  Journal of Hydroinformatics | 07.4 | 2005

User influences on knowledge structuring on the Internet: literature survey and case study of a global network of water professionals

Juliette Butcher, Paul Jeffrey and Richard Stuetz

ABSTRACT

This paper presents a review of the literature on information retrieval and knowledge structuring on the Internet and how end-user factors influence these processes. The review demonstrated that variations in users' professional and educational backgrounds are likely to have an important influence on Internet-based knowledge acquisition and structuring. A good knowledge of users is therefore necessary for the successful design of an on-line information service, particularly for multidisciplinary domains such as the water sector which cover a wide range of subject disciplines, and scientific, technological and industrial domains. The implications of the variety of individuals involved in the water sector for structuring information/knowledge on the Internet is demonstrated by analysing the International Water Association's membership database as a function of several variables, including title, job function, organisation type, nationality and fields of interest. The database revealed a wide range of cultural and professional backgrounds, as well as diverse interests between different organisations. As well as having a good understanding of their backgrounds and interests, analysis of how the different users' use the Internet, for example, by looking at web server log files, is recommended for the successful design of web based information/knowledge sharing structures for the water sector.

Key words | Information search process, Internet, knowledge structuring, multi-disciplinary, user backgrounds, water sector

Juliette Butcher (corresponding author)
Paul Jeffrey
School of Water Sciences,
Cranfield University,
Cranfield, BEDS MK43 0AL,
UK
Tel: +44 1234 750111;
E-mail: *j.butcher@cranfield.ac.uk*

Richard Stuetz
Centre for Water and Waste Technology,
School of Civil and Environmental Engineering,
University of New South Wales,
Sydney, NSW 2052,
Australia

INTRODUCTION

An important advantage of the Internet is that it can help tackle our information challenges by making information more widely known and used by facilitating discourse on topics of information needs as well as by forming links between information providers and users. The Internet is becoming "a major repository of human knowledge" (Lesk 1997) as the amount of knowledge that is generated, the speed at which it is distributed and user access to the Internet all continue to increase. Information, however, needs to be well structured to enable users to understand, manipulate and deploy it to effective use. Information retrieval methods and knowledge structuring are known to be influenced by a range of variables. These include age, gender, ethnic or cultural background, education and language. The user's subject knowledge, environmental context, language used and preferences all influence the success with which a message is retrieved (Rowley & Farrow 2000). The implication here is that the design and structure of an Internet web site, or specific web pages, would have to be adapted in order to cope with these user variations to make the information more accessible to end users. Rowley & Farrow (2000) point out that an important preliminary to knowledge-based systems design and creation is an understanding of the different kinds of users and the way in which users may wish to search or identify information.

It is imperative to understand the process of knowledge acquisition via the Internet particularly when designing extensive Internet systems such as an information gateway or portal which are based on selection criteria, abstracting and indexing language principles, and the structure and representation of subjects and categories included in the website. We need to understand the variability of knowledge acquisition behaviours that will influence site access, searching/browsing processes and knowledge exploitation. Various factors influence such behaviours but professional and educational background is a key feature as formal knowledge acquisition skills are generally obtained during education and training courses. Web sites that present scientific or technical information provide an interesting framework within which to explore some of these factors, especially as the information on these sites is generally well defined and accessed by individuals from both industry and academia. Communication and information is intrinsic to the practice of science; research stimulated often by new information is sustained by the continuing flow of information and, when completed, again yields new information (Palmer 1991). Scientific research is becoming more broadly based and collaborative and increasingly, information, techniques and tools are being imported and exported across disciplinary boundaries (Palmer 1996).

Little research has been carried out on the information needs and information-seeking behaviour of individuals in multidisciplinary domains such as the water sector, which cover a wide range of subject disciplines, and scientific, technological and industrial domains. Individuals involved in these domains come from a variety of educational and professional fields, and therefore there may be dissimilarities in the way they structure information/knowledge as a result of their personal learning process, and cultural and organisational differences. In this paper we will:

- present a review of what influences information retrieval and structuring on the Internet,
- characterise the IWA (International Water Association) network of water professionals,
- demonstrate the implications of the variety of individuals involved in the IWA for structuring information/knowledge on the Internet.

It is hoped that the knowledge gained from this study will subsequently assist in the development of successful Internet websites or portals for the water sector.

KNOWLEDGE AND INFORMATION STRUCTURING

It is useful to clarify the difference between knowledge and information. Knowledge resides in people, both individually and in a group. It is created in the process of interaction between people sharing their experiences and information sources (Meadows 1991). There are two basic types of knowledge, "tacit" and "explicit". Tacit knowledge is defined by Vail (1999) as that which resides in people as mental models, experience and skills, and is difficult to communicate externally. Explicit knowledge is defined as that which can be communicated externally and captured in formal models, rules and procedures. Explicit knowledge is mostly stored in printed and electronic archives of societies and organisations, and is generally more accessible than tacit knowledge, depending on the method/type of storage and communication (Rowley & Farrow 2000). All knowledge is a product of the society and cultural environment in which it is created. Information is structured data that can be stored, in print or in a computer, independently of human beings. It can be disseminated by computer systems to help people who use that information and apply it to the situation in which it is required. Information is data ordered in a meaningful fashion and knowledge refers to the next stage in which the information is absorbed and understood by new users (Meadows 1991). Therefore knowledge and knowledge creation start with a human being, and information is a description of knowledge which can be transmitted and stored by non-human means.

The Internet has turned into an immense information space with diverse and often poorly organised content as a result of the accelerated growth of the World Wide Web (Holscher & Strube 2000). Owing to the poor structure of knowledge on the Internet, users face a chaotic information environment in which they have to find a range of appropriate sources as well as locate bits of information within these sources (Rowley & Farrow 2000). Koniger & Janowitz (1995) express that information is only valuable to the extent that it is structured. If there is a lack of structure in the creation, dissemination and reception of information, the information may not arrive where it is needed and, therefore, becomes useless. A lack of structure makes it more difficult for users to draw the knowledge they want from the information they receive. The expectancy structure of the user and the actual information structure need to be coherent. Gaines & Shaw (1994) state that the knowledge processes of scientific communities can be supported, via electronic communication, more directly through overtly represented knowledge structures, and that this will result in the systematic acceleration of scientific research.

Koniger & Janowitz (1995) express that the classical methods of information handling are not sufficient for the growing amount and new forms of information and, therefore, that in today's information society the individual needs a more comprehensive system of information management. They recommend the application of four universal structuring dimensions – selection, time, hierarchy and sequence – to information regardless of the information carrier (see Koniger & Janowitz (1995) for more information on the dimensions and how they can be applied to information). They also discussed the importance of physical features, labels and stylistic features as additional symbols and organising methods for structuring the new forms of information that have emerged. They also believe that user competence is important in information management and that the personality of the information-processing individual is crucial to implementing the correct, formal structures. Rowley & Farrow (2000) state that knowledge structuring needs to be able to cater for both directed searching and browsing processes. Also most users need to use multiple information sources and knowledge structures to locate information and documents, so they have to negotiate a complex maze of different subject terms and subject relationships, as different sources are structured using a different controlled language (Rowley & Farrow 2000).

Bibliographic classification systems are used to systematise knowledge for information storage and retrieval purposes in libraries and bibliographies, both manual and online (Beghtol 1998). Classification schemes have a role in aiding information retrieval in a network environment, especially for providing browsing structures for subject-based information gateways on the Internet (Koch & Day 1997). They vary in scope and methodology, but can be divided into universal, national general, subject-specific and home-grown schemes (Table 1). The type of scheme used depends on the size and scope of the service being designed. The major classification systems that predominated during the twentieth century were originally predicated on the academic disciplines, but as Beghtol (1998) pointed out, this structural principle is no longer adequate because multidisciplinary knowledge production has overtaken more traditional disciplinary perspectives and created various discourse communities whose documents cannot be accommodated in a disciplinary structure. There appears to be no internationally recognised classification scheme for water related information. Some classification codes have, however, been developed for two large water related publication databases, Aqualine and Water Resources Abstracts both based on the Cambridge Scientific Abstracts Internet Database Service (http://www.csa.com).

In her paper, Beghtol (1998) reports some attempts to revise the existing classification systems to accommodate multidisciplinary works more appropriately. This is important as the various collaborative scientific communities which conduct interdisciplinary research are moving their knowledge acquisition processes to the Internet (Chen & Gaines 1996). The interdisciplinary nature of modern scientific research, where researchers in departments publish in journals across a range of fields outside their normal disciplinary affiliation, is an acknowledged "norm" in the university research community (Bourke & Butler, 1998). Tomov & Mutafov (1996) state that knowledge from different scientific fields and technological know-how contribute, either directly or indirectly, to the research in publications. In their paper, they demonstrate the use of bibliometric methods/indicators to assess the structure and dynamics of interdisciplinary links between different fields and subfields.

INFORMATION RETRIEVAL

The rapid increase in the number of information sources is causing an exponential growth in the quantity of information on the Internet, making it increasingly difficult for users to find relevant information. Information retrieval is recognised to be a major problem on the Internet, with the extensive range of databases in various formats and many types of search and retrieval software set up on different computers providing access via a variety of interfaces to the databases. Rowley & Farrow (2000) point out that

Table 1 | Classification schemes (information from Koch & Day, 1997)

	Universal schemes	National general schemes	Subject-specific schemes	Home grown schemes
Scope	Universal subject coverage designed for international use	Universal subject coverage designed for use in single country	Single subject coverage designed for use by particular subject community	Devised for use in a particular service
Examples	DDC (Dewey Decimel Classification), UDC (Universal Decimel Classification)	BC (Nederlandse Basisclassificatie), SAB (Sveriges Allmam Biblioteksforening)	Ei (Engineering Information), MSC (Mathematics Subject Classification)	Yahoo
Advantages	-Available in machine-readable form	-Well known within country	-Particular user group	-Relatively flexible
	-Widely supported	-Multilingual not necessary	-More up to date than universal	-Easy to change
	-Known to more users		-Some well known subject schemes	-Can quickly absorb new topics
	-Permits multilingual access		-Controlled structure and terminology	
Dis-advantages	-False ontology	-Not well known externally	-Interoperability more difficult	-May lead to lack of consistency
	-Poor assimilation of new topics	-External support limited	-May be difficult to learn structure	-Less well known than universal
		-Geographic bias	-Fringe topics not well covered	-May need frequent revision
			-Only best if well-established	-Lack of cooperation

information retrieval requires considerable skill if the best information or document for the purpose is to be extracted. Despite the enhanced availability of information on the Internet, the applicability of it remains far behind the potentials and information overload is one feature of this problem (Lechner *et al.* 1999). Excessive information impedes its assimilation and therefore does little to improve knowledge and understanding (Dodge 2000). There is now an increasing awareness of the problem of "information overload" (e.g. Berghel 1997; Edmunds & Morris 2000). This has stimulated research and development in information retrieval and dissemination on the Internet. A great deal of effort is being directed by researchers in a range of disciplines to cope with the problem of information retrieval and information overload, through filtering, structuring, analysing and visualising information to aid the limited human capacity to search for, absorb and comprehend information (Dodge 2000).

Research into improving technology for information searching and retrieval via the Internet is now increasingly abundant. Many recent publications have focused on the effectiveness of search engines, information validation and quality, the design of interfaces, data structures and metadata, classification, and indexing and search agents (Chowdhury 1999). Successful information retrieval depends on the effectiveness of indexing[1] and searching processes (Rowley & Farrow 2000). A variety of indexing and search tools have been developed which help different users locate information among the vast array of databases, other information resources and services available on the Internet. Some tools rely on human involvement for the selection, evaluation and indexing processes, others totally, or almost entirely, depend on automatic search and retrieval techniques. Nowadays search engines[2] (e.g. Google, Yahoo!, Lycos) are a central part of information seeking on the Internet; they are becoming the main tools for indexing and retrieving Internet resources. The effectiveness of the various web search tools is discussed in detail by Lawrence & Lee Giles (1999). Web search engines are limited in terms of coverage, recency, how well they rank query results and the query options they support. The development and application of most services has been very rapid with little central planning. There is a lot of information that traditional search engines cannot or do not index, for example, content that resides in searchable databases. This problem is discussed in detail by Pedley (2001) who states that search engines may be able to index homepages of databases but are unable to index individual records within a database: therefore an enormous amount of valuable content/information on the web is "invisible".

Allen (1970) claims that failure in information retrieval of scientific and technical information is a result of the nature and complexity of the information itself, uncertainty and the user's individual nature of information needs. There have been some recent studies focusing on improving the access to, and the dissemination and retrieval of, scientific information on the Internet. Lechner *et al.* (1999) studied the use of an encyclopaedic concept as a knowledge medium for scientific communities on the Internet. The encyclopaedia can be reconstructed into a new interactive carrier of information to structure and organise scientific knowledge on the Internet. Lawrence & Lee Giles (1999) discuss the use of autonomous citation indexing in digital libraries, for example Citeseer (http://citeseer.nj.nec.com/cs). There has also been some work on the development of intelligent tools that automate the structuring, maintenance, search and retrieval tasks by generating profiles of user interests and web documents with minimal developer/user interaction. Kapur & de Vel (1997) demonstrate a system called VWeb (Virtual Web) which draws upon the nature of information, the structure of the human brain and natural language. Gaines *et al.* (1997) point out that it is important to model and support the processes by which knowledge is acquired and disseminated through the Internet. In order to improve the knowledge processes of geographically dispersed communities collaborating via the Internet, they developed a number of conceptual models to characterise the existing types of Internet services in terms of their utility for computer-mediated communication, access to services or search processes (Gaines & Shaw 1994; Chen & Gaines 1996; Gaines *et al.* 1997).

[1]Indexing is described by Rowley & Farrow (2000) as the process whereby structure can be added to knowledge in order to support more effective and more efficient retrieval.

[2]A *search engine* is a retrieval mechanism that performs the basic retrieval task, the acceptance of a query, a comparison of the query with each of the records in a database and the production of a retrieval set as output (Rowley & Farrow 2000).

INFORMATION SEARCH PROCESS

Understanding users' information seeking behaviours has become a major concern in information science because this knowledge is essential in order to serve users better and to design suitable information systems (Kim 1998). The design of web sites is currently based on comprehensive investigations about the interests of web site users and on less studied assumptions about their exact behaviour (Spiliopoulou *et al.* 1999). When looking for information on the web, people make use of a variety of techniques. They also behave differently towards information in diverse situations. They may wish to retrieve a document or piece of information for different reasons. Information seeking and processing is not the result of a simple, linear process where a discrete need for information arises. First, a method of retrieval is decided on to access the information, and then the information is acquired (Hess 1999). An information search is a process of construction, which involves the whole experience of the person, feelings, as well as thoughts and actions. It begins with the user's problem. Information need is the gap between the user's knowledge about the problem or topic and what the user needs to know in order to solve the problem. The user's state of knowledge is dynamic, not static, and changes during the search process (Hess 1999).

Rowley & Farrow (2000) defined two basic search processes: "directed searching" and "browsing". Directed searching is performed by users when they know what they are looking for. Some characteristic of the information or document (e.g. author) is usually used as the basis of a specific search. Browsing is performed when users have a less precise view of the information or documents that may be available and are not sure whether their requirements can be met or how they might be met. They also defined two other types of search processes: "current awareness search", which is performed when a well-informed user is concerned to locate the latest information on their topic of interest, and "retrospective search", which is performed when a user searches for information to update their knowledge on a specific subject. They point out that success in searching will not be achieved solely through the identification of an appropriate source, but also depends on skills in extracting the information from that source.

The first significant investigations on the behaviour of users when navigating the web mainly investigated aspects of *browsing* (e.g. Catledge & Pitkow 1995). Since the development of web-based services and the migration of "traditional" on-line search services to the web, there is a growing body of research exploring information-seeking processes over time. Results from these studies show that humans progress through a series of stages, adopt different strategies and exhibit different information behaviours at different stages of their information-seeking process (Ellis 1989; Kuhlthau 1993). Ellis (1989) developed a general model of information seeking behaviour to improve the design of information retrieval systems. He used this model to study the information seeking patterns of social scientists, research physicists and chemists (Ellis *et al.* 1993), as well as engineers and research scientists in an industrial firm (Ellis & Haugan 1997). In his model, he defines the following different characteristics of information seeking behaviour: Starting, Chaining, Browsing, Differentiating, Monitoring, Extracting, Verifying, and Ending. Kuhlthau (1993) developed an Information Search Process Model after studying the information seeking process of library patrons. She found that the process occurred in six clearly defined stages related to the cognitive, affective states and search activities of the users, including task initiation, topic selection, pre-focus exploration, focus formulation, information collection and search closure.

More recently, Choo *et al.* (1999) investigated the information-seeking behaviour of knowledge workers over a two-week period. A number of web user behaviours were characterised by combining the results of surveys, interviews and client-side logging. They developed a model combining both behavioural modes (strategies & reasons for viewing/searching) and moves (tactics used to find/use information based on Ellis' (1989) work). Table 2 shows the four distinct modes of scanning described by Choo *et al.* (1999) based on different information needs, information seeking processes and the purpose of information use.

In his paper Wilson (1999) discusses the various models of information seeking behaviour and search processes in the literature. Some researchers (e.g. Jansen *et al.* 1998) have studied immense data sets derived from the logs of Internet search engines (e.g. Excite) that provide a comprehensive view of how average web users approach

Table 2 | Modes of scanning (from Choo *et al.* 1999)

Scanning modes	Information need	Information seeking	Information use
Undirected viewing	General areas of interest; specific need to be revealed	"*Sweeping*" Scan broadly a diversity of sources, taking advantage of what's easily accessible	"*Browsing*" Serendipitous discovery
Conditioned viewing	Able to recognize topics of interest	"*Discriminating*" Browse in pre-selected sources on pre-specified topics of interest	"*Learning*" Increase knowledge about topics of interest
Informal search	Able to formulate simple queries	"*Satisfying*" Search is focused on area or topic, but a good-enough search is satisfactory	"*Selecting*" Increase knowledge on area within narrow boundaries
Formal search	Able to specify targets in detail	"*Optimizing*" Systematic gathering of information about an entity, following some method or procedure	"*Retrieving*" Formal use of information for decision- and policy-making

a search service. An important disadvantage of these studies is that, because the data is anonymous, there is no information on the context of the "individual" user (e.g. experience, information need, etc.). The analysis of user behaviour has two main aspects, one relating to the interests of the users and the information they access and the other concerning the *way* of accessing this information (Spiliopoulou *et al.* 1999). The first aspect may be resolved by methods used to establish user profiles and the second one may be determined by means of analysing web server logs. Spiliopoulou *et al.* (1999) believe that these two aspects are complementary since a web user is characterised by their interests *and* by their navigational behaviour. Lancieri (1999) states that a user's behavioural profile is related to the way that the user acquires knowledge when he "surfs" on the web, for example, a young student and an experienced university researcher "surf" the web differently because the student's need is not the same as that of the researcher, even if they have the same interest. Lancieri demonstrates that by studying the user's web access, it is possible to obtain knowledge relating to his behaviour and his field of interest. In his research, he analysed proxy caches[3] to monitor (or "follow") users' activities on the Internet as their contents reflect the interests, habits and needs (i.e. profiles) of the users. This method helps characterize, for example, experienced users from beginners and shows what type of information users regularly need, for example, different/superficial or more detailed/specific information.

Data mining on the navigational behaviour of web users has gained considerable importance in recent years as the significance of the web for information dissemination has been recognised and understanding users' navigation preferences became necessary for improving the design and quality of electronic services (Spiliopoulou *et al.* 1999). When web users interact with a site, data recording their behaviour is stored in web server logs and, as the log data is collected in a raw format, it can be analysed by automated tools (Borges & Levene, 2000). At present there are quite a few commercial log analysis tools available (e.g. Net Tracker Pro, Funnel-Web). Some of these tools, however, have limited analysis

[3] All the web sites that a user accesses go to the proxy and all the information downloaded is copied in the cache memory.

capabilities producing only simple results such as summary statistics and frequency counts of page visits. Some applications are also rather slow, inflexible and difficult to maintain. Meanwhile researchers have been studying data mining techniques to take full advantage of information available in the log files. Several research communities, such as those involved in databases, information retrieval, artificial intelligence, psychology and statistics, are involved in web mining research (Wang 2000).

To date, two main approaches have been used in mining for user navigation patterns from log data. The first approach is the mapping of log data onto relational tables and an adapted version of standard data mining techniques, such as mining association rules, statistical methods or clustering analysis, are used (Borges & Levene 2000; Wang 2000). The second approach involves developing techniques that can be used directly on the log data (Spiliopoulou *et al.* 1999). Spiliopulou *et al*. used a Web Utilisation Miner tool to provide knowledge on the way users navigate in a web site by discovering access patterns with "interesting" statistical properties. Borges & Levene (2000) propose a data mining model that captures user navigation behaviour patterns by using the N-grammar approach, which has potential use in helping web site designers understand the preferences of their sites' visitors and individual users to understand better their own navigation patterns, as well as increase their knowledge of the web's content. One problem with using data mining techniques is that sufficient data is needed; otherwise it will be very difficult to identify the users. User cooperation is needed to improve data quality but this is not easy as a result of the privacy needs of individual users (Wang 2000). Borges & Levene (2000) discuss the use of heuristics to improve the quality of mining for user navigational patterns.

DETERMINANTS OF VARIATION IN INFORMATION RETRIEVAL BEHAVIOUR

Relatively few studies have been carried out on the influence of individual characteristics on the information search process compared to research that focus on factors such as search tactics, technologies for searching the web, performance of information systems, etc. As Internet use continues to increase, it is now widely recognised that users' personal differences influence their search/navigational behaviour on the web, and an increasing number of studies are being carried out to understand end-users with diverse backgrounds, skills and needs. The user variables that may influence information retrieval include gender, age, personal experience, state/level of present knowledge, cognitive styles, culture, etc. Table 3 shows some examples of previous studies carried out on the influence of some user variables on information-seeking processes.

Several recent studies have shown that previous experience has an important influence on a user's success in information retrieval. Holscher & Strube (2000) investigated the types of knowledge structures or strategies that are relevant for web-based information seeking and that help improve searching on the web, based on a better understanding of user characteristics. They looked at the effects of web experience and domain-specific background knowledge on a series of search tasks in an economics related domain, and their results showed that successful search performance requires the combination of the two types of expertise. Research by Hseih-Yee (1993) also demonstrates that technical expertise in using bibliographic database systems needs to be combined with background knowledge about the subject area to be searched for successful information retrieval. Bower & Hilgard (1981) state that prior knowledge increases the ability to acquire more knowledge and the ability to recall and use this knowledge. This notion is supported by Trott (1993) who studied the inward technology transfer processes in an organisation and argues that the ability to evaluate and utilise external knowledge is related to prior knowledge and expertise. O'Reilly (1982) states that decision-makers who have more tenure, education and motivation tend to seek and use more updated information, but it is not clear how he collected the data for his research. According to Allen (1977), more experienced individuals tend to be more aware of specialised information sources than those who are less experienced. Gaines *et al.* (1997) created a "human factors" framework for analysing the utility, usability and likeability of Internet services, and to analyse it in greater depth they used a layered protocol model for human–computer interaction. This model is shown in Figure 1 (note the term "agent" is used to cover both human users and computer services). It shows that both the user's knowledge and skills, related to their background knowledge gained

Table 3 | Previous studies on the influence of user variables on information search processes

User variable	Notes	Author(s)
Gender		Qiu (1993); Bayer & Jahoda (1979)
Age		Bayer & Jahoda (1979)
Cognitive style	Represents the manner in which an individual receives, processes and uses information (Palmer 1991).	Kim (2000); Hess (1999); Navarro-Prieto *et al.* (1999); Korthauer & Koubek (1994); Allen (1991); Palmer (1991).
Experience/ Expertise	E.g. Computer experience, online search expertise, subject knowledge	Holscher & Strube (2000); Kim (2000); Hseih-Yee (1993); Marchionini *et al.* (1990); O'Reilly (1982); Fenichel (1981)
Academic background		Borgman (1986)
Culture		Iivonen & Durnite (2001)

through experience/training and to their capability to translate intentions into actions using their background knowledge, both influence usability.

There have been a few studies on the information seeking behaviours of scientists and engineers. Research by Allen (1977) into the acquisition of information within technological environments showed that scientists and engineers differ in sourcing techniques. Scientists tended to get information from literature whereas engineers were inclined to get information from personal contacts. Kim (1998) identified three important types of factors that influence an engineer's information seeking behaviour: organisational factors, work task factors and personal characteristic factors. He reports that, based on a survey of existing literature on engineers' information seeking behaviour, engineers' perception on accessibility and quality of information influence the way in which they search information. Pinelli (1991) found that, when engineers were confronted with a subject unfamiliar to them, they tended first to look for colleagues or other contacts within their personal network that possessed the knowledge. Also they chose their information channels based on their own experience and knowledge, through consultation of

Figure 1 | Layered protocol model for human-computer interaction (from Gaines *et al.* 1997).

personal contacts, or both methods. In her research based on a psychometric test of the information seeking behaviour of a group of agricultural research scientists, Palmer (1991) found that discipline or subject area had a more important influence than individual differences. She concludes that no population of information users, even though they work together and share common goals, can be regarded as an undifferentiated whole and that different sectors of the population require different treatment depending on their discipline, the length of time they have spent in the subject field and organisation, their task role and their cognitive style.

CHARACTERISATION OF INTERNATIONAL WATER ASSOCIATION (IWA) MEMBERSHIP

As a part of ongoing research into research collaboration in the water sector, we analysed the IWA's "individual" membership database which contained data on 4516 members. The IWA is a global network of water professionals "spanning the continuum between research and practice and covering all facets of the water cycle" (quoted in www.iwahq.org.uk). The main objective of this analysis is to describe the membership as a function of a number of variables, including title, job function, organisation type, nationality and field of interest (specialist group). Our aim is to characterise the variety or diversity of characteristics exhibited by the individual members. We applied simple frequency and cross-tabulation techniques to the data set using SPSS (Statistical Package for Social Sciences) to describe the major characteristics of the IWA membership.

Looking at the percentage of membership by title, which roughly indicates the professional qualifications the members have, the majority (over half) of the members do not show a qualification in their titles (e.g Mr/Mrs; see Figure 2). Approximately 35% of the members have PhDs (Doctor of Philosophy). We also received some information on the members' age, classified into three different age groups: under 35, 36–50 and 51 plus. Their age roughly indicates the number of years of experience they have (level of knowledge). Of the 1417 members who indicated their age, the majority (48.6%) are between 36–50 years old. A small proportion of members (16.5%) are under 35 years old and 34.9% are over 50 years old.

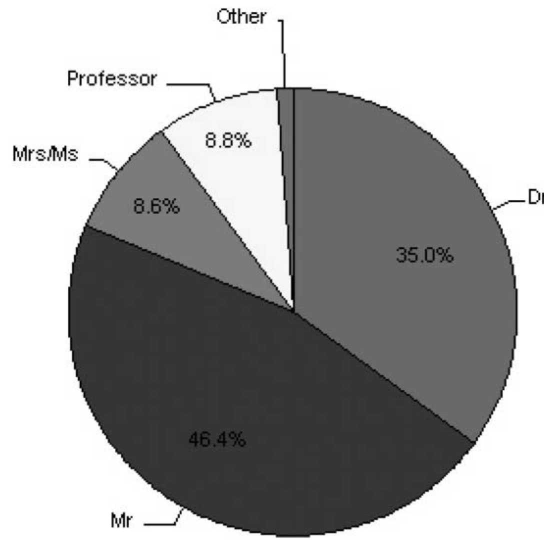

Figure 2 | Percentage of IWA membership by title (of 4068 members).

Figure 3 shows the variety of work roles of IWA members and Figure 4 illustrates the diversity of institutions that they work in. Although we can clearly see that just over half of the members are researchers or scientists and that nearly half of the members work at a research institute or organisation, just over half of the IWA members work in an "industrial" environment, the majority of these in consultancy. Further analysis of the data, using cross-tabulations, indicates that the majority of the members who work in the "industrial" organisations (regulator, equipment manufacturer, industry and utility) describe themselves as "managers".

We also used cross-tabulations to look at the qualifications of the members within the different types of

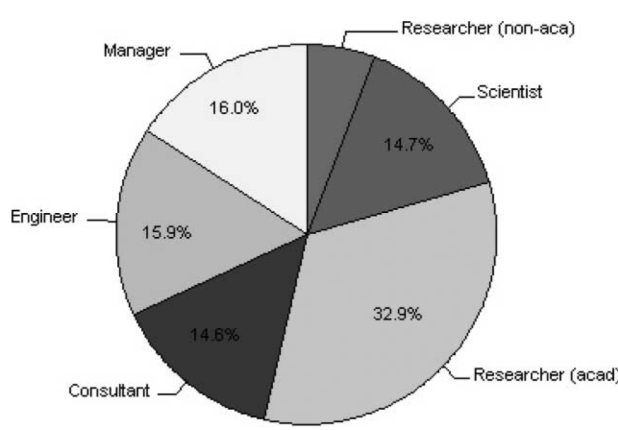

Figure 3 | Percentage of IWA membership by primary work role (of 3007 members).

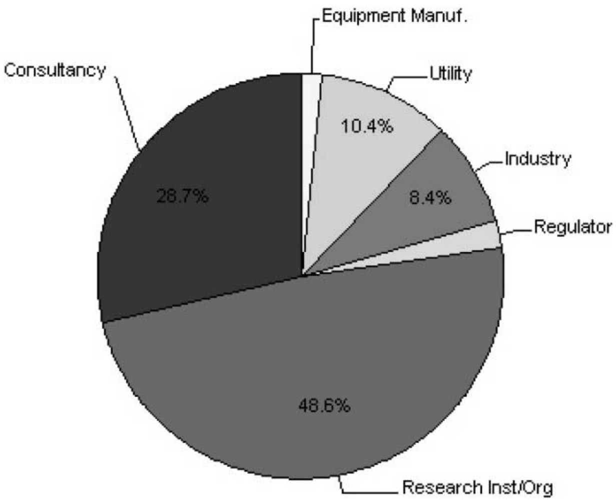

Figure 4 | Percentage of IWA membership by organisation type (of 2671 members).

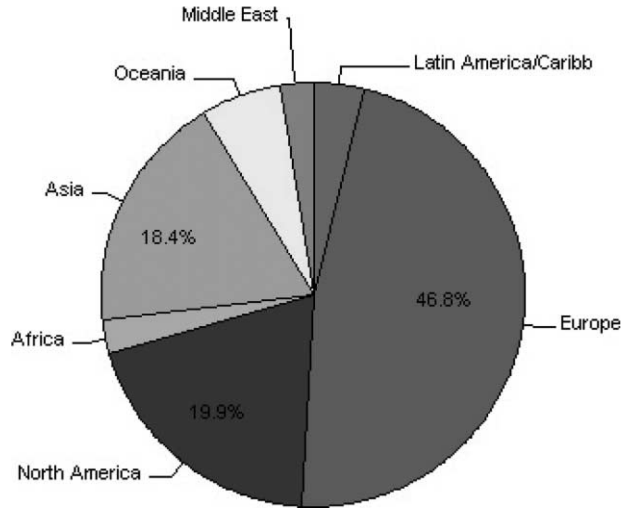

Figure 5 | Percentage of IWA membership by continental region.

organisation. Just over 45% of the members who work in a research institute/organisation have a qualification in their title (e.g. Dr/Eng.), whereas on average a quarter of the members who work in the "industrial" organisations have a qualification in their title. By analysing the titles for each work role, we found that half of the scientists/academic researchers have a qualification in their title, whereas less than 40% of the non-academic researchers and just 21% of the engineers have a qualification in their title.

Looking at the nationalities of the members, the IWA membership represents 118 different countries spanning all the continental regions, of which seven countries (USA, UK, Japan, Germany, Australia, Italy and Canada) account for over 50% of members. Figure 5 illustrates the percentage of IWA membership by continental region.

We also explored variations in the percentage of membership by title, age, work role and organisation type between the different continental regions. The percentage of members which show a qualification in their title was lowest in Africa (24.8%) and highest in the Middle East (48.5%). The age group cross-tabulation revealed that in all the continental regions the majority of members are in the "36–50" age group, except for North America which has more members in the "51 plus" age group. Compared to the other continental regions, Europe has the highest proportion of members in the "under 35" age group. By analysing the members' work roles in the different continental regions, we found that Oceania has a low proportion (36.1%) of researchers/scientists. On the other hand, the Asian membership has a very high proportion (71%) of researchers/scientists. With regard to organisation type, in Europe, North America, Africa and Oceania more than half of the members work in an "industrial" environment, whereas in Asia, Latin America and the Middle East the majority of the members work at a research institute/organisation.

The IWA membership database also included information on the Specialist Groups which members choose to be involved in. Specialist Groups represent the core vehicle for issue-based interaction on scientific, technical and management topics, and every IWA member can join up to 5 out of the 46 active groups which are classified into five main fields: i) Drinking Water, Wastewater and Solid Waste Treatment Processes, ii) Education, Health, Management and Training, iii) Network Management, iv) Source Management and v) Water Quality Management. Of 4516 members in the database, 3807 are members of one or more specialist group. This data provides useful information on the members' subject interests and it is useful to look at the variations in the interests of members from different organisations and continental regions. By calculating the number of members involved in each specialist group, we were able to see which groups were the most popular. The three most popular groups which involve more than 20% of the 3807 members were (in order): i) "Wastewater Reclamation, Recycling and Reuse", ii) "Nutrient Removal and Recovery" and iii) "Small Water and Wastewater

Systems". All these three groups are within the "Drinking Water, Wastewater and Solid Waste Treatment Processes" field. Only one group from the four other fields shows significant interest by the members (20% of members) and that is the "Watershed and River Basin Management" group which is in the "Source Management" field.

We also looked at the percentage of members from each of the seven continental regions involved in each of the 46 specialist groups. We found some interesting patterns that reflect the circumstances of the different continental regions. The most popular specialist group in Europe, Asia and North America was the "Nutrient Removal and Recovery" group. In Oceania and the Middle East, "Wastewater Reclamation, Recycling and Reuse" was the most popular group and in Latin America/Caribbean and Africa, "Water/Waste Technology and Management Strategies for Developing Countries" was the most popular group. Table 4 shows the specialists groups that show interest from more than 25% of the members in any one continental region.

Looking at the other groups in the table, Latin America and Africa show high interest in "Anaerobic Digestion", Latin America and the Middle East show high interest in "Small Water and Wastewater Systems" and only Africa shows high relative interest in the "Source Management" field group "Watershed and River Basin Management".

We also analysed the variations in interests of the members from the six different organisation types. Table 5 shows the specialists groups that show interest from more than 25% of the members in any one type of organisation.

We can see a greater diversity of interests between the different organisation types. The regulators only show particular interests in the management field groups, in particular the Source Management field, "Watershed and River Basin Management". Those members working in Utilities show most interest in the "Design, Operation and Costs of Large Wastewater Treatment Plants" group and those from consultancy show most interest in the "Small Water and Wastewater Systems" group. The members from both industry and equipment manufacturers have a high interest in "Membrane Technology". Both the research organisations and equipment manufacturers show interest in "Wastewater Reclamation, Recycling and Reuse".

Even within a specialist group, there is a diversity of backgrounds and, to give an example, Figures 6 and 7 show the work roles and organisation types for members involved in the "Membrane Technology" specialist group. The patterns are quite similar to those of the whole IWA membership (Figures 3 and 4), except that there are fewer researchers/scientists (47.2%) and more engineers (22.7%) within the membrane group and a higher proportion work in an "industrial" environment (~60%) with one third of the members from consultancy.

The analysis of the IWA membership database revealed a diversity of educational (in terms of qualification and age) and professional backgrounds, with members working in a variety of roles within a number of different organisations, both industrial and research. There is also a mixture of nationalities and interests (specialist group subject fields). Although we have found that different organisations and continental regions have particular interest in different specialist groups, in general, they all also show some interest in many other specialist groups. The variety in the IWA's membership is, however, no real surprise as it has a policy of increasing the diversity of its membership as the following statement from its website (http://www.iwahq.org.uk) demonstrates:

> "The ultimate strength of IWA lies in the professional and geographic diversity of its membership that includes academic researchers and research centres, utilities, consultants, regulators, industrial water users and water equipment manufacturers. IWA members from each of these communities represent the leading edge in their fields of specialty, and together are building new frontiers in global water management through interdisciplinary exchange and collaboration."

DISCUSSION AND CONCLUSIONS

The literature review revealed that much research was devoted to providing "access" to knowledge and less work was done on understanding the needs of those acquiring knowledge. Researchers are, however, now acknowledging that a solid knowledge of users is necessary for the successful design of an on-line information service, as variations in their backgrounds are likely to have an important influence on Internet based knowledge acquisition and structuring. These variations result from the education and training individuals

Table 4 | Specialist groups showing interest by more than 25% of members in any one of the seven continental regions (values are percentage of members from region)

Specialist group	Europe	Asia	North America	Latin America /Caribbean	Africa	Oceania	Middle East
Anaerobic Digestion (DWWSTP)*	19.7	19.1	18.8	36.4	27.7	14.2	14.4
Nutrient Removal and Recovery (DWWSTP)*	26.1	30.9	26.0	22.4	20.2	28.3	13.5
Wastewater Reclamation, Recycling and Reuse (DWWSTP)*	22.3	28.9	25.4	34.3	29.8	45.0	36.9
Water/Waste Technology and Management Strategies for Developing Countries (DWWSTP)*	17.0	20.5	15.5	40.6	44.7	20.8	28.8
Small Water and Wastewater Systems (DWWSTP)*	25.2	19.6	16.1	37.1	20.2	25.4	29.7
Watershed and River Basin Management (SM)*	20.3	19.5	21.0	18.2	27.7	17.5	14.4

*Field: DWWSTP = Drinking Water, Wastewater and Solid Waste Treatment Processes; SM = Source Management

have undertaken which lead to differences in, for example, their understanding of, and skills in, online search techniques; their knowledge of the subject(s) in their domain (and perhaps in other fields of interest); their level of awareness/-knowledge; and their information management skills. All these features have an effect on the way users acquire knowledge from Internet based resources and the structure of the knowledge on these resources.

The wide range of cultural (organisation and country) and professional (work roles, qualifications and subject

Table 5 | Specialist groups showing interest by more than 25% of members in any one of the six organisation types (values are percentage of members from organisation type)

Specialist group	Research inst/org	Utility	Consultancy	Industry	Regulator	Equipment Manuf.
Design, Operation, Costs of Large Wastewater Treatment Plants (DWWSTP)*	11.7	30.9	28.5	21.9	12.5	11.9
Nutrient Removal and Recovery (DWWSTP)*	24.2	25.7	31.6	30.8	16.1	23.8
Wastewater Reclamation, Recycling and Reuse (DWWSTP)*	27.5	18.1	27.4	24.4	23.2	42.9
Small Water and Wastewater Systems (DWWSTP)*	21.6	18.5	32.0	22.4	14.3	26.2
Membrane Technology (DWWSTP)*	15.4	19.3	21.6	31.3	3.6	38.1
Advanced Treatment / Membrane Technology (DWWSTP)*	17.0	18.5	17.0	31.3	8.9	42.9
Watershed and River Basin Management (SM)*	22.5	20.1	18.5	10.4	35.7	7.1
Health-Related Water Microbiology (EHMT)*	17.6	16.9	7.9	10.0	26.8	16.7
Urban Drainage (WQM)*	11.3	10.0	14.2	6.0	26.8	7.1
Diffuse Pollution (WQM)	14.1	7.6	7.7	2.0	26.8	2.4

*Field: DWWSTP = Drinking Water, Wastewater and Solid Waste Treatment Processes; SM = Source Management; EHMT = Education, Health, Management and Training; WQM = Water Quality Management

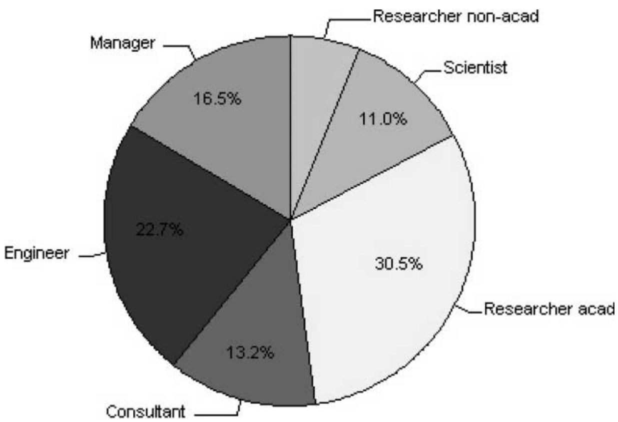

Figure 6 | Percentage of members by primary work role in Membrane Technology Specialist Group.

knowledge) backgrounds of the individuals involved in the water sector, as demonstrated by our analysis of International Water Association's membership database, clearly makes it very difficult to design successful water-related web based information/knowledge sharing structures. You would need to take into account the various backgrounds (both educational and professional), cultures, ages, languages and subject interests. It should be remembered that "water" provides a thematic point of reference for our interests, not a common way of thinking about problems and issues. There is a need to recognise and respect disciplinary, professional and specialist perspectives on our theme.

The literature review revealed that engineers and scientists, and even young students and experienced university researchers, have different information seeking behaviours and ways of structuring knowledge, due to factors related to their work task, organisation and subject knowledge. It is therefore recommended that, as well as having an understanding of their backgrounds, an investigation is made of how the different users approach water-related websites by using some of the recent methods shown in the literature review which show users' navigational patterns, interests, etc., for example, by analysing web server log files and combining the results with user profile data.

In conclusion, this study has demonstrated a variety across a number of dimensions within the IWA network and within a specialist group – dimensions which the literature states are important determinants of attitudes to knowledge/information retrieval and structuring. We propose that the design of dissemination and knowledge support frameworks, not just web sites, needs to take this variety into account and that further work which seeks to record or capture variations in knowledge needs, presentation formats, etc., would be a useful first step in the design of such frameworks.

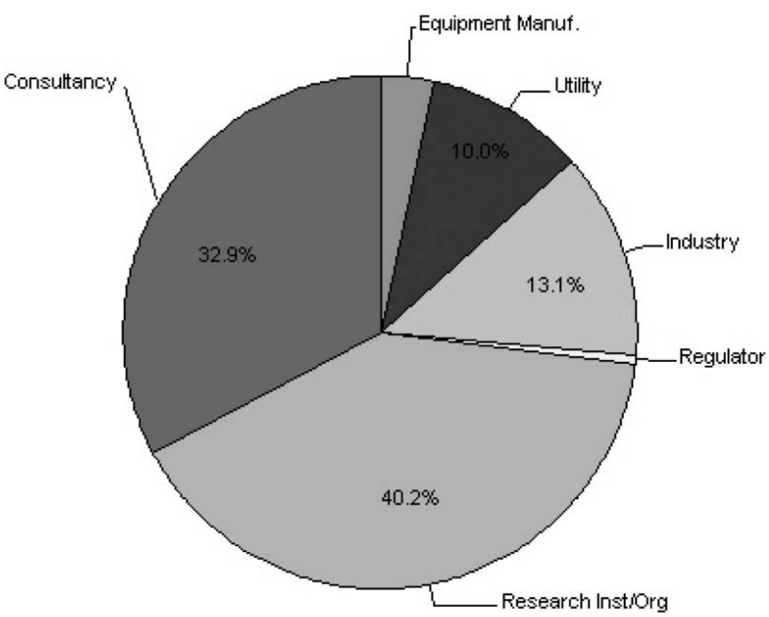

Figure 7 | Percentage of members by organisation type in Membrane Technology Specialist Group.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the IWA (International Water Association) for providing the membership data and the Engineering and Physical Sciences Research Council (EPSRC) for their financial support.

REFERENCES

Allen, B. L. 1991 Cognitive research in information science: implications for design. In *Annual Review of Information Science & Technology*, vol 26 (ed. M. E. Williams), pp. 3–37. Learned Information. Medford, NJ.

Allen, T. J. 1970 Communication networks in R & D laboratories. *R & D Mngmnt.* **1** (1), 14–21.

Allen, T. J. 1977 *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R & D Organisation*. MIT Press. Cambridge, MA.

Bayer, A. E. & Jahoda, G. 1979 Background characteristics of industrial and academic users and nonusers of online bibliographic search services. *Online Rev.* **3** (1), 95–105.

Berghel, H. 2000 Cyberspace 2000: dealing with information overload. *Commun. ACM.* **40** (2), 19–24.

Beghtol, C. 1998 Knowledge domains: multidisciplinarity and bibliographic classification systems. *Knowledge Org.* **25** (1/2), 1–12.

Borges, J. & Levene, M. 2000 Data mining of user navigation patterns. In *Web Usage Analysis and User Profiling* (eds B. Masand & M. Spiliopoulou), Springer, Berlin, pp. 92–111.

Borgman, C. 1986 The user's mental model of an information retrieval system: an experiment on a prototype online catalog. *Int. J. Man-Machine Stud.* **24**, 47–64.

Bourke, P. & Butler, L. 1998 Institutions and the map of science: matching university departments and field of research. *Res. Policy.* **26**, 711–718.

Bower, G. H. & Hilgard, E. R. 1981 *Theories of Learning*. Prentice-Hall. Englewood Cliffs, NJ.

Catledge, L. D. & Pitkow, J. E. 1995 Characterizing browsing strategies in the World Wide Web. *Comput. Net. ISDN Syst.* **27**, 1065–1073.

Chen, L.L.-J., Gaines, B.R. 1996 Knowledge Acquisition Processes in Internet Communities. (eds Gaines, B.R. and Musen, M.A.), Proceedings of Tenth Knowledge Acquisition Workshop, Banff, Canada, November, pp. 43:1–43:18 available at http://ksi.cpsc.ucalgary.ca/KAW/KAW96/chen/ka-chen-gaines.html

Choo, C. W., Detlor, B. & Turnbull, D. 1999 Information seeking on the Web - an integrated model of browsing and searching. In *Proceedings of the 62nd Annual Meeting of the American Society of Information Science (ASIS), 'Knowledge: creation, organisation and use' Washington, D.C., 1999* (ed. L. Woods), Information Today, Medford, NJ pp. 187–199.

Chowdhury, G. G. 1999 The Internet and information retrieval research: a brief review. *J. Doc.* **55** (2), 209–225.

Dodge, M. 2000 Accessibility to information within the Internet: how can it be measured and mapped? *Information, Place, and Cyberspace* (eds D. G. Janelle & D. C. Hodge). Springer Verlag, Berlin, pp. 187–204.

Edmunds, A. & Morris, A. 2000 The problem of information overload in business organisations: a review of the literature. *Int. J. Inf. Mngmnt.* **20**, 17–28.

Ellis, D. 1989 A behavioural model for information retrieval system design. *J. Inf. Sci.* **15** (4/5), 237–247.

Ellis, D., Cox, D. & Hall, K. 1993 A comparison of the information seeking patterns of researchers in the physical and social sciences. *J. Doc.* **49** (4), 356–369.

Ellis, D. & Haugan, M. 1997 Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *J. Doc.* **53** (4), 384–403.

Fenichel, C. H. 1981 Online searching measures that discriminate among users with different types of experiences. *J. Am. Soc. Inf. Sci.* **32**, 23–32.

Gaines, B. R., Chen, L. L.-J. & Shaw, M. L. G. 1997 Modelling the human factors of scholarly communities supported through the Internet and World Wide Web. *J. Am. Soc. Inf. Sci.* **48** (11), 987–1003.

Gaines, B. R. & Shaw, M. L. G. 1994 Using knowledge acquisition and representation tools to support scientific communities. AAAI'94: *Proceedings of the Twelfth National Conference on Artificial Intelligence*, AAAI Press/MIT Press, pp. 707–714. Menlo Park, California.

Hess, B. 1999 Graduate student cognition during information retrieval using the World Wide Web: a pilot study. *Comput. Educ.* **33**, 1–13.

Holscher, C. & Strube, G. 2000 Web search behaviour of Internet experts and newbies. *Comput. Net.* **33**, 337–346.

Hseih-Yee, I. 1993 Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *J. Am. Soc. Inf. Sci.* **44** (3), 161–174.

Iivonen, M. & Durnite, M. D. 2001 The choice of initial web search strategies: a comparison between Finnish and American searches. *J. Doc.* **57** (4), 465–491.

Jansen, B. J., Spink, A., Bateman, J. & Saracevic, T. 1998 Real life information retrieval: a study of user queries on the Web. *SIGIR Forum* **32** (1), 5–17.

Kapur, S. & de Vel, O. 1997 Knowledge structuring, navigation and retrieval in a virtual World Wide Web. In *Proc. International Conference on Computational Intelligence and Multimedia Applications*. Watson Ferguson & Company, Australia, pp. 177–183.

Kim, K. S. 2000 Individual differences and information retrieval: implications on Web design. In *Proceedings of the 6th Conference on Content-Based Multimedia Information Access RIAO 2000, Paris, France, 2000* (eds J. Mariani & D. Harman), CID, Paris, pp. 885–893.

Kim, S.-Y. 1998 Measuring the impact of information on work performance of collaborative engineering teams. In

Collaboration Across Boundaries: Theories, Strategies, and Technology: Proceedings of the 1998 Asis Midyear Meeting, Orlando, Fl. May 16-20, 1998 (ed. K. Lieberman, D. H. Sonnenwald & B. M. Wildemuth), Information Today, Medford, NJ, pp. 54–63.
Koch, T. & Day, M. 1997 The role of classification schemes in Internet resource description and discovery. *EU Project DESIRE.* Deliverable D3.2.3.
Koniger, P. & Janowitz, K. 1995 Drowning in information, but thirsty for knowledge. *Int. J. Inf. Mngmnt.* **15** (1), 5–16.
Korthauer, R. D. & Koubek, R. J. 1994 An empirical evaluation of knowledge, cognitive style, and structure upon the performance of hypertext task. *Int. J. Human-Computer Interaction* **6** (4), 373–390.
Kuhlthau, C. C. 1993 *Seeking Meaning: A Process Approach to Library and Information Services*. Ablex Publishing, Norwood, NJ.
Lancieri, L. 1999 Description of internet user behaviour. In *Proc. IEEE IJCNN'99, Washington DC. Vol 4*, IEEE, Los Alamitos, CA, pp. 2514–2519.
Lawrence, S. & Lee Giles, C. 1999 Searching the Web: general and scientific information access. *IEEE Commun.* **37** (1), 116–122.
Lechner, U., Schmid, B., Schmid-Isler, S. & Stanoevska, K. 1999 Structuring and Systemizing Knowledge on the Internet – Realizing the Encyclopaedia Concept as a Knowledge Medium. In *Proceedings of the 10th international Conference of Information Resources Management Association IRMA* (ed. M. Khosrowpour). Idea Group Publishing, Hershey, PA.
Lesk, M. 1997 Going digital. *Sci. Am.* **276** (3), 58–60.
Marchionini, G., Lin, X. & Dwiggins, S. 1990 Effects of search and subject expertise on information seeking in a hypertext environment. In *Proceedings of the 53rd Annual Meeting of the American Society for Information Science* (ed. D. Henderson), Learned Information ASIS, Silver Springs, MD, (Toronto, Ontario, November 4-8, 1990) pp. 129–142.
Meadows, A. J. 1991 *Knowledge & Communication: Essays on the Information Chain*. Library Association, London.
Navarro-Prieto, R., Scaife, M. & Rogers, Y. 1999 Cognitive strategies in web searching. In *Proc. 5th Conference on Human Factors and the Web, June, Gaithersburg*, pp. 43–56.
O'Reilly, C. A. 1982 Variations in decision makers' use of information sources: the impact of quality and accessibility of information. *Acad. Mngmnt. J.* **25** (4), 756–771.
Palmer, C. L. 1996 Information work at the boundaries of science: linking library services to research practices. *Library Trends* **45** (2), 254–275.
Palmer, J. 1991 Scientists and information: II Personal factors in information behaviour. *J. Doc.* **47** (3), 254–275.
Pedley, P. 2001 *The Invisible Web*. Aslib-IMI, London.
Pinelli, T. E. 1991 The information-seeking habits and practices of engineers. *Sci. Technol. Libraries* **12**, 5–16.
Qiu, L. 1993 Markov models of search state patterns in a hypertext information retrieval system. *J. Am. Soc. Inf. Sci.* **44** (7), 413–427.
Rowley, J. & Farrow, F. 2000 *Organising Knowledge: An Introduction to Managing Access to Information*, 3rd edn. Gower, Aldershot.
Spiliopoulou, M., Faulstich, L.C., Winkler, K. 1999 Proceedings of the Workshop on Machine Learning in User Modeling, Advanced Course on Artificial Intelligence (ACAI '99), Chania Greece 1999: http://www.iit.demokritos.gr/skel/eetn/acai99/Workshops.htm
Tomov, D. T. & Mutafov, H. G. 1996 Comparative indicators of interdisciplinarity in modern science. *Scientometrics* **37** (2), 267–278.
Trott, P. 1993 *Inward Technology Transfer as an Interactive Process – A Study of ICI* Unpublished thesis. Cranfield University UK.
Vail, E. 1999 Mapping organisational knowledge. *Knowledge Mngmnt. Rev. May/June*, 10–15.
Wang, Y. 2000 Web Mining and Knowledge Discovery of Usage Patterns. CS 748T Project (Part 1). [http://db.uwaterloo.ca/~tozsu/courses/cs748t/surveys/wang.pdf]. [28/01/01].
Wilson, T. D. 1999 Models in information behaviour research. *J. Doc.* **55** (3), 249–270.