBRARY VALUE-ADDED SERVICES ON THE INTERNET:

- OCLC, Internet cataloguing project, NetFirst, FirstSearch
- hypertext thesauri
- Library of Congress subject headings and class schedules

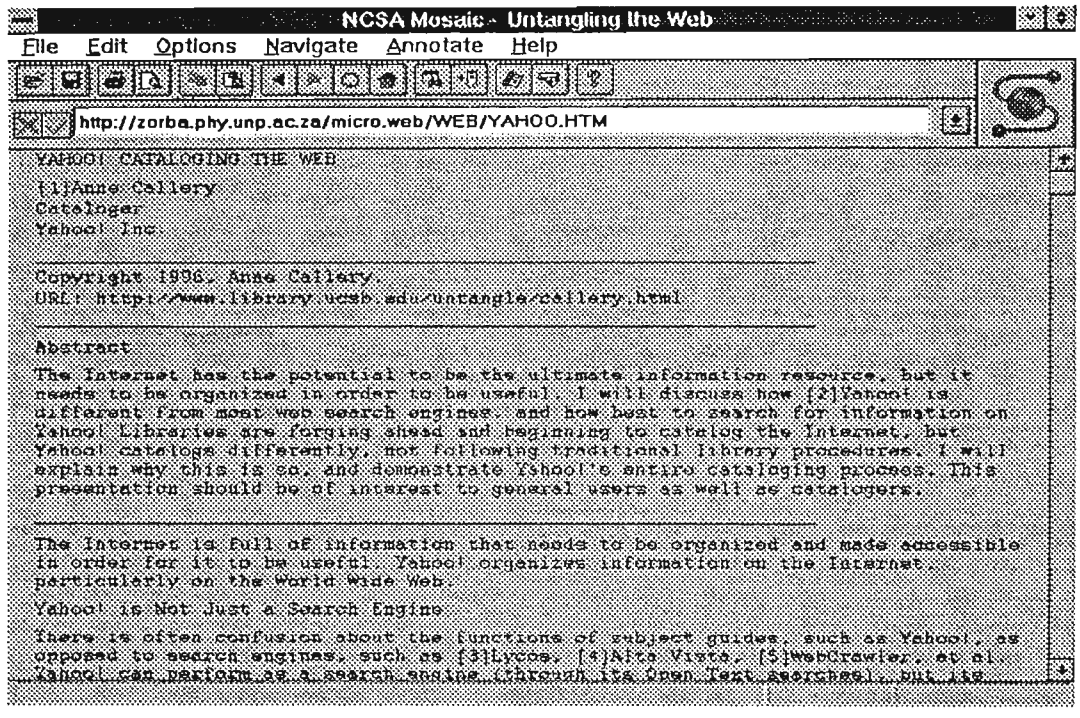
Back to homepage.

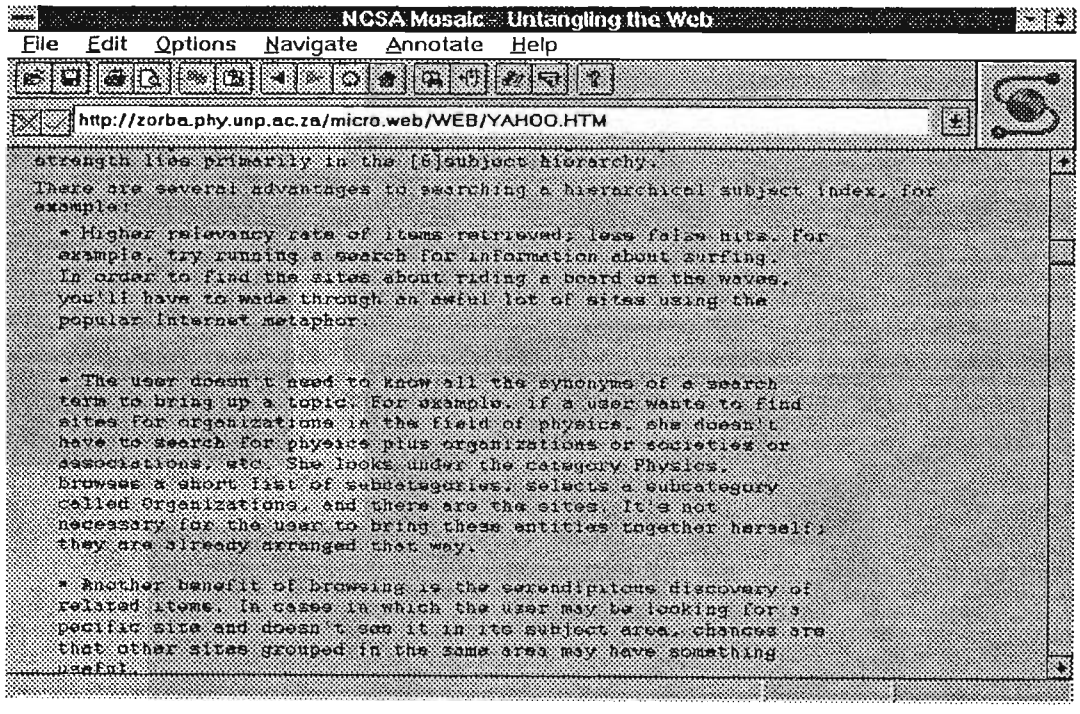
Appendix C - Target 1



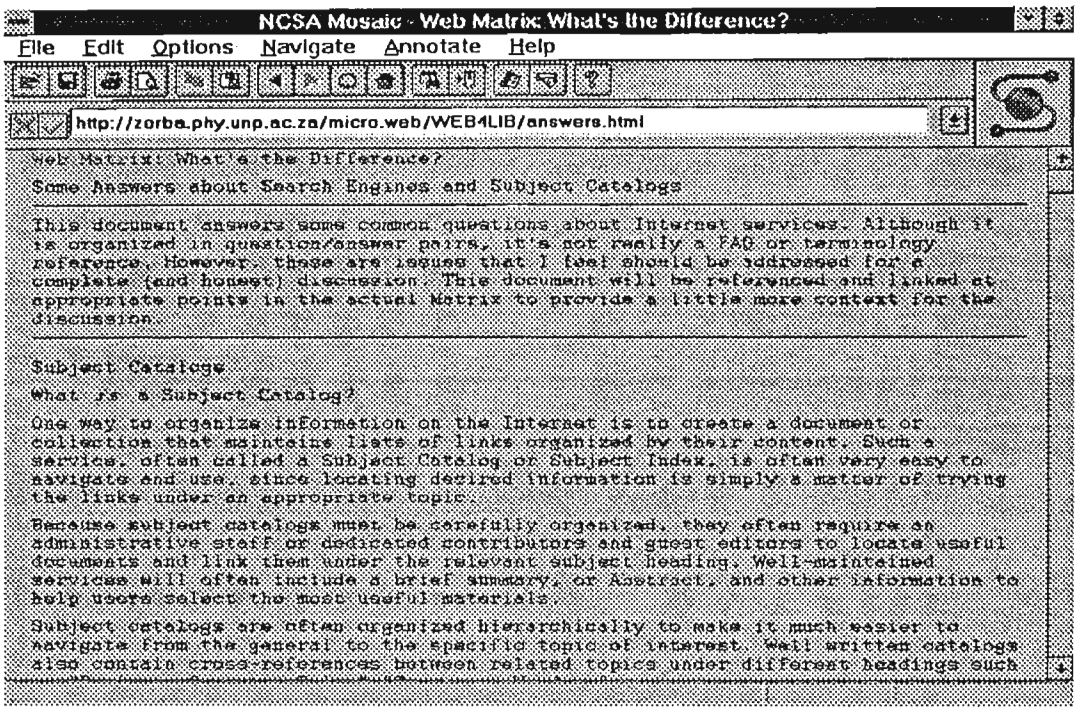


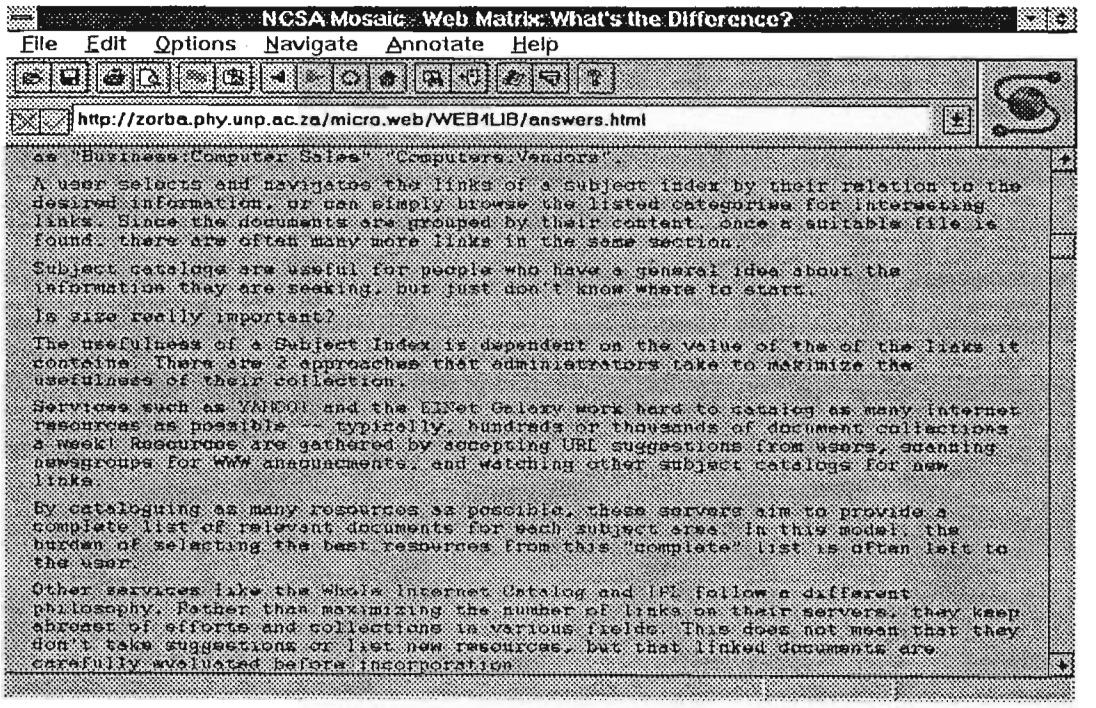
Appendix D - Target 2



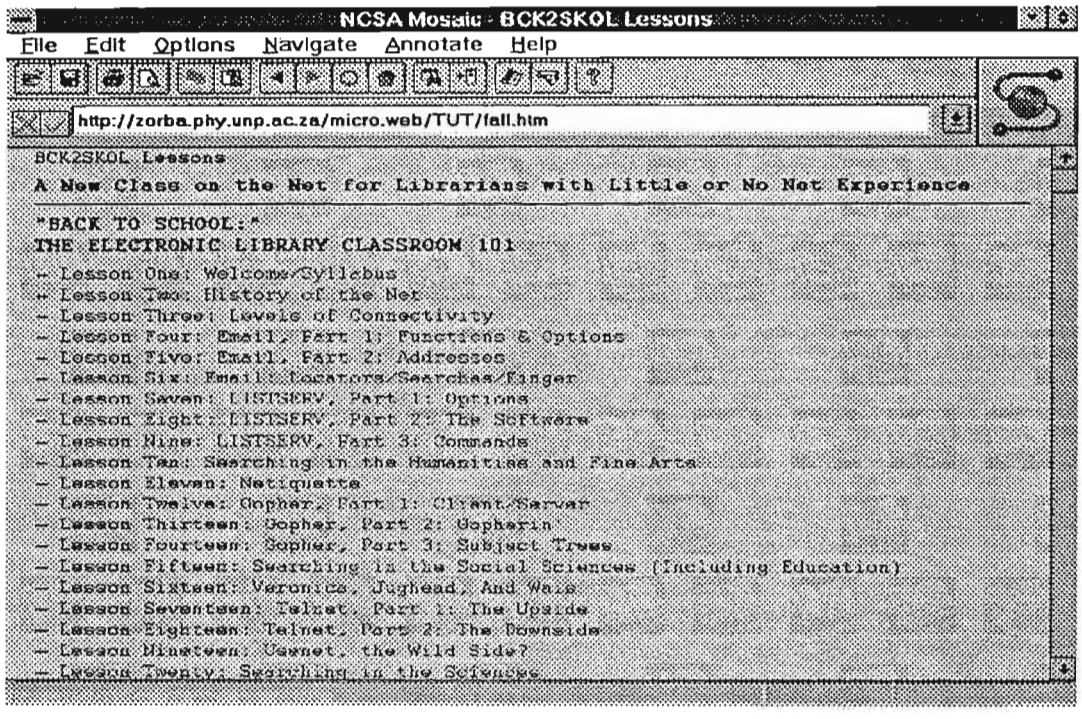


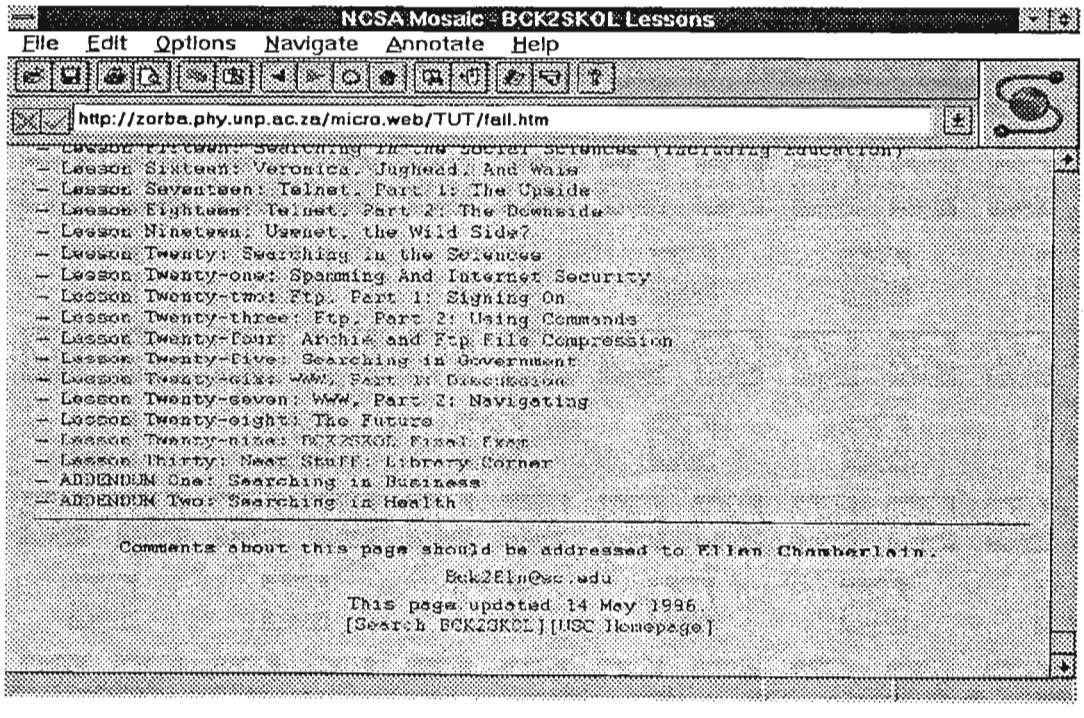
Appendix E - Target 3



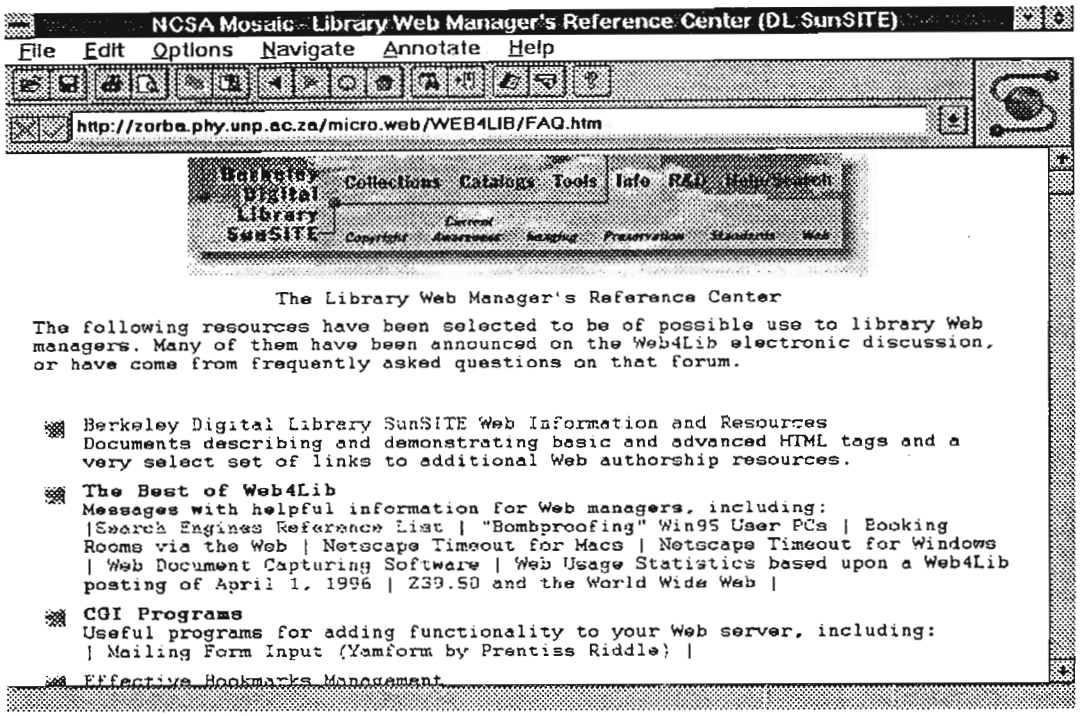


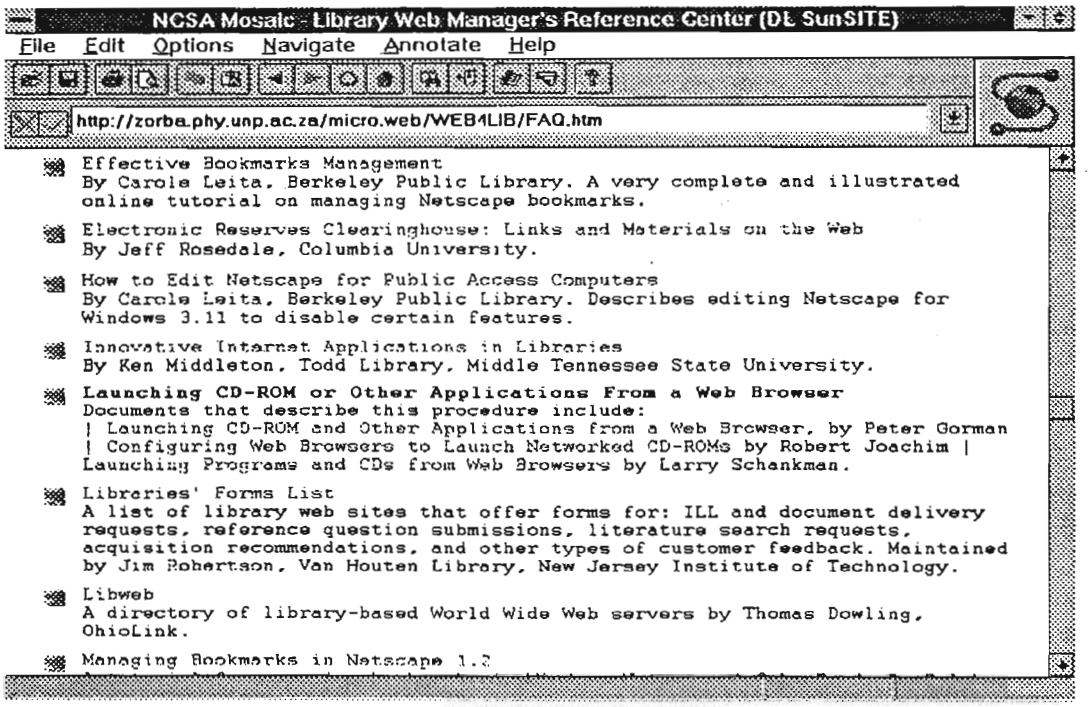
Appendix F - Nav. 1

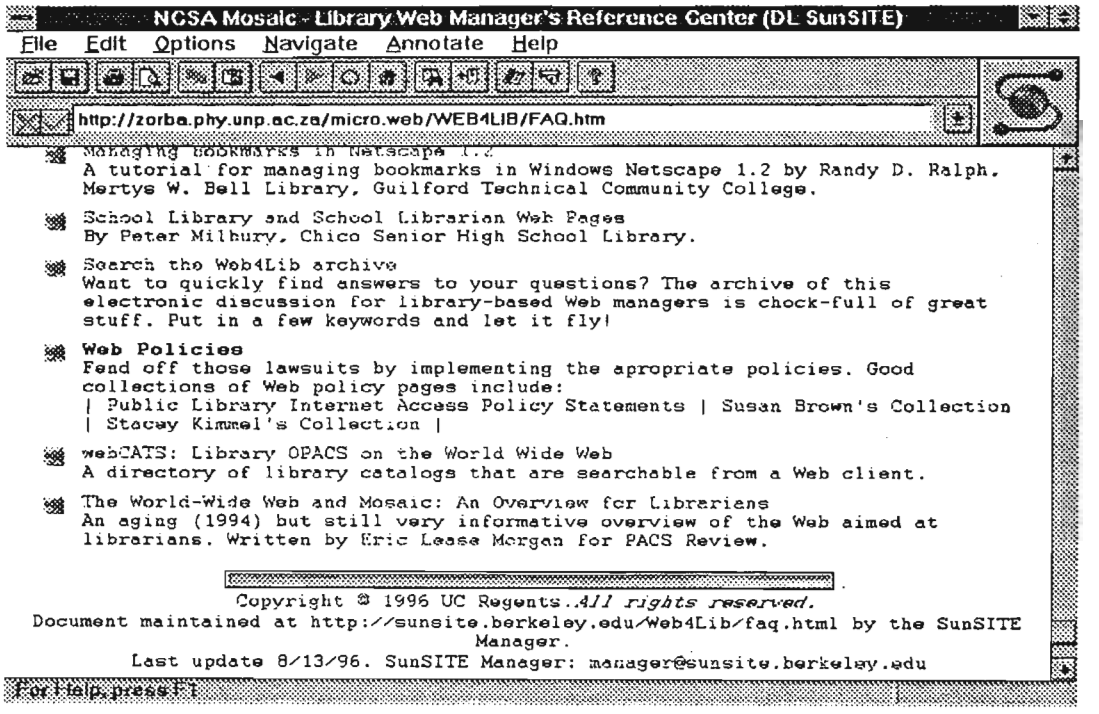




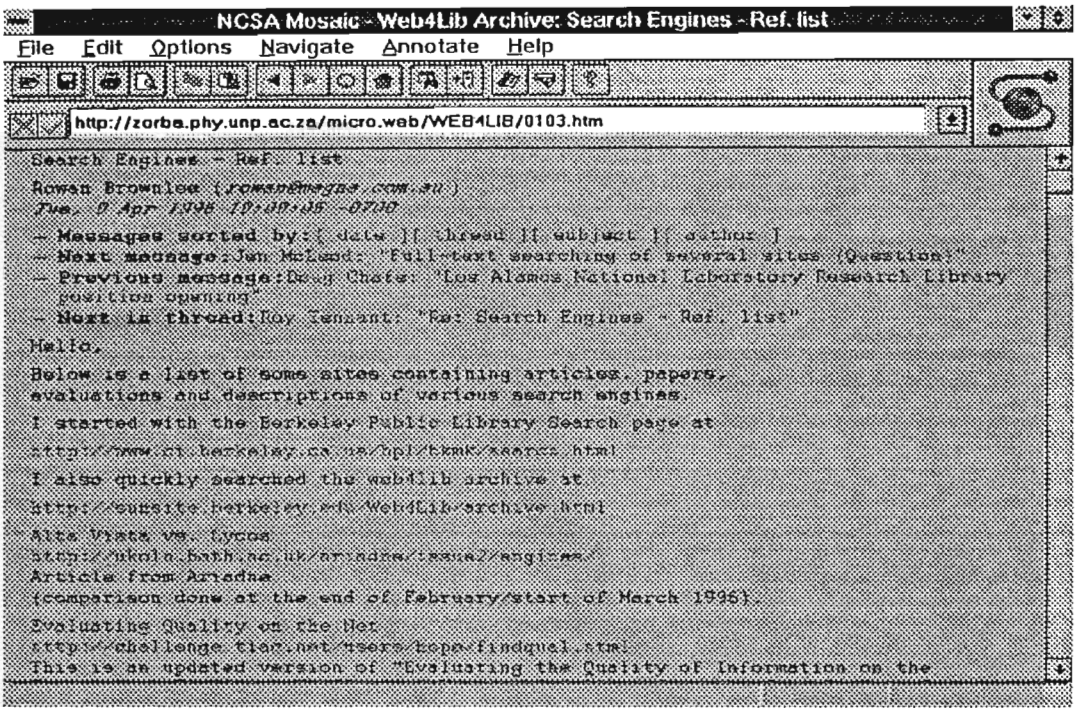
Appendix G - Nav. 2

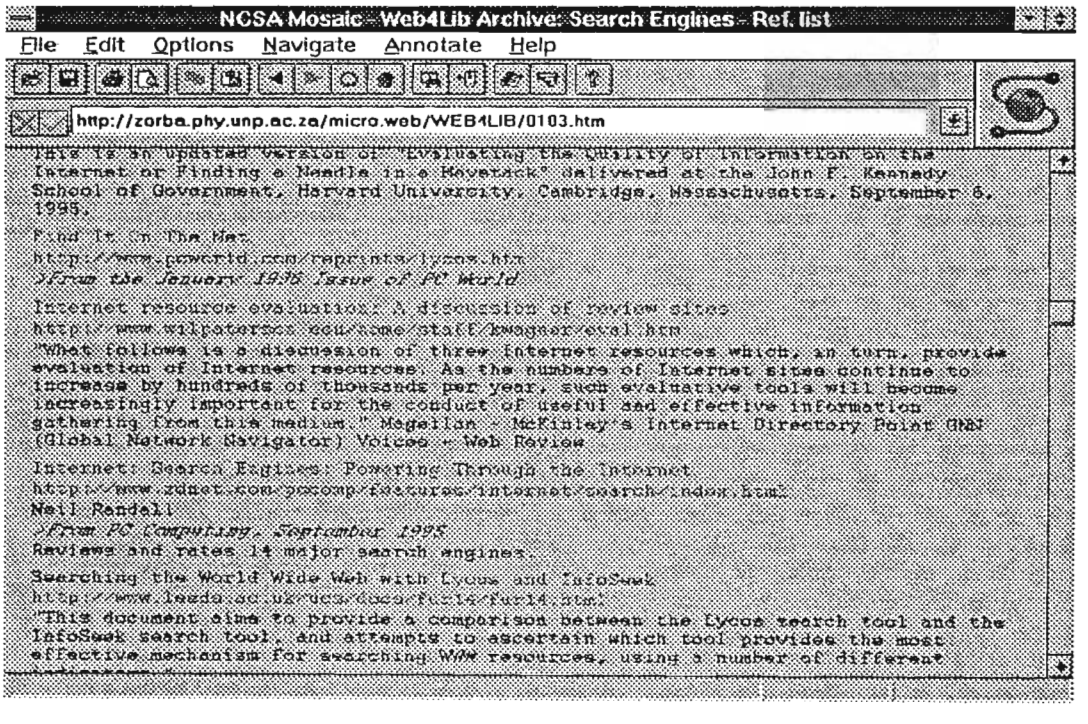


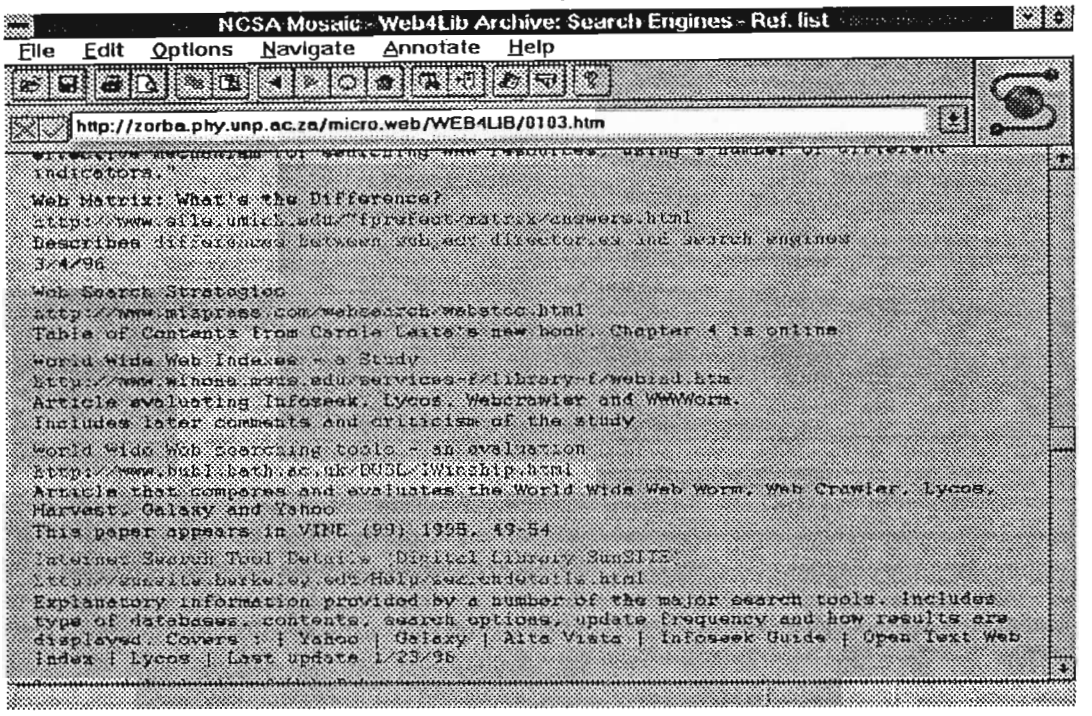


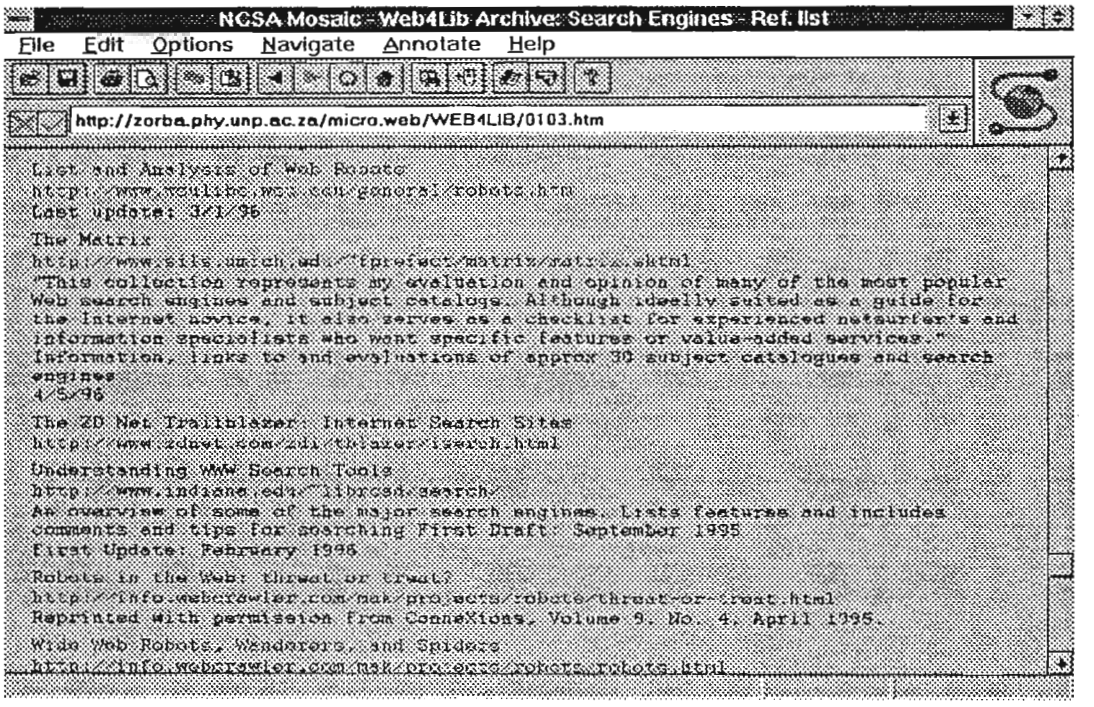

NCSA Mosaic - Library Web Manager's Reference Center (DL SunSITE)
File Edit Options Navigate Annotate Help
http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/FAQ.htm
Managing bookmarks in Netscape 1.2
A tutorial for managing bookmarks in Windows Netscape 1.2 by Randy D. Ralph, Mertys W. Bell Library, Guilford Technical Community College.
School Library and School Librarian Web Pages
By Peter Milbury, Chico Senior High School Library.
Search the Web4Lib archive
Want to quickly find answers to your questions? The archive of this electronic discussion for library-based Web managers is chock-full of great stuff. Put in a few keywords and let it fly!
Web Policies
Fend off those lawsuits by implementing the apropriate policies. Good collections of Web policy pages include:
| Public Library Internet Access Policy Statements | Susan Brown's Collection | Stacey Kimmel's Collection |
webCATS: Library OPACS on the World Wide Web
A directory of library catalogs that are searchable from a Web client.
The World-Wide Web and Mosaic: An Overview for Librariens
An aging (1994) but still very informative overview of the Web aimed at librarians. Written by Eric Lease Morgan for PACS Review.
Copyright © 1996 UC Regents. All rights reserved.
Document maintained at http://sunsite.berkeley.edu/Web4Lib/faq.html by the SunSITE Manager.
Last update 8/13/96. SunSITE Manager: manager@sunsite.berkeley.edu
For Help, press F1

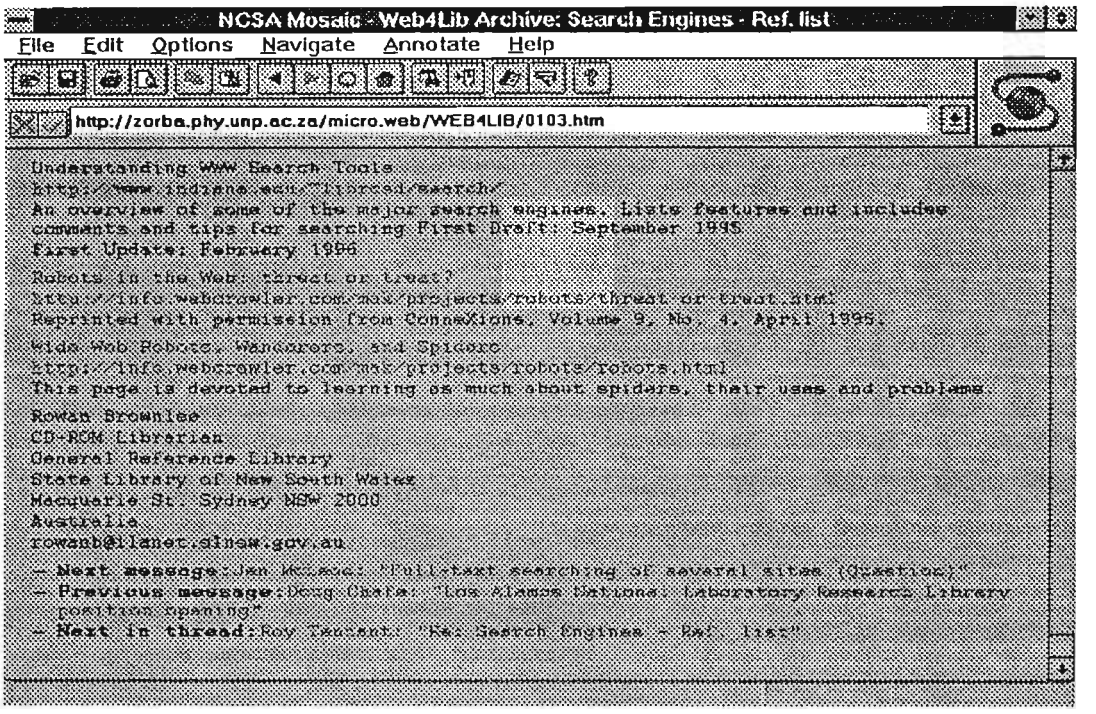
Appendix H - Nav. 2a






NCSA Mosaic - Web4Lib Archive: Search Engines - Ref. list
File Edit Options Navigate Annotate Help
http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/0103.htm
indicators."
Web Matrix: What's the Difference?
Describes differences between web.edu directories and search engines
3/4/96
Web Search Strategies
Table of Contents from Carole Leita's new book. Chapter 4 is online
World Wide Web Indexes - a Study
Article evaluating Infoseek, Lycos, Webcrawler and WWWorm.
Includes later comments and criticism of the study
World Wide Web searching tools - an evaluation
Article that compares and evaluates the World Wide Web Worm, Web Crawler, Lycos, Harvest, Galaxy and Yahoo
This paper appears in VINE (99): 1995, 49-54
Internet Search Tool Details (Digital Library SunSITE
Explanatory information provided by a number of the major search tools. Includes type of databases, contents, search options, update frequency and how results are displayed. Covers : | Yahoo | Galaxy | Alta Vista | Infoseek Guide | Open Text Web Index | Lycos | Last update 1/23/96


NCSA Mosaic - Web4Lib Archive: Search Engines - Ref. list
File Edit Options Navigate Annotate Help
http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/0103.htm
List and Analysis of Web Robots
Last update: 3/1/96
The Matrix
"This collection represents my evaluation and opinion of many of the most popular Web search engines and subject catalogs. Although ideally suited as a guide for the Internet novice, it also serves as a checklist for experienced netsurfer's and information specialists who want specific features or value-added services."
Information, links to and evaluations of approx 30 subject catalogues and search engines
4/5/96
The ZD Net Trailblazer: Internet Search Sites
Understanding WWW Search Tools
As overview of some of the major search engines. Lists features and includes comments and tips for searching First Draft: September 1995
First Update: February 1996
Robots in the Web: threat or treat?
Reprinted with permission from ConneXions, Volume 9, No. 4, April 1995.
Wide Web Robots, Wanderers, and Spiders


NCSA Mosaic - Web4Lib Archive: Search Engines - Ref. list
File Edit Options Navigate Annotate Help
http://zorba.phy.unp.ac.za/micro.web/WEB4LIB/0103.htm
Understanding WWW Search Tools
http://www.indiana.edu/~librcsd/search/
An overview of some of the major search engines. Lists features and includes
comments and tips for searching First Draft: September 1995
First Update: February 1996
Robots in the Web: threat or treat?
http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html
Reprinted with permission from ConneXions, Volume 9, No. 4, April 1995.
Wide Web Robots, Wanderers, and Spiders
http://info.webcrawler.com/mak/projects/robots/robots.html
This page is devoted to learning as much about spiders, their uses and problems.
Rowan Brownlee
CD-ROM Librarian
General Reference Library
State Library of New South Wales
Macquarie St. Sydney NSW 2000
Australia
rowanb@ilanet.slnsw.gov.au
— Next message: Jen McCane: "Full-text searching of several sites (Question)"
— Previous message: Doug Cutts: "Los Alamos National Laboratory Research Library position opening"
— Next in thread: Roy Tennant: "Re: Search Engines - Ref. list"

Appendix I - The Three Targets Task Questions

1. Target 1 Task Questions

1. I have come to depend upon email because it is:

   a. the most basic function of networks, and available to users with even the lowest levels of connectivity
   b. the means by which I communicate electronically on the same network and, through 'gateways', to other networks
   c. my access to academic listservs
   d. the means by which I can deliver BCK2SKOL to you as a distribution listserv
   e. all of the above

2. Lately, I've been busy flooding the Net with files advertising my personal business. What have I been up to?

   a. lurking
   b. hacking
   c. fingering
   d. spamming
   e. surfing

3. When I telnet, I am:

a. subscribing to a newsgroup or discussion list
b. logging on to a remote computer
c. transferring files from a remote computer to my own
d. all of the above

4. I know that Usenet newsgroups are read heavily because they are:

a. email-based discussions like listservs
b. censored across-the-board
c. free-wheeling and carried on most Internet sites
d. owned by Ted Turner, but being bought out by Rupert Murdock

5. The World Wide Web allows me to:

a. jump from document to document via hypertext
b. hear audios and see videos
c. search all of cyberspace via gopher, telnet, ftp, etc. without having to remember/use arcane commands
d. create my own documents with my choice of audio, video and textual links
e. all of the above

6. T F The Internet was designed by the U.S. government in

the 1950's to spy on individuals suspected of being communists.

7. T F I know email messages are secure, so I can safely go ahead, let my hair down, and say anything I want.

8. T F To finger someone is to report them to the police.

2. Target 2 Task Questions

1. Yahoo is different from other efforts to organise the information resources on the Internet in that it combines:
   a) a search engine and
   b) .....
2. How is the material organized in 1(b)?

.....

3. Does Yahoo try to catalog everything?

.....

4. Does it conform to standard library classification schemes?

.....

5. Who determines the information categories?

.....

6. What 7 bibliographic description fields are maintained on Yahoo?

.....

.....

7. In what 2 ways can you search for information on Yahoo?

.....

8. What 3 parts do the Yahoo search results display?

.....

3. Target 3 Task Questions

1. What is better to use if you have only a general idea about the information you are seeking on the Internet and do not know where to start:

a) a subject index/directory

b) a search engine

2. How do you locate desired information on the Internet if you are using:

a) a subject index/directory

b) a search index

3. What must be added to the size of the database to ensure the usefulness of an Internet indexing service in the form of:

a) a subject index

b) a search engine

Appendix J - Raw Data Tables

Table 1: Task 2 browsing sessions - detailed navigational moves data

Subject	Total number of navigat-ional moves	Number of links chosen					Number of backtracking moves		
		Total	From Cont. Page	Sole links in docs	First links in docs	Other links in docs	Total	Backtrack. to Cont. Page	Internal path backtrack.
1	47	32	15	1	6	10	15	13	2
2	34	27	5	4	6	12	7	3	4
3	44	31	12	4	5	10	13	10	3
4	36	25	11	1	4	9	11	9	2
5	31	24	7	5	2	10	7	5	2
6	24	16	5	1	5	5	8	3	5
7	26	18	7	1	5	5	8	5	3
8	27	18	10	0	3	5	9	7	2
9	93	57	32	7	5	13	36	30	6
10	91	49	39	1	6	3	42	36	6

Table 2: Task 2 browsing sessions - detailed subjects' browsing behaviour data

Subject	Inter-document browsing behaviour: Contents Page reading tactics					Across-document browsing behaviour			
	Total number of CP visits	Number of false starts	Number of useful path starts	Number of reading direction changes	Number of entry points repetitions	Total number of internal docs visited	Number of short documents	Number of long documents	Average length of useful paths
1	14	9	5	4	0	33	19	14	4.8
2	4	1	3	2	0	30	27	3	9.67
3	11	7	4	5	0	33	22	11	6.5
4	10	4	6	3	2	26	15	11	3.67
5	6	2	4	1	0	25	19	6	5.75
6	4	2	2	3	0	20	19	1	9
7	6	2	4	3	2	20	5	15	4.5
8	9	6	3	2	2	19	0	19	4.33
9	31	23	8	17	16	62	44	18	4.88
10	38	32	6	14	26	54	37	17	3.67

Table 3: Task 2 browsing sessions - subjects' searching strategies types and success rates

Sub ject	Session duration	Subject seaching strategies types	Success rate - number of navigational pages and target hits			
			Total	Nav.2 hits	Nav.2a hits	Target 2 hits
1	45'	operationalist methodical	4	1	3	0
2	47'	conceptualist	2	1	1	0
3	53'	conceptualist	5	3	2	0
4	43'	conceptualist	2	2	0	0
5	50'	conceptualist	3	2	1	0
6	40'	conceptualist	1	1	0	0
7	68'	conceptualist	3	1	1	1
8	40'	conceptualist	5	2	2	1
9	70'	operationalist overview	7	4	2	1
10	62'	operationalist overview	4	3	1	0
Totals:			**36**	**20**	**13**	**3**

Appendix K - Definitions of Mean Value, Standard Deviation and Correlation Coefficient

In the statistical analysis of the experimental data, the mean values and standard deviations for the quantities of interest were calculated from the following general definitions:

$$\textit{mean value of } x = \langle x \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i \quad ; \quad i=1,N$$

$$\textit{standard deviation of } x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$$

The correlations between various quantities derived from the experiment were estimated numerically by means of the correlation coefficients according to the definition:

$$\textit{correlation coefficient for any two variables } x \textit{ and } y = \frac{\langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle}{\sqrt{\langle (x - \langle x \rangle)^2 \rangle}\sqrt{\langle (y - \langle y \rangle)^2}}$$