Intelligent Information Retrieval from the World Wide Web using Fuzzy User Modelling

Gabrielle Joanne Mooney

Ph.D.

De Montfort University

September 1999

Acknowledgements

The author thanks Professor J Haythornthwaite and Roy Adams for providing advice and assistance, especially during the early years of this project.

Many other people have offered support and succour, all of which was gratefully received. My thanks go to: Usha Asi, Vicki Bell, Ranbir Bhalla, Nicky Bhamra, Steve Bradford, Professor Mel Collier, Dave Houghton, Rob Knott, Raj Pandit, Anil Sharma, Louise Taylor, Angela Watkins, Chris Webb, Owen Williams, and everyone else who helped me along the way.

Finally, and mostly, I would like to express a deep gratitude to my main supervisors, Robert John and Professor Derek Teather for offering and providing their time, advice, expertise and support, and for engendering in me the belief that the project could be finished.

Contents

List of Figures

Appendices

List of Tables

Appendices

Abstract

This thesis investigates the application of fuzzy logic techniques and user modelling to the process of information retrieval (IR) from the World Wide Web (WWW). The research issue is whether this process can be improved through such an application.

The exponential rise of information itself as an invaluable global commodity, coupled with accelerating development in computing and telecommunications, and boosted by networked information sources such as the WWW, has led to the development of tools, such as search engines, to facilitate information search and retrieval. However, despite their sophistication, they are unable effectively to address users' information needs. Also, as the WWW can be seen as a dynamic, continuously changing global information corpus, these tools suffer from the problems of irrelevancy and redundancy.

Therefore, in order to overcome these problems and remain effective, IR systems need to become 'intelligent' in some way. It is from this premise that the focus of this research has developed.

Initially, theoretical and investigative research into the areas of IR from electronic sources and the nature of the Internet (including the WWW) revealed that highly sophisticated systems are being developed and there is a drive towards the integration of, for example, electronic libraries, CD-ROM networks, and the WWW. Research into intelligent IR, the use of AI techniques to improve the IR process, informed an evaluation of various approaches. This revealed that a number of techniques, for example, expert systems, neural networks and semantic networks, have been employed, with limited success. Owing to the nature of the WWW, though, many of the previous AI approaches are inapplicable as they rely too much on extensive knowledge of the retrieval corpus.

However, the evaluation suggested that fuzzy logic, with its inherent ability to capture partial knowledge within fuzzy sets, is a valid approach. User modelling research indicated that adaptive user stereotypes are a fruitful way to represent different types of user and their information need. Here, these stereotypes are represented as fuzzy sets, ensuring flexibility and adaptivity.

The goal of the reported research, then, was not to develop an 'intelligent agent' but to apply fuzzy logic techniques and user modelling to the process of user query formulation, in order to test the research issue. This issue was whether the application of these techniques could improve the IR process. A prototype system, the Fuzzy Modelling Query Assistant (FMQA), was developed that attempts intelligently to assist the user in capturing their information need. The concept was to refine the user's query before submitting it to an existing search engine, in order to improve upon the IR results of using the search tool alone.

To address the research issue, a user study of the FMQA was performed. The design and conduct is reported in depth. The study results were analysed and the findings are given. The results indicate that, for certain types of user especially, the FMQA does provide improvement in the IR process, in terms of the results. There is a critical review of the research aims in the light of the results, conclusions are drawn and recommendations for future research given.

CHAPTER 1 - INTRODUCTION

1.1 Background

It was suggested 17 years ago that the world was undergoing a third and greatest social revolution - a telecommunications and information revolution, after the agricultural and industrial revolutions of the preceding centuries (Masuda, 1982). That revolution has continued to such an extent that information is now seen as a marketable commodity to be globally exchanged. There is increasing need for timely, relevant and accurate information, and librarians and other information professionals are at the forefront of providing access to information sources. In response to this growing need, and in tandem with technological and computing advances, different delivery platforms, e.g. CD-ROM, On-line databases, OPACs (On-line Public Access Catalogues) (Heller, 1990; Herther, 1985; VUBIS, 1992; BLCMP, 1998), have been developed to provide access to sources. There has been an exponential increase in the number and types of sources of information, and perhaps the greatest boost to this has been the development, over the past 20 - 30 years, of the global network of networks known as the Internet (Lane-Lawley & Summerhill, 1994, Stephenson, 1996).

This ever-changing nature of information sources has raised users' expectations and increased demand for library and information services. Moreover, users wish to access these information sources directly, without recourse to a 'professional' intermediary. These factors, coupled with the financial pressures libraries are facing, have made it difficult to provide efficient and sufficient access (Werckert & Cooper, 1989). The combined effect of these trends has led to a recognition that, for some information services, libraries and other information providers need to put in place information retrieval (IR) systems that can incorporate the expertise of the information professional and gather knowledge about the user's experiences and preferences (Cavanagh, 1989; Morris, 1990). The goal is to produce an 'intelligent' IR system which works with the user to satisfy his/her information needs and improve the results of the IR process, thus easing the burden on the information professional.

Therefore, application of Artificial Intelligence (AI) techniques seems a likely approach to the problem (Morris, 1990). Also, as working with the user implies understanding the user's experiences, knowledge and information needs, application of user modelling techniques seems also appropriate (Rich, 1979; Thompson & Croft, 1989).

Thus, the basis of this research project is to investigate the effects of applying AI and user modelling techniques to the process of IR from the World-Wide Web (WWW). The WWW is a networked distributed hypertext (Berners-Lee *et al.*, 1992) and is a fairly recent addition to the connected family of networks known as the Internet.

The Internet began as a US Department of Defense project in the 1960s (known as ARPAnet – Advanced Research Projects Agency network (Stephenson, 1996)) and was developed through the academic community in the 1970s (Lane-Lawley & Summerhill, 1994). The advent of the WWW caused an exponential shift in the growth of the Internet such that it has become a truly global network of networks (Stephenson, 1996). Since its development in 1991-92, the WWW has quickly become the prime means of delivering information across the Internet. This is due in part to the on-going development of Graphical-User-Interface (GUI) WWW browsers (Microsoft, 1996; Netscape, 1998) and of software protocols which allowed previously heterogeneous networks to exchange information.

These two aspects have allowed researchers and developers to pursue the goal of producing WWW-based interfaces which allow relatively easy access to the different and diverse Internet information sources (Emtage & Deutsch, 1992; McCahill, 1992; Machovec, 1993; Kahle *et al.*, 1993; WAIS, 1995; Excite, 1997). In turn, these interfaces have boosted the popularity of the WWW and fuelled the integration of sources. This integration is developing so quickly that many people now incorrectly use the terms Internet and WWW interchangeably.

The amount and type of information available via the WWW is staggering - over 350 million distinct documents as of May 1999 (Alta Vista, 1999a). As it grew so network tools were designed and developed to facilitate information search and retrieval (Archie, 1995; WAIS, 1995). However, despite the tools' increasing sophistication (Alta Vista, 1999b), they are unable to capture effectively the information needs of users (Dern, 1995). Also, as the Internet can be seen as a continuously dynamic, unstructured information corpus, these tools suffer from the problems of data irrelevancy and redundancy (Ferry, 1996; Mauldin, 1991). A crucial aspect of this research is the assumption that, in order to overcome these problems and remain effective, IR systems need to become 'intelligent' in some way. Thus, the needs of the user, the financial pressure felt by libraries and other information providers, the explosion in information sources and types of access, the development of networks and the limitations of IR tools, all point to the requirement of 'intelligent' IR tools which work with the user to satisfy their information queries.

1.2 The Focus of the Research

At the heart of the discussion in the previous section is an important research issue. This is whether, and to what extent, the performance of IR from the WWW *can* be improved through the application of AI and user modelling techniques?

In order to address this issue, investigative research in a number of areas was required. These areas included: IR from electronic sources; AI and IR; user modelling; and user study evaluation. The results of this research then informed the design and implementation of a prototype IR system. This prototype was then evaluated through a user study of its performance and the results of this evaluation allowed the research question to be successfully answered.

To decide the AI approach to employ in this research, an investigation of AI and IR was required. This research was predicated on earlier background research looking at IR and electronic sources of information. The results from both areas of research implied that many of the previous AI approaches to IR - those which relied on extensive knowledge of the document retrieval set (Thompson & Croft, 1989) or which employed an AI technique that is brittle and inflexible such as expert systems (Cavanagh, 1989) - would be inapplicable, because of the ever-changing nature of the information which resides on the WWW. However, the research indicated that so-called soft computing techniques such as neural networks (McCulloch & Pitts, 1943) and fuzzy logic (Zadeh, 1965) might be useful approaches and these were further evaluated. This evaluation indicated that fuzzy logic might be the most fruitful approach to adopt.

An investigation of user modelling was required in order to decide which modelling approach might be adapted for IR from the WWW. User modelling research and development is far-reaching, not only does it find currency in the fields of AI and information systems (Kok, 1991a; McTear, 1993) but also in areas such as the psychological and cognitive sciences (Allen, 1990). Researchers employ user models for a number of reasons, including mapping users' mental models, aiding user interface design, and creating adaptive systems (Ackerman *et al.*, 1997; Borgman, 1986; Chang *et al.*, 1993; Kobsa, 1990). The field of user modelling in AI systems has rapidly expanded (Kok, 1991a, Guilfoyle, 1997; Maes, 1995). In this field user modelling attempts to deal with the characteristics of users working with these types of systems .
In the present research, the modelling of WWW users is designed to provide information about their knowledge, experience and opinion of the WWW and the information domain in order to improve their IR performance. The user models provide a partial representation of

the users' conceptual model of the WWW and information which can be used within an adaptive IR system. Results from the user modelling research indicated that adaptive models based on the techniques of stereotypes (Rich, 1979) and profiles (Cullen, 1997) would be the productive technique to employ.

This combination of fuzzy logic and adaptive user models was employed to develop the prototype. Evaluation of the prototype was influenced by the research which indicated that only certain types of WWW user would show improvement in performance. Additionally, resource constraints on the research limited the extent of the evaluation to one test-bed information domain. Thus the general research question was narrowed to a particular focus, expressed in the following statement:

> Whether the application of the prototype system can improve the IR results from the WWW, for novice and intermediate users, employing queries in one information domain.

This narrowing of the issue does not undermine the whole research. Answering the above question establishes proof of concept for one particular set of criteria and thus, for this set, the research question is answered. However, as argued in Chapter 9, the AI and user modelling techniques used, and the characteristics of the test-bed subject domain are such as to allow the scaling necessary for the more general question to be answered. The evaluation was performed through a user study. The methodology and implementation of this were informed by general research on user studies of IR systems.

The research has spanned four and a half years, during which the first six to twelve months were spent researching relevant literature in the fields of IR from electronic sources, AI and IR, user modelling and user studies. The next year was spent on design, implementation and testing of the prototype system, and on organising and designing the user study. In the third year, the first six months were spent implementing the user study and collecting, collating and analysing the results. The remainder of the time was spent writing up the research results in this thesis. A more detailed exploration of the thesis contents is now provided.

1.3 Overview of the Thesis

The research initially focused on IR from the electronic sources of information in the last 20-30 years. There has been continuing interest in the research, development and use of electronic sources of information, and any new progress in computing has been exploited by the Library and

Information Science (LIS) field (Bansler *et al.*, 1985; DMU, 1997; Saltzer, 1992). Chapter 2 examines IR from electronic sources in the light of this research. Particular interest is paid to the nature of networked IR sources as this has the greatest bearing on this research.

In Chapter 3, AI and IR is examined in detail and this chapter also covers the assessment of AI approaches in the light of the research question considered in this project. Researchers and developers have always been concerned with improving IR from electronic information resources (Arens *et al.*, 1996; Alta Vista, 1999b; Laine-Cruzel *et al.*, 1996; Morris, 1990; TREC, 1996). Often, this has been attempted through the application of some facet of AI (Maes, 1994a; Miyamoto, 1989; Radecki, 1976). These attempts have met with limited success and, in Chapter 3, it will be seen that many have relied on extensive knowledge of the retrieval corpus (CIIR, 1997; Gordon, 1988; Terano *et al.*, 1992).

As the WWW is ever changing, it is effectively unbounded and unknowable *in toto*. It is argued in Chapter 3 that this rules out most of the previously used AI approaches but that Fuzzy Logic (Zadeh, 1965), with its inherent ability to capture and represent partial knowledge, can be considered to be a valid approach. The crux of this research is that by modelling the users' WWW knowledge and experience, then users' performance in interacting with the WWW to retrieve information will be improved. User modelling has been employed in a number of AI systems (Kok, 1991b), including intelligent tutoring (Kass, 1989), dialogue systems (Kobsa & Wahlster, 1989) and intelligent information retrieval (Brajnik *et al.*, 1987). Chapter 4 reviews the literature on user modelling, specifically in regard to AI and IR. There has been much research into the employment of user stereotypes (Rich, 1979) as a way to capture and represent user information, and recently into the use of user profiles (Cullen, 1997) to assist IR. It is argued in Chapter 4 that the approach adopted here is related to both these techniques.

This combined application of fuzzy logic and user modelling to IR from the WWW is novel and the approach is discussed fully in Chapter 5. At the heart of the application is a Fuzzy Knowledge Base (FKB) and a number of fuzzy sets. Each set is effectively an adaptive user model representing some information about a particular user.

Chapter 6 covers the design, development and implementation of the FMQA (Fuzzy Query Modelling Assistant) - a prototype system which was developed using the approach described in Chapter 5. It will become evident that the FMQA is neither an intelligent search engine nor an intelligent agent. (The notion of agency and distributed artificial intelligence (DAI) is explored in Chapter 3.) Rather, it seeks to intelligently assist the user in improving their IR results through query refinement. The query is modified before it is submitted to an *existing* WWW search tool. In this manner, the FMQA effectively sits between the user and the existing search engine, and it is more correct to view the FMQA as an intelligent assistant (Maes, 1994b).

The design and implementation of the user study are reported in Chapter 7. An examination of the research literature on user studies and IR revealed that there are merits in both tightly controlled studies (Turtle & Croft, 1991; Belkin *et al.*, 1993) and open or 'live' studies (Morita & Shonoda, 1996). It is argued in Chapter 7 that because an effective commercial intelligent WWW IR system would have to allow global access to all, then any user study of a prototype system should be carried out under 'live' conditions. However, to allow the results to be compared, some factors were initially controlled. Thus, the user study performed can be seen to have two distinct parts - one controlled and one totally live or ad-hoc. To answer the research question, the following sub-questions had to be answered:

How are the categories of novice and intermediate user characterised?

How is 'improvement' to be measured?

What are the characteristics of the test-bed domain?

Here, the design of the FMQA influenced the methodology of the user study and Chapter 7 describes how these issues were resolved. The evaluation of the user study involved capturing and collating various data concerned with the users' interaction with the FMQA. This data was collated and analysed in the light of the design of the user study and the issues raised in the questions above. This evaluation is reported in Chapter 8. The study results are interpreted and discussed with regard to the overall research aims. It is argued that, within the parameters of the research and the user study, the results show an improvement in user IR performance when they employ the FMQA as opposed to just using an existing WWW search tool.

The overall conclusions of the research are given in Chapter 9. The implications for, limitations of, and possible future directions for the research in the field of Intelligent IR and Library and Information Science are discussed. This novel combination of fuzzy logic and user modelling has shown it is possible to improve IR performance for one particular set of conditions. It is argued in Chapter 9 that this approach is generalisable and would also, because of its basic concept, be portable to other areas of current IR research, such as digital libraries (DMU, 1997;elib, 1998) and intelligent agents (Maes, 1994a; AgentWeb, 1998).

CHAPTER 2 - INFORMATION RETRIEVAL FROM ELECTRONIC SOURCES

2.1 Introduction

The field of IR from electronic sources is quickly evolving, with changes driven partially both by technological advancements and by developments in the organisation and presentation of information. This chapter investigates research conducted into this field – the investigation was performed to underpin and provide support for the focus of the present research.

The sustained and growing interest in the research and development of electronic sources of information (Bansler *et. al*, 1984; Bourne, 1980; Follet *et. al*, 1993; Hartley *et. al*, 1987; Parfell, 1987; Saffedy, 1993; Saltzer, 1992) is one of the major reasons why IR tools need to become intelligent, as will become apparent in this chapter. This interest has fuelled the creation and exploitation of such information sources as On-line Public Access Catalogues (OPACs), CD-ROM databases, On-line databases and networked databases, and the development of sophisticated user interfaces with which to access these sources. Other potential sources, specifically focused on libraries and their users, are the subject of on-going research such as De Montfort University's (DMU)'electronic library' projects (Collier, 1993; DMU, 1997).

These advances have had great effect on how libraries and other information providers offer access to the information sources (Follet *et. al*, 1993). Certain technologies have been and are currently suitable for delivering particular forms of information, because of a temporal favourable conjunction of the two aspects. For example, OPACs - containing bibliographic data - developed on-line because that was the available medium, although nowadays they are available on CD and the WWW. CD-ROM has been developed as the media for delivering multi-media information, but efficient retrieval of this sort of data over networks is becoming easy (*NCSA*, 1997; *Realplayer*, 1998). It is expected that the recent developments in optical storage capabilities, such as Digital VideoDisc (DVD), will revolutionise the delivery of data in all its forms (Weinzerl, 1998).

It has been recognised that, potentially, every type of medium could be used to deliver every type of information (Adams, 1990). This research does not seek to address what has been called the 'delivery platform versus content debate' – there is on-going research into which platform of delivery, e.g. CD-ROM, WWW, is the best for which type of content, e.g. video, sound, multi-media. It is probable that a situation in which there was multi-platform delivery of different information types would exacerbate rather than alleviate the current problems regarding IR from electronic sources, because of the increase in different information formats and different user interfaces. This chapter examines IR from electronic sources as it currently stands, in order to highlight the diverse nature of existing information sources and to provide the background to this research.

The following subsections investigate each source in turn, considering the relative advantages and disadvantages. Section 2.2 examines *Library-based sources*, and includes sections on OPACs, CD-ROM databases, and on-line databases. . In section 2.3 the concept of the *Electronic Library*, a fairly recent development within libraries, and its implications for information retrieval are discussed. Section 2.4 investigates *Networked Information Retrieval* and, bearing in mind the focus of this research, concentrates mainly on the *World Wide Web* Following this, future trends in IR from electronic sources are identified, and implications for this research are summarised.

2.2 Library-based sources

2.2.1 Overview

Enormous changes have taken place in the provision of information sources within academic libraries over the past twenty years (Bourne, 1980; Clark & Mischo, 1992; Hall, 1986; Heller, 1990; Herther, 1985), and this has included the development of new electronic sources, such as OPACs, CD-ROM and On-line databases - this section investigates and compares these and considers the implications for information retrieval.

2.2.2 OPACs

OPACs are on-line catalogues of library stock, giving users access to information about the current books and periodicals held. This information is held and displayed in the form of bibliographic database records whose fields can include: title, author, publisher, number of copies held, location, availability and library class number. These records are accessed usually from a terminal via keyboard commands and a number of terminals provide, through a server, a networked system (VUBIS, 1992, BLCMP, 1998). Search facilities are provided with the user able to search each entry on such fields as author, title and keyword and the OPAC returning exact and nearest matches to the search string in the form of a browsable index. Boolean combination is allowed. Some libraries, e.g. DMU, have extended the usability of their OPACs by adding options such as placing reservations.

OPACs replaced the old card catalogues, which represented a serious step forward in the automation of library function and in providing greater and more open access to the stock. Since their introduction in the early 1980s they have proliferated rapidly (Clark and Mischo, 1992). Ironically, this proliferation has created one of the disadvantages of OPACs - there being many different types with different interfaces which the untrained user must learn to navigate. However, recent developments integrating networked OPACs under a WWW interface (see section 2.4) are beginning to reduce this problem (British Library, 1997; CURL, 1997; DMU, 1998a; Scott & MacDonald, 1998). Other disadvantages of OPACs are the lack of content information about titles and the inability to browse contents of periodical issues. This last drawback has been alleviated to some extent by the introduction of CD-ROM technology which is discussed next.

2.2.3 CD-ROM

CD-ROM technology became commercially available during the latter half of the 1980s and its impact accelerated during this decade (Herther, 1985; *Online and CD Notes*, 1994a). This adaptation of optical disc technology revolutionised the storage and manipulation of data through the ability to store on one disc a much larger amount of information (650 Mb) than on the old floppy discs. This capacity meant that CD-ROM became the medium for the delivery of multi-media information, not only in academic and business areas but also in the fields of education and entertainment (*SilverPlatter*, 1998; *Microsoft*, 1997).

This storage capacity has been exploited by database manufacturers in producing large databases of specialised information, such as LISA - the Library and Information Science Abstracts - and INSPEC - a science-based database, for research use in libraries, information centres and businesses. *The CD-ROM Directory* lists nearly 13000 CD-ROM titles available in 1996 (TPFL, 1996). By 1999 (Waterlow New Media, 1999), an on-line multimedia product database which is derived from the earlier directory lists more than 31000 multimedia titles. Each database is contained on one or more discs and each presents the user with a unique interface which allows searching of the material. This searching is more sophisticated than that of OPACs as the user is generally allowed to use Boolean operators and predicate logic to produce quite complicated search strings in order to limit the number of results returned and focus the topics to which they refer. Most CD-ROM databases employ a thesaurus to aid keyword searching.

Initially, libraries installed these CD-ROM sources as stand-alone entities, with a PC dedicated to the task and the CDs kept separate from the machine (Adams, 1990). This is still the system employed by some libraries although, as the popularity of CD-ROMs has increased and with hardware improvements, many others have moved towards networking all their CDs so as to provide for a multi-user network of CD-ROM databases (DMU, 1998b).

Networking the CD-ROMs has its own implications (Van Brakel, 1993) as it involves taking what has been developed as a single-user device and making it available to many simultaneously. The CDs are usually kept permanently in their drives and the series of drives are controlled by a dedicated computer (called a 'server') which also controls access. Access is via a series of terminals connected to a server on a network. The user normally chooses the CD title to access from a menu screen. Most CD search software allows for the saving and printing of retrieved information.

The initial networking of CD-ROM databases led to considerations that they might replace an older form of electronic information retrieval - On-line databases (Stratton, 1994) - which are discussed next.

2.2.4 On-line databases

On-line databases are massive collections of information held on magnetic tape or disc which can be accessed remotely by users through a suitably attached terminal. These databases cover every field of endeavour, from the sciences, arts, media and financial areas, and provide information such as journal contents, abstracts, statistical data, patents, news stories, and reports. They developed rapidly in the 1960s and 1970s (Hall, 1986) and have continued to grow in number - 5511 databases in 1996 according to one source (GALE, 1996). A distinction must be made between *hosts* and *providers*, in that the information resides at the host computer. These databases are commercially operated by companies - one of the first and biggest is The Dialog Corporation (DIALOG, 1998) (formerly DATASTAR-DIALOG) - which acquire information from the *providers*, e.g. journal information from publishers in a certain field, and then make it available at a cost to the prospective users. In the case of academic institutions, the library service acquires the on-line databases on behalf of the users - research staff, students and others - and makes the service available to them.

Undertaking the task of producing an on-line database is quite a project. The *Beilstein Handbook of Organic Chemistry* is a classic work of organic chemistry and consists of more than 350 printed volumes and 275 000 pages of text. Its conversion to an on-line database took four years and its development continues (Heller, 1990). The producers of the Beilstein now provide access to its database, which now contains information about 8.5 million compounds, via the WWW (Beilstein, 1999). The major providers each have their own system and many databases are available in different systems - the Beilstein being an example. Also, a host like Knight Ridder Information offers many databases under its system, the advantage to the user being that they have only to learn one unique search interface. On-line database interfaces allow for quite sophisticated searches involving Boolean constructions. Companies are constantly trying to improve their products and introduce more databases on-line to maintain their share of the market under competition from other technologies (*On-line and CD Notes*, 1994b). Recently, the impact of the Internet is beginning to be felt. Some companies are starting to offer access to their databases through the WWW (Medline, 1998) and this is another example of the increasing integration of electronic information resources.

2.2.5 Discussion

Library-based electronic sources of information continue to change and develop under the influence of accelerating computing and telecommunications advances and a changing financial situation. OPACs are now available on CD-ROM and via the WWW (DMU, 1998a; Scott & MacDonald, 1998). The networking of multi-CD databases and the continuous development of on-line databases has increased enormously the amount of information available to potential users.

Some have considered CD-ROM will replace on-line services, as a huge number of on-line databases are also available on CD-ROM (Stratton, 1994). There are advantages and disadvantages to both.

CD-ROM can provide free ' access ' without the telecommunications and access costs involved in using on-line databases. CDs can offer full-text documents from the database which can be downloaded or printed out - a very expensive procedure on-line.

Against CD-ROM is its relative lack of currency. A number of on-line databases are upgraded daily whereas upgrade discs are released perhaps monthly or quarterly which in some fields of endeavour - for example certain areas of business and finance for which timely information is crucial - is not good enough. Some hosts, e.g. Knight Ridder, have solved this by offering their services on CD-ROM as complementary to on-line. Compared with on-line, CD-ROM fares badly in terms of storage capacity as, although 650 Mbs is impressive for a small portable disc, it cannot cope well with large databases such as MEDLINE which needs 13 discs. Simultaneous access to multi-disc databases can be achieved through the networking of

CD-ROMs but the incumbent costs might outweigh the savings obtained from moving away from on-line (Atkinson & Yokley, 1993; Stratton, 1994).

In recent years, there has been a number of developments in CD technology, such as CD-I, CD-TV and Photo-CD (Kroeker, 1997) which imply that as a delivery medium, CD technology is still a force to consider. Additionally, the advent of CD-RW, rewriteable CDs, means that this form of optical technology may be able to compete as an information storage medium. However, other optical storage developments such as WORM (Crane, 1991) and DVD (Weinzerl, 1998) may eventually supersede CD-ROM. Also, the accelerating technological developments in networking (see Section 2.3) mean that CD-ROM is under pressure as the most effective delivery medium for multimedia data. However, it seems that the continuous growth in the number of CD-ROM titles available (Waterlow New Media, 1999) and its obvious success in audio format mean that it is still a potent force. Hosts of on-line databases are turning to providing access via the WWW. Some publishers of CD-ROMs, though, have been taking advantage of the explosion of the WWW, by providing in-built links to websites within the data presented. These web sites offer information about upgrades, errata and new products (*RS Components*, 1998).

These last two points are examples of the drive towards integration of electronic sources of information. This drive perhaps means that neither CD-ROM nor On-line databases will replace the other totally as the way to provide users with database information. Rather, the idea of a personal information resource is crystallising. This concept is one of a desktop computer through which all electronic information resources are available under a transparent seamless interface. A great impetus to this idea has been the development of the 'electronic' library, which is beginning to have an impact upon library-based sources of information. This development is discussed next.

2.3 The Electronic library

2.3.1 Overview

The concept of the electronic library is one of a teaching, learning and research environment in which information is held primarily in electronic form. It has been born out of technological developments offering academic institutions the opportunity to develop and provide a radically different type of information service and of financial constraints preventing institutional growth along traditional lines (Collier *et al.*, 1993). Electronic libraries are being developed in many places (Manoff *et al.*, 1993; Story *et al.*, 1992; Troll, 1993), not least at De

Montfort University in Leicester where a number of research projects has been performed (DMU, 1997).

The main objective of an electronic library is to provide the user with their own personal library through the medium of computers. But the final goal entails more than this - it also means providing the services offered by a traditional research librarian. Whereas this goal of providing the user with their own 'electronic' librarian may be distant, recent improvements in hardware, increases in the bandwidth of networks and the development of multimedia applications have meant that some aims of the electronic library are achievable now.

An electronic library differs from another source, say an OPAC, in that it stores the whole contents of a book or journal, not just a short record. It differs from networked sources because they are concerned with providing information over large areas whereas electronic libraries are designed to provide a service similar in function to libraries today, at a local level (although 'local' may apply to a multi-site institution such as DMU).

2.3.2 The Electronic Library at DMU

DMU's interest began with the goal of developing an electronic library within five years and a research project called ELINOR (Electronic Library Information On-line Retrieval) was set up in March 1992 as a two-year pilot (Collier *et al.*, 1993). The project was broadly successful (Wu *et al.*, 1995) and since then other electronic library projects have been initiated at DMU, e.g. ELSA, ELISE I and II, MINSTRIL, ELIVIL, Phoenix and ERCOMS to name some. Phoenix was a project designed to provide on-demand electronic publishing of materials such as lecture notes and journal articles (Phoenix, 1997). ERCOMS is concerned with providing information on copyright management (ERCOMS, 1997). ELISE is an image retrieval and display system (Eyre, 1997). Currently, many of the electronic or digital library projects are carried out under the auspices of the *Digital Library Group* (Digital Library Group – DMU, 1999). Projects developed by the group include ExamNet (1999), a collection of examination papers available via the electronic library, and GASS (1999), an extension to a previous project which looks at delivering electronic journals via the electronic library. Many of the projects were carried out in consortia, with other universities and companies. The Follet and FIGIT reports (Follet *et al.*, 1993; FIGIT, 1994) gave great momentum to the development of electronic libraries and there are currently over 50 projects on-going as part of the E-Lib programme (elib, 1998).

When ELINOR was conceived at DMU, it employed a commercial Database Management System (DBMS) and its own unique Windows-based interface (Wu *et al.*, 1995). With the rise of the WWW, a great number of current electronic library projects are being developed with WWW interfaces. Thus, the drive towards integration manifests within the realm of electronic libraries as well. Using the hypertextual approach (see section 2.4) of the WWW allows the production of a standardised user interface to all the electronic library services on offer. It also allows efficient networking of the resource and allows librarians and other information providers to offer this resource to users along side the more established other resources, such as OPAC, thus going some way to achieving the stated goal of providing the user with their own personal library.

As new technological developments occur and attitudes towards the provision of information and what it is change, then so the definition of an electronic library has changed and will change (Collier, 1995). Whilst existing electronic library projects are moving some way to providing the user with access to the immense and ever increasing amount of information that exists in all its myriad forms, soon - because of the exponential increase - to access this information *effectively* electronic libraries will need to become *intelligent* in their behaviour and the way they interact with both the user and the information source. Only then will we begin to move from electronic library towards *electronic librarian.*

Recently, the goals of the electronic library (or digital library as it has become increasingly known) are being promulgated through the rise of the Internet and the WWW. Many projects are focusing on the global network as a way to deliver these goals. Bishop *et al.* (1997) report on progress in building a digital library at the University of Illinois which aims to deliver material from commercial and professional society publishers. Fox *et al.* (1997) discuss a networked digital library of theses and dissertations and (Salampasis *et al.* (1996) argue for co-operative IR in digital libraries through communication between distributed databases via information agents (see chapter 3 for discussion of agents). It seems that the present and future success of library-based resources lies in the fact that they can be networked and it is this form of information source that is discussed in the next section.

2.4 Networked Information Retrieval

2.4.1 Overview

The most important information resource in the near future for libraries and information users will undoubtedly be networked information sources (Saltzer, 1992). Numerous networks developed during the 1970s and 1980s - for example, JANET (Joint Academic Network) linking Universities in the UK.

The development of efficient protocols to allow data packaging, control, transmission and exchange (Lane-Lawley & Summerhill, 1994) aided this growth of networks: two of the most well-known being the TCP/IP suite (Transmission Control Protocol/Internet Protocol) and the OSI (Open Standards International) seven layer suite. These protocol suites allowed for the linking of networks across the globe and the result has been given the catch-all term of the INTERNET. It began as a US Government defence project in the 1960s (Stephenson, 1996), designed to show the viability of storing sensitive information on a number of geographically distributed computers (Lane-Lawley & Summerhill, 1994). The discovery that the network offered a cost-effective mechanism for academics to communicate and exchange information boosted its growth and use, and University College in London was the first non-American link. Users with access to the Internet can exploit the vast information resources residing on hundreds of networks around the globe. It is the interoperability of the different networks that comprise the Internet that give it its strength.

As the Internet has grown, so has software developed to support the possible activities . Amongst the transactions possible are:

- *E-mail* - a very popular application used to send messages across global networks (Winder, 1994).
- *Mailing lists* - an extension of e-mail in which a group of people communicate with each other on a specific topic.
- *USENET Newsgroups* - similar to mailing lists but with a bulletin board structure. Anyone can post and read messages to hierarchically arranged topics (Braun, 1994). Many groups have a FAQ (Frequently Asked Questions) which is a special file containing answers and information regarding questions that have been raised often.
- *FTP and Archie* - FTP (File Transfer Protocol) is the Internet protocol and application used to move files between Internet hosts. Archie is a tool designed to index and search FTP resources (Emtage & Deutsch, 1992).
- *Gopher and Veronica* - Gopher is a menu-based system for searching and retrieving information on the Internet which is organised hierarchically from directories down to files (McCahill, 1992). Veronica (Very Easy Rodent-Oriented Net-wide Index to Computerised Archives) is designed as an index and access tool to the hundreds of Gopher databases that exist (Machovec, 1993). Recently, this way of organising networked information has become considerably less popular, its demise fuelled by the rise of the WWW.

- *WAIS* - WAIS (Wide Area Information Server) is another way to search and retrieve information on the Internet (Kahle *et al.*, 1993). WAIS databases are globally distributed and organised so that a user can perform full-text searching of any databases (WAIS, 1995).

By far the most popular to date and the most quickly developing part of the Internet is the WWW. As the focus of this research is IR from the WWW it warrants closer scrutiny and it is discussed in greater detail next.

2.4.2 World-Wide Web

The World-Wide Web, developed at CERN by Berners-Lee *et al.* (1992), is the aspect of the Internet where information search and retrieval and hypertext techniques merge. The attraction of hypertext is the potential it shows for providing tools to aid information seeking that go beyond those available in simple text-based media. Whilst text can be read non-sequentially in simple text documents, hypertext allows this to take place easier by providing pointers which link different portions of text (and now other media) non-sequentially (Nielsen, 1990). Basically, pages of information are linked together by the employment of pointers within the pages. For example, there maybe a pointer in page 1 of a document which points to page 34 of the same document. Additionally, these hypertext pointers allow information to be accessed in different ways, at different points within a document, or from pointers from different documents. This ability of hypertext encourages the linking of documents which perhaps enriches the information provided in the original document and exhibits a relationship between documents which otherwise might not be readily be apparent. These pointers or links are often, in the case of a WWW document, highlighted text but they can also be images and other media. A reader of a hypertext document can use the links to jump from one place to another within the document. In the WWW this is normally achieved by placing the mouse pointer over the link and clicking the left mouse button once or twice and jumps can be made to other documents as well as within documents.

Within the WWW, then, information is organised into a distributed hypertext system consisting of documents, links and indexes. Indexes are special documents which are searched, the result being another 'virtual' document containing links to the found documents (Berners-Lee, 1995a). Hypertext documents contain links to other documents, or places within documents. All documents are contained within the same addressing scheme and are given a unique identifier called the URL (Uniform Resource Locator), for example *http://www.dmu.ac.uk/* which is the address of the 'home page' of De Montfort University.

It is also perfectly possible to create home page and other documents (and create links to other parts of the WWW therein) which are unique to a particular WWW server. In figure 2.1 the author's Ph.D. project page is shown, as it appears using the WWW Graphical User Interface (GUI) *Netscape* (see below). The interfaces to the WWW are often referred to as browsers, as they allow the user to view or 'browse' the documents.

As the WWW developed, WWW clients were designed to allow access to Gopher, FTP and USENET information sources directly. This is one of the reasons why the WWW has grown exponentially over the last 5 years. As of May 1999, one source of statistics estimated the number of URLs to be over 350 million! (Alta Vista, 1999a). All these are theoretically accessible via the use of a WWW client browser or navigation tool. The WWW is continually being given new impetus from new developments. Some of the latest are the growth in on-line commerce – financial transactions via the WWW (*Mecklermedia*, 1998) - and the development of corporate Intranets (Earthweb, 1998; Gralla, 1996).

Intranets are private networks designed along the same structure as the WWW. They allow the information resources within a company to be made readily available to staff via a easy-to-grasp and user-friendly interface. Passwords are often employed to authorise access, and using security measures allows the company to offer staff access to the WWW via the Intranet whilst barring access to the Intranet from the WWW. Recently, some firms have considered the possibility of including others, such as suppliers and customers, within an Intranet – in these cases the network is termed an *Extranet* (Jones, 1998).

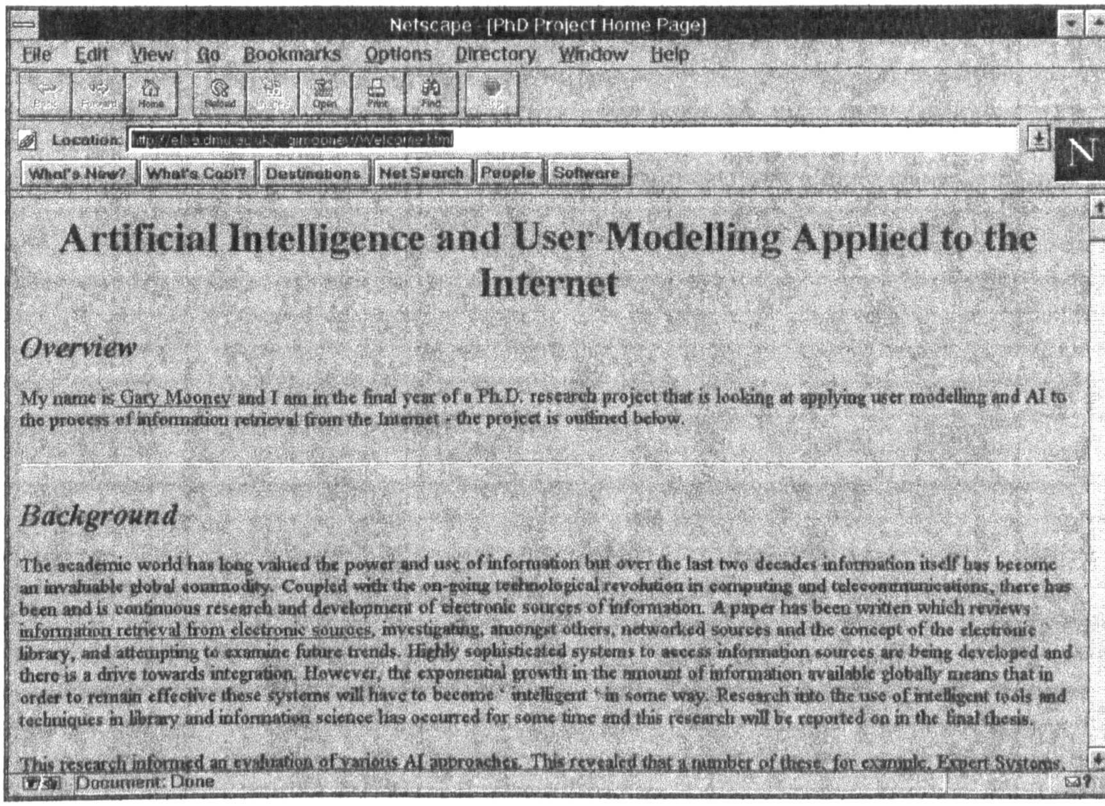


Figure 2.1 A WWW ' home ' page with URL displayed at the top.

The WWW was initially designed to disseminate textual information amongst the high-energy physics community and early browsers, such as Agora (Secret, 1995) and Lynx (Montulli, 1995), were text-based. Continuous developments in network technology, such as increasing bandwidth capabilities (Gilder, 1994), and increased user expectation for multi-media information (due, in part, to the explosion of CD-ROM sources) has made it possible for non-textual information such as sound and video to become part of the WWW. Consequently, text-based browsers have given way to GUI browsers such as Mosaic, Netscape (*Netscape*, 1994; 1995; Smarr, 1995) and, Microsoft's Internet Explorer (*Microsoft*, 1996).

These browsers have become quite sophisticated and continue to develop and increase their functionality. Formats such as RealAudio, RealVideo (*Realplayer*, 1998) and Videoconferencing (Fourcher, 1997) are now being supported through the WWW and the effects of the recent development of the Java programming language (*Sun Microsystems*, 1998) - effects which could change the whole emphasis of the WWW - are yet to be fully absorbed. Java is a computer language based on the language C++ which affords a developer and builder of WWW pages the opportunity to create programs and applications which will execute when the WWW page containing the program – often called an *Applet* – is accessed by a WWW user. These programs can range from the quite simple, e.g. animated images and pop-up windows, to the quite sophisticated, e.g. accessing, searching and displaying results from a database. The use of Java-based programs in the WWW may change the emphasis of the network from the static – pages of multimedia information that the user accesses and views – to the dynamic – pages which display information that is dependent upon the interaction between the user, the applet and that upon the which the applet acts, e.g. a database. This latter interaction may be influenced by the characteristics of the user, e.g. their level of access at that particular WWW page. The growing use of Java is evidenced by the continuing development of the language and its exploitation in different ways. Recently, Sun Microsystems have introduced Jini (*Sun Microsystems,* 1999), Java-based technology designed to make the networking of a broad range of devices – computers, printers, phones, cameras – as simple as possible.

Researchers and developers have used the increasing popularity of the WWW to develop links (called gateways) between it and other heterogeneous information sources such as WAIS and Z39.50 databases (Govert & Pfeifer, 1997; Nassar, 1996). This activity is partially driven by a desire to combine the excellent browsing capabilities of the WWW with the sophisticated search and retrieval capabilities of said databases. It is another example of the drive towards the integration of the various electronic resources under one interface. Over the past five years, these different developments have confounded predictions of a fall off in the growth rate of the WWW - it continues to grow at a prodigious rate.

As the WWW has grown the problem of where to find required information has increased in tandem. An additional problem, not addressed by the browsers, is one of moved, broken and lost links within the web structure. What were needed were search tools and not surprisingly a series of different kinds have been developed to address these issues. World-Wide Web Robots, Wanderers, Spiders and Search Engines are all names given to WWW search tools - computer programs that traverse the web automatically (Alta Vista, 1999a; Excite, 1997; Fielding, 1994; Lycos, 1997; Nierstrasz, 1995). These software tools are designed for different functions but mainly the purpose is to keep track of the ever expanding WWW and to provide information *about* the information contained therein (Koster, 1995). All search tools exist on servers at WWW hosts around the globe and are thus theoretically accessible by anyone with a WWW client.

All WWW search tools employ different methods in an attempt to perform the same action. One disadvantage of this is that different tools performing the same query are highly unlikely to return the same set of matching documents. This is due to number of factors, including the different search strategies employed by tools and the differing sizes of databases they build, index and search. This problem, and that of maintenance, is likely to reduce user confidence in the information search and retrieval capabilities of WWW search tools (Ferry, 1997; Catledge & Pitkow, 1995). The pre-search query refinement abilities of WWW tools are limited. This can lead to too much information being returned, rendering the search results difficult to use or, at worse, unusable, or to the user believing that essential information has been missed. Despite this, the WWW is the fastest growing and most popular part of the Internet. It continues to develop with navigation and search tools being continually improved (Alta Vista, 1999b).

2.3.3 Summary

Over the last 30 years an interconnected network of networks has spread out across the globe to provide an enormous wealth of accessible information. The development of the Internet and tools with which to access it have also raised network users' expectations with regard to information retrieval from networks. The increasing number of users and amount of information transmitted has undoubtedly meant that networks are the most important information resource today, but users want and expect more , e.g. faster access, faster retrieval, tools to aid searching, the ability to transfer sound, video and large software files.

Whilst developments in hardware and software might cope with issues such as increased network traffic, the call for greater bandwidth, and the move from text-only documents to pictures, sound , video and large encoded software files, the problem of where to find the right information and how to search for it successfully will remain and become exacerbated. Current search tools (Alta Vista, 1999a; Excite, 1997; Lycos, 1997), despite their increasing sophistication, are unable to cope with IR problems such as irrelevancy – information returned by the tool which is inapplicable in terms of satisfying the user's need - and redundancy – the tool returns repeatedly returns the same information as separate documents. Arguably, *intelligent* tools are needed to prevent users from losing confidence, not only in the search and retrieval capabilities of the tools, but also in the *worth* of the networked information itself (Dern, 1995). This point is discussed further in the last section of this chapter, the conclusions, which also briefly examines possible future trends in IR from electronic sources.

2.5 Conclusions

This chapter examined IR from electronic sources and found a rapidly developing field, driven by computing and technological improvements, changing user expectations and social and economic changes. There have been enormous changes to library-based and networked information sources in the last three decades.

The added usability provided by OPACs has, through raising expectations, increased the pressure to provide access to more detailed information than is traditional, such as citations to periodicals at article level. It seems that the future of library OPACs lies in integration with other information sources such as on-line databases, CD-ROM databases and the WWW (Brown, 1995; Clark & Mischo, 1992; Lund University, 1997), in order to meet new demands and develop new services.

On-line databases have been available to users since the late 1960s, and their numbers continue to grow. In answer to the introduction of CD-ROM, direct end-user searching is becoming commonplace and sophisticated user interfaces have been developed (Brook Long *et al.*, 1992) - this trend looks set to continue with developments in on-line databases occurring regularly (Basch, 1995; Bjorner, 1995) and a move towards providing WWW-based interfaces (IOP, 1999; ROSAT, 1997).

As noted, the networking of CD-ROM has delivered multi-user access to numerous databases (Van Brakel, 1993) and user interfaces have become fairly sophisticated (Zink, 1992), but CD-ROM cannot presently compete with on-line for frequency of upgrades, access time or storage capacity for the largest databases (Stratton, 1994). CD-ROM seems set to make an impact on print, as CD-ROM publications whose origins lie in print gain in number. There are advantages in this for librarians, as networking publications could provide wider access than achievable through print alone.

CD-ROM may find a niche in the continuing development of the electronic library as there is potential for networking some CD-ROMs (Stratton, 1994) and making them available through electronic library software. Also, a combination of CD-ROM, the electronic library and networks threatens the future of periodical journals and reference works. Perhaps though its real future lies as a database delivery medium, for as a storage device it may become obsolete as the new emerging optical techniques, e.g. WORM and DVD, become widely available (Crane, 1991; Weinzerl, 1998). There have been for some time combined CD-ROM/WORM systems in operation (Thiel, 1992).

The electronic library provides a way to offer the library resources and services on-line through the user's computer. Funding from national governments and the EU are fuelling major electronic library projects in many places (Elib, 1998). Again, the primacy of the WWW is offering a way to provide efficient access to electronic library resources and this trend may continue.

The amount of information available over networks is quite staggering. Networked information is set to be *the* resource in the future. The WWW continues to grow and other ways of organising and delivering information, such as WAIS and other Z39.50-based systems, are developing apace (Cheney, 1995; Russel, 1994). There is and will continue to be a drive to integrate access to all electronic information sources in one multi-functional usable personal workstation (Brown, 1995; Clark & Mischo, 1992; Spring, 1993). The user interfaces employed on these workstations will be crucial to the sources. Interfaces have become quite sophisticated, with complex searching and browsing functions. Presently, these tend to be developed as WWW interfaces but the WWW may eventually be overtaken as the premier delivery medium within the Internet (Nadeau, 1997).

Any investigation can only be a temporal snapshot. It is hard to predict the future but it can be said, with confidence, that electronic sources per se are set for massive growth. This will be driven by users wanting quicker and better access to all kinds of multimedia information, situated both locally and remotely, and to tools to search, retrieve and manipulate this information. Existing tools, despite their sophistication, are unable to cope with the classic problems of IR - redundancy, overload and irrelevance.

All the evidence points to the need to develop *intelligent* tools which aid users in the process of IR and which help to satisfy their information requirements (Dern, 1995). This need to develop intelligent information retrieval has been recognised for some time and research has taken place in the fields of librarianship and information science, as well as part of mainstream AI research. There has also been some research into solving these problems in different parts of the Internet (Jennings *et al.*, 1993). This research is investigated in chapter 3.

The other major conclusion from the research in IR from electronic resources is that the integration of sources is being driven by the WWW. Hence, the reason why the research is focusing on applying AI and user modelling techniques to improving IR from the WWW. This necessitated an investigation of AI and IR in general, which led to an evaluation of which AI techniques might be fruitful for this research. These are both covered in chapter 3 as well.

CHAPTER 3 - INTELLIGENT INFORMATION RETRIEVAL

3.1 Overview

The desire and need to improve the results of the IR process has long been recognised (Bush, 1945; Salton, 1983). Over the past 20 years there have been many efforts to build systems that attempt to improve the results of IR through assisting the user (Arens *et. al*, 1996; Alta Vista, 1998a, Bassano *et. al*, 1992; Botman *et. al*, 1987; Emtage & Deutsch, 1992; Frei & Stieger, 1995; Laine-Cruzel *et. al*, 1996; Lycos, 1996; Maes, 1994a; Miyamoto, 1989; Morris, 1990; Radecki, 1976; Thompson & Croft, 1989; TREC, 1996; Werckert & Cooper, 1989; Yahoo!, 1998). Whether these and other IR systems can be deemed to be intelligent is debatable (Anick, 1997; GANNET, 1997; Gilbert & Janca, 1996). This research is not concerned with the debate about what intelligence is or whether human intelligence can be effectively captured, represented and employed within information systems. Rather, it examines the application of AI and user modelling techniques to IR from the WWW. The aim is to perform 'intelligent' IR. Here, the word 'intelligent' is understood to mean the application of some AI technique and not an attempt to mimic or reproduce human thought and reason. Therefore, the investigation of intelligent IR in this chapter is an examination of the application of AI to library and information systems in general, with particular focus on examples concerned with IR.

The rest of the chapter is constructed as follows: section 3.2 briefly discusses the definition of AI; section 3.3 explores the use of AI in libraries and information science before section 3.4 widens this investigation to cover the use of different AI techniques for IR. Section 3.5, building on the results reported in the previous sections, focuses the investigation on the present research, and covers an evaluation of a number of AI techniques. The focusing of the research necessitated an investigation of query refinement techniques and this is reported in section 3.6. The final section, section 3.7, offers conclusions and indicates which AI technique is the most promising to use to propagate the aims of the present research.

3.2 Artificial Intelligence

Although the development in 1949 of the stored program computer provided the technology that led to the development of AI, it was not until the 1950s that the link between human intelligence and machines was observed. Weiner (1948) theorized that all intelligent behaviour was the result of feedback mechanisms and that these mechanisms could possibly be simulated by machines. In 1950, Turing's seminal work on the development of intelligence within computing was published (Turing, 1950). In 1955, Newell, Simon and Shaw developed *The Logic Theorist*, (Newell *et al.*, 1956) considered by many to be the first

AI program. However, it is generally accepted that AI came into being, as a separate entity within computing, at the Dartmouth Conference of 1956. Artificial intelligence is the branch of computer science involved in the design and construction of computers to perform tasks which, in humans, require intelligence. Winston (1984) defines AI as *'the study of ideas that enable computers to be intelligent'* (p.1). One major problem with this definition is the present inability to understand and define intelligence completely, even though it is often recognisable. Thus there are basically two goals of AI: to make computers act more like intelligent humans and to understand the principles that make intelligence possible (Smith, 1986). Artificial intelligence itself has spawned different branches, such as Expert systems and Neural Networks, and these, in turn, have all been thought suitable for improving the performance of the IR process.

3.3 The use of AI in libraries

In the past, the application of AI to library functions has mainly been limited to the use of Expert Systems (ES) or Knowledge-based Systems (KBS). These are the oldest class of systems designed to behave 'intelligently'. They basically consist of a *knowledge base*, a repository of knowledge about some domain, and an *inference engine*, which makes inferences using the knowledge and based on the conditions which prevail. Expert systems often also possess a user interface and an explanation facility (Fu, 1994). Feigenbaum (1977) points out that the power of an ES derives from the knowledge base rather than the inference method employed. Therefore, since they rely heavily on domain knowledge, they are by nature very domain specific. Expert Systems are designed to imitate experts, often through the employment of *rules* of the form *if...then*, and thus attempt to encode how an expert would react in certain situations. Expert systems have found many applications, for example in systems such as MYCIN and DENDRAL (Buchanan & Shortliffe, 1984), and their numbers continue to grow.

Morris (1990) reported on two surveys into the use of ES in the Library and Information Science (LIS) environment and discussed numerous research projects tackling the various library functions, including IR. PLEXUS (Vickery & Brooks, 1987) is a limited-domain system for answering queries about horticulture and gardening. Although this project did not get further than the research stage, it is worth mentioning as its focus is concerned with search formulation and it is an example of an early attempt to satisy user information needs intelligently, which is one way to view the process of intelligent IR.

Werckert and Cooper (1989) in their review of ES in libraries concluded that most ES research in libraries was concerned with IR. They describe an attempt called Answerman

(Waters, 1986) which was developed to retrieve information from library reference books. The system used a series of increasingly specific menus to refine the query until the required information was found, not unlike the old Internet Gopher system (Gopher, 1995).

The National Library of Medicine's thesaurus projects are probably the largest-scale effort that uses the knowledge in existing thesauri. In one of the projects, Rada and Martin (Martin & Rada, 1987; Rada *et. al*, 1989) conducted experiments for the automatic addition of concepts to MeSH (Medical Subject Headings) by including the CMIT (Current Medical Information and Terminology) and SNOMED (Systematized Nomenclature of Medicine) thesauri. Access to various sets of documents can be facilitated by using thesauri and the connections that are made among thesauri. The Unified Medical Language System (UMLS) project is a long-term effort to build an intelligent automated system that understands biomedical terms and their interrelationships and uses this understanding to help users retrieve and organize information from machine-readable sources (Humphreys, 1989; Lindberg & Humphreys, 1990; McCray & Hole, 1990). The UMLS includes a metathesaurus, a semantic network, and an information sources map. The difficulty of applying these approaches to an information system like the WWW lies in preventing the semantic network from becoming obsolete and keeping the thesauri up to date.

Cansearch (Pollit, 1987) was an ES designed to interact with the large on-line database MEDLINE, specifically in the area of cancer therapy, but it did not go further than the research prototype stage. More recently there has been ERSE - An Expert Retrieval System for Electronic Databases (Shoval *et al.*, 1990) - designed to support engineering professionals in formulating proper queries for submittal to electronic databases, and DIALECT 2 - a multi-experts system for information retrieval (Bassano *et al.*, 1992). This system is worthy of note as it employed a number of ES working together (using the 'blackboard' system) and a natural language interface.

The electronic and information revolution of the last 15-20 years has impacted upon the LIS field in numerous ways. Although much research and development has been conducted in producing usable and efficient interfaces to these new sources and platforms, these interfaces are not 'intelligent' in the sense of being able to act independently like an expert user. The explosion of networked information sources such as the WWW has focused the need for such 'intelligent' interfaces. The impact of the WWW has had considerable effect and there have already been huge changes to how information is made available to the prospective library user (see section 2). It has led in part to the development of digital, virtual and electronic libraries over the past four years (Collier, 1993; DMU, 1997). In turn there has been research into improving IR from these 'libraries'. Srihari *et. al* (1994) describe a system designed to allow intelligent data retrieval from raster images of documents which are scanned or undergo the process of optical character recognition (OCR). These images are stored electronically and made available through a client-server architecture. The IR module is only briefly described and no evaluation of the proposed system is reported. Salampasis *et. al* (1996) consider a digital library from a computing perspective to be a distributed information system. Taking advantage of the pre-eminence of the WWW, they describe an open agent-based (see below for discussion of agents) hypermedia model for distributed digital libraries. The paper describes a formalism only and there is no evaluation of any prototype built for the purposes of IR. However, in his PhD thesis (Salampasis, 1999), Salampasis does report a user-centered evaluation of a prototype system based on their model. This prototype was designed to employ agents to test a collection fusion strategy (the fusing of IR results from different and distributed sources) and the evaluation found that information seekers using parallel searching strategies are more effective than those who employ single strategies.

3.4 AI and IR in general

Different AI approaches to IR have been attempted - some involving a combination of existing techniques, others finding a novel way to attack the problem.

3.4.1 Knowledge-based Systems, Expert Systems and Semantic Networks

There have been many attempts to capture information specialists' domain knowledge, search strategies, and query refinement heuristics in information retrieval systems design. For example, CoalSORT (Monarch & Carbonell, 1987), a knowledge-based system, facilitates the use of bibliographic databases in coal technology. A semantic network, representing an expert's domain knowledge, embodies the system's intelligence. GRANT , developed by Cohen and Kjeldsen (1987), is an expert system for finding sources of funding for possible research proposals. Its search method - constrained spreading activation in a semantic network - makes inferences about the goals of the user and thus finds information that the user did not explicitly request but that is likely to be useful. Fox's CODER system (Fox, 1987) consists of a thesaurus that was generated from the Handbook of Artificial Intelligence and Collin's Dictionary.

RADA (Tan *et al.*, 1990) - A Research and Development Advisor incorporating artificial intelligence techniques and ES approaches - used a combination of AI techniques to address the IR problem. RADA is designed to have a user modelling component in the form of a KB. The main component is another domain KB which reflects the information contained in the document database. Although well designed, it appears that RADA was not actually tested upon a real document database. However, it does show the feasibility of employing user models in the IR process and using KBS to create these user models.

Celentano *et. al* (1996) have described the Kabria system, a system of office document retrieval based on the representation of semantic contents of the documents in KB. Obviously, their system relies heavily on knowledge of the structure and content of the documents which are acquired and classified by a separate system. At the time of the article evaluation was in progress but the authors believed the approach was extendable to more complex information systems such as the WWW.

The FIRE (Flexible Information Retrieval Environment) (Mizzaro, 1996) project is an attempt to design and develop intelligent interfaces to bibliographic databases. The FIRE prototype includes a KBS module which contains domain-specific knowledge. The prototype allows the user to reformulate a query if they are not satisfied with the initial IR results, by proposing a set of alternative or additional terms. This process is iterative and can continue until the user is satisfied. Evaluation of this prototype has produced some impressive results (Brajnik *et al.*, 1996) but the system is working with fixed databases.

Werckert and Cooper (1989) have noted that ES are suitable for designing systems to emulate an expert in a *narrow* domain - a fact highlighted by Cavanagh (1989) when discussing two KBS projects: POINTER (Smith, 1986) and PISCES (Cavanagh, 1989). ES are, by their nature, *brittle*, in that they do not work well outside their chosen domain - it is difficult to adapt them to general situations. Perhaps the ultimate failure of the systems described above can be attributed to this. In any case, all these systems were designed to be used with fixed, fairly static databases whereas a fundamental aspect of this research is the nature of the WWW itself. This nature - the fact that the WWW is a distributed, heterogeneous information store which is constantly changing - seems to preclude the use of approaches which rely on extensive knowledge about the database domain.

However, at the 5th WWW Conference (WWW5, 1996), a number of position papers were presented which explore the possibility of KB access to WWW information. Kirk (1996) describes a system called Information Manifold, a KB IR and information management system that allows interaction with the WWW in terms of knowledge about information. It is based around a system developed by Brachman *et al.* (1991). To work, information sources need to be added to the KB by being described. These descriptions have to be fairly detailed, including specification of semantic content and references to other documents. Kirk argues that as the KB grows, it becomes in effect a user profile which can be used to improve IR. The system still appears to be in the development stage and no evaluation is reported.

Barcaroli *et al* (1996) outline another KB approach to WWW IR using description logics (Woods, 1994). The idea is to represent knowledge by building hierarchical structures, sometimes called terminologies, where items are classified according to their properties. The basic elements are concepts and roles and in Barcaroli *et al.*'s approach they link WWW pages to these elements by regarding pages as an instance of a concept. The whole approach is one of building a KB from knowledge extracted from a 'training' set of pages.
No system has been implemented yet. Two other KB approaches described at the conferences are the MIHMA project (Hoppe, 1996) and Euzenat's (1996) proposal to use a KB as a WWW server. Despite the sophistication of these latest KB approaches to IR, Werkert and Cooper's remarks still apply. The hardest part for these KB WWW IR systems will be to build a sufficiently detailed KB for the subject domain and to ensure it is flexible and adaptable enough to meet the ever-changing nature of the WWW.

Recently, at the eighth WWW conference (WWW8, 1999), two KB approaches to information retrieval were presented. Froliich *et al.* (1999) describe *KBS-HYPERBOOK – An Open Hyperbook System for Education* in which they use a domain modelling language based on the techniques of object-orientation to create an open hypermedia information system. The concept is that the system would be part of a computer-supported learning environment. The domain model relies upon a semantics that includes rules and constraints. Important also within the domain model is the identification of domain objects and relationships between them. An example they give is a computer science course for a domain and a Java applet as an object. Actual links between pages must obey the constraints imposed by the model. A prototype has been built and tested but it is difficult to see how the concept would scale-up easily to cover the whole of the WWW though it may be a useful approach for small domains placed on the WWW as a set of pages.

Guan and Wong (1999) describe an approach to information extraction or mining which involves using an algorithm which employs keywords, patterns and samples of text to extract the desired information. Basically, a heuristic set of KB rules is applied to, for example, the keywords after they are identified in order to extract the information related to the keyword automatically. Experimental results to extract the name and email addresses of all academic Professors in an institute showed, according to the authors, that it is more efficient than other methods. However, they only considered 10 WWW sites in their experiments.

3.4.2 Case-based Reasoning

Another approach that has been applied to IR is case-based reasoning. In case-based reasoning a system adapts and retrieves old data to produce new data that solves the current need (Kolodner, 1993). In IR this has been applied, for example, to the retrieval of archived meteorological data (Jones & Roydhouse, 1995) in order to assist intelligently meteorologists in their work. As this approach relies on an extensive database of information so as to satisfy the user need, it is difficult to see how the approach could be applied to something which is ever changing like the WWW. It would require heavy processing resources in the shape of continuous indexing of new documents to fit the representation required by the system.

3.4.3 Inference and Bayesian Networks

Probability-based methods of handling uncertainty in AI are based on Wright's (1921) analysis of crop failure. They have since been reintroduced by a number of researchers under various names, such as 'causal network' and 'influence diagram' (Good, 1961; Howard & Matheson, 1981). Two of these methods that have been applied to IR are *inference networks* and *Bayesian networks.*

An inference network, introduced by Turtle and Croft (1990), is similar to other probability - based simpler methods, in that it follows a probability ranking principle. However, rather than ranking the documents by a calculated probability of relevance (given document and query), it ranks them based on the probability that a document satisfies the user's information need. In IR, an inference network consists of two component networks – a document network and a query network. The document network is a series of nodes that represent a collection of documents, it is developed once for a particular collection and does not change during the process of query satisfaction. The query network consists of a single node which represents a user's information need and one or more query representations which express that need. A query network is constructed for each need and can be modified during the query processing as the query is refined or additional representations are added in an effort to capture the better the information need. The document and query networks are joined by links between representation concepts and query concepts. All nodes in the inference network can be *false* or *true.*

The retrieval inference network is intended to capture all of the significant probabilistic dependencies amongst the information represented by the nodes in the document and query networks. Given the prior probabilities associated with the documents, the posterior probability or belief associated with each node of the network can be calculated. The network, taken as a whole, represents the dependence on the belief that a user's information need is met by the documents in a certain collection, where that dependence is mediated by document and query representations. When the query network is first constructed and attached to the document network, the belief associated with each node in the query network is computed.

The initial value at the network node representing the information need is the probability that the need is met given that no specific document in the collection has been observed and all documents are equally able (or not able) to satisfy the need. If a single document, d_i is then observed and evidence to the effect that d_i is true is attached to the network, then a new belief for each node of the network can be computed given this evidence. That is, the probability that the information need is met given that d_i has been observed can be found. Then, removing this evidence and asserting instead that some other document, d_j, $i \neq j$, has been observed, the process can be repeated for each document in the collection. Thus, the probability that the information need is met given each document in the collection can be found and the documents can be ranked accordingly.

Haines and Croft (1993) have proposed the incorporation of relevance feedback to improve the performance of inference networks and have found significant increases overall in their experiments. With or without relevance feedback, inference networks are trained and constructed using sample sets of documents, and existing studies have been carried out on fixed static databases, meaning the approach is unlikely to be suitable for improving IR from the WWW. However, recently Lu *et al.* (1996) have attempted to apply inference networks to multiple collection searching. They argue that IR from the WWW is a matter of IR from multiple collections. They demonstrate the effectiveness of their approach in experiments with the INQUERY IR system (Broglio *et al.*, 1996) and the TREC sets of documents (TREC, 1996). However, the TREC sets, though large, are nowhere near as large WWW, and the approach still relies on building inference networks for documents and thus on knowledge about those documents. Also, Lu *et. al* themselves question the efficiency of their approach in dealing with information residing on wide-area networks.

The development of Bayesian networks (Fung & Del Favero, 1995) has given a boost to probabiltiy-based AI techniques. Savoy (1993) describes a project which tackled the searching for information in hypertext systems using multiple sources of evidence (Savoy, 1993) and employed Bayesian Networks. A Bayesian Network is an annotated directed graph that rigorously describes probabilistic relationships between variables in a domain (e.g. in IR, documents and queries) but also includes a qualitative structure that facilitates communication between system and user (Heckerman *et al.*, 1995). Fung and Del Favereo describe a fundamental approach to the use of Bayesian networks in IR. Their approach supposes that a user specifies one or more topics of interest and identifies some features of documents that can be used as evidence for the topics of interest. The three main steps involved in applying Bayesian networks to the IR process are then: (a) build the network representing the query, (b) score each document by extracting the relevant features from the document, instantiating the features in the network and calculating the posterior probability of relevance, and (c) rank the documents according to the probability values. A simple Bayesian IR network is shown in figure 3.1.

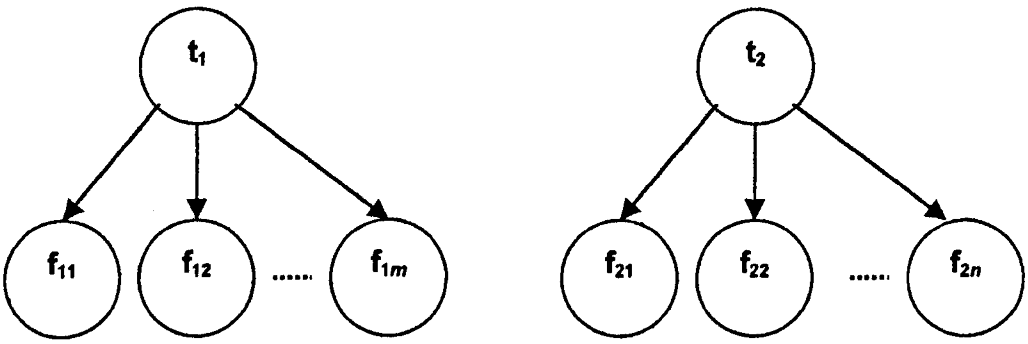


Figure 3.1 A Two-level Bayesian Network Model of Information Retrieval.

In figure 3.1 there are one or more topics of interest (t_i) and one or more document features to examine (f_{ij}). The feature sets for the topics are shown as being different though this is not required to be true. The nodes t_i represent the events "the document is relevant to topic t_i" and the nodes f_{ij} represent the events "the feature f_{ij} is present in the document. Given two probabilities, which need to be specified - the prior probability that the document is relevant to topic t_i and the conditional probability that feature f_{ik} is present in a document, given that the document is relevant to topic t_i - then a posterior probability, the probability that the document is relevant to t_i given the presence or absence of all the features f_{ij} for each document, can be calculated for each document. This can be used to rank the documents in a set and, although there are different approaches to building the initial Bayesian Network, this is the basic procedure involved in each case.

Savoy's system employed multiple indexing schemes to structure a document set, managed by using Bayesian networks. Different search policies were employed and favourable results produced, especially when relevance feedback techniques were also applied (Salton & Buckley, 1990). However, as is the case with inference networks, since the Bayesian trees were built from a fixed hypertext it is doubtful if this approach would prove useful in tackling the problems of IR from an ever-changing, distributed hypertext like the WWW.

However, Bayesian networks have been employed as well in the INQUERY system, developed as part of the TIPSTER project at the Centre for Intelligent Information Retrieval (CIIR, 1997). The approach involves changing the representation of text and information needs through the development of Bayesian nets for each. It also requires the parsing and indexing of retrieval documents which may limit its effectiveness in dealing with the WWW. The authors report evaluations which show efficient and effective IR when tested against fixed, if very large, databases. A new project is looking to extend the INQUERY approach to aid the user to browse and search large distributed databases. Again the technique involves building Bayesian networks from examining documents (Croft, 1997).

3.4.4 Relationship lattices

Pedersen (1993) reported on the application of a branch of mathematical lattice theory, called relationship lattices, to bibliographic IR. Essentially, the BRAQUE (Browse and Query Environment) project sought to hide the user from the complexities of database structure and query languages. Relationship lattices are used to model a convenient GUI and an extensible thesaurus. Commercial versions have been developed (*Satellites International Ltd*, 1995).

3.4.5 Neural Networks

There have been numerous attempts to use Neural Networks (NN) to address the problems of IR and some of the approaches are outlined below. This a branch of AI which has its foundations in both the early days of computing and neurobiology. It adopts a metaphor which suggests that intelligence *emerges* through a large number of elements connected together. Each element, though, is performing a simple computation. The knowledge of a network is encoded as a set of weights on connections between elements or nodes (Haykin, 1994; Fu, 1994). An artificial neural network mimics a simplistic view of a biological network of neurones and research progress in the field of neurobiology has allowed the mathematical and computational modelling of NN (see figure 3.2).

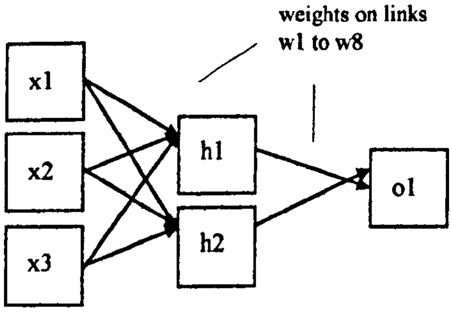


Figure 3.2 A Simple Neural Network showing three input nodes (x1 - x3) and a required output (o1). The second layer is called the hidden layer, used to train the net. The links between the nodes have a set of weights (w) which are adjusted during the training.

The idea dates back to the early 1940s, with the perceptron introduced by McCulloch and Pitts (1943), and Hebb (1949) proposing a learning law for NN. However, by 1969, the field had become dormant, owing to limitations pointed out by Minsky and Papert (1969). The major limitation voiced by Minksy and Papert was the inability of the early neural networks to satisfactorily compute a simple function, the exclusive-or (XOR) problem. The NN approach was resurrected in the 1980s, notably by Hopfield (1982) and Rumelhart *et al.*(1986) who developed an approach which dealt with the early limitations. Neural networks originally found favour with their ability to recognise, classify, convert and learn patterns, but have also been applied to other fields, such as IR.

The work of Belew is probably the earliest example of a NN model adopted in IR. In his system (Belew, 1989) he developed a three-layer NN of authors, index terms and documents. The system used relevance feedback from its users to change its representation of authors, index terms and documents over time, the result being a representation of the consensual meaning of keywords and documents shared by a group of users.

Doszkocs (1991), in his discussion of potential library applications of NN, notes that this ability to learn is crucial. Neural networks lend themselves to the notion that intelligent behaviour is emerging and that NN could be applied successfully to achieve intelligent IR from dynamic sources such as information networks (Doszkocs *et al.*, 1990).

Iwadera and Kimoto (1992) explored the effects of using a Dynamic Word Network (DWN) on IR. This network was being used, at the time of the article, to develop a system called AIRS (Associated Information Retrieval System). The authors found that, in comparison with other more conventional methods, this DWN approach improved the precision of IR. Kwok (1991; 1995) reports on a project which looks at extending a NN approach for probabilistic IR to allow the possibility of automatic query expansion. Kwok found improved performance but again this system was evaluated against fixed document sets about which a lot was known.

Gallant *et al.*(1992) report on a system called *Matchplus* whose goals include high precision and recall, ease of use and incorporation of machine-learning algorithms. High dimensional (~ 300) vectors, called context vectors, represent word stems, documents and queries in the same vector space. This allows one type of NN algorithm to generate the stem context vectors and a more standard NN algorithm performs routing and automatic query modification based on user feedback. Queries can take the form of terms, part of or full documents and conventional Boolean expressions. Initial experiments with MatchPlus showed a 3-15 % improvement over more traditional methods of IR. However, again this was on fixed databases. Problems with this approach for the present research would include: how are unknown documents to be incorporated into the system and, with an ever-changing database, what are the computational difficulties in constructing the context vectors.

Scholtes (1992) work has had bearing on the research reported in this thesis. He reports on an implemented Kohonen NN method for free-text information filtering and retrieval. The method is compared with some classical statistical IR algorithms and found to improve on performance.

His NN approach adopts a statistical *n*-gram vector method. Basically, an *n*-gram is an *n*-length sequence of characters occurring in a word. For example, the *trigrams* (*n*=3) in the word 'neural' are: --n, -ne, neu, eur, ura, ral, al-, l-- (where '-' is a space). In such a way documents can be uniquely identified. N-gram vectors are very powerful, easily manipulated, self-learning and language independent. For the purposes of training the Kohonen net, the n-gram method is interpreted as a window of size *n* moving over the words. By shifting this window over the learning text (query) only frequent n-grams form clusters in the Kohonen net. A coding procedure converts the n-grams into a suitable format for Kohonen learning. Once the training is done, the IR filtering is a matter of matching the database information with what is represented on the Kohonen net or feature map. One problem is anomalies emerging in the feature map. Scholtes (1992) points to a conceptual problem in that this approach lacks the ability to incorporate real meaning. However, in tests and comparisons with statistical IR algorithms, this NN approach performed well, and because it is a method for performing IR filtering on a dynamic database, it did form the basis of a further evaluation of NN (see section 3.5).

Also interesting in the light of the present research is Jennings *et al.* (1992) use of a User Model NN to develop a personal news service. Here a user model of preferences is constructed based on articles read and rejected. The NN adapts over time to represent better the user's interests and once in place can rank articles according to this model. User modelling and IR in general is examined in chapter 4.

In tests against selected newsgroups of the networked USENET, this system performed well. One advantage of this approach is that it is applicable to dynamic networked databases such as the corpus represented by USENET. It avoids the problems of trying to model semantically a whole information store - which is impossible with something like the WWW. However, although it can provide for changes to the user's preferences, Jennings *et al.*

acknowledge that it is difficult to do - there is no way to directly destroy nodes and links, only a general decay that occurs if the user continually rejects articles that contain the words they represent. Another problem is the computational cost involved with producing the user model NN at the outset. This paper and that of Scholtes (1992) do show, though, that NN is one AI technique that can address the problems of IR from dynamic sources such as the WWW.

3.4.6 Genetic Algorithms

There has been little research into the application of Genetic Algorithms to IR. During the past decade there has been growing interest in the use of algorithms which rely on analogies to natural processes and Darwinian theory of evolution. Genetic algorithms are based on these principles (Shaffer, 1997) and in such an algorithm a population of individuals (potential solutions) undergoes a sequence of transformations such as mutation and cross-over. The individuals strive for survival and there is a selection process, biased towards selecting fitter individuals, which produces individuals for the next generation. After an iteration of generations, the algorithm should converge and produce the best individual as the optimum solution. Genetic algorithms are part of the relatively new field of Artifical Life which attempts to study and understand biological life by synthesizing artificial forms. Langton (1995) provides an overview of the research questions studied within the field.

Gordon (1988) presented a GA-based approach for document indexing in which competing document descriptions (keywords) are associated with a document and altered over time by using the usual genetic mutation and crossover operators. It appears the work did not get much further than initial research. The GANNET system (GANNET, 1993) presents a hybrid NN/GA approach to IR. The approach is to use Gordon's GA methods to manipulate a set of input documents and associated keywords in order to generate an optimized set of initial keywords.

This set is then used by a Hopfield NN to produce other relevant keywords (in a form of spreading activation). These are then included in GAs for the next round of optimization and the process repeats itself until there is no overall improvement in fitness. Though the authors report this improvement in what they call the search space and note that they believe the approach could be used to develop adaptive IR, it appears that no testing against real databases has been performed.

Maes (1995) has argued the need for a more interdisclipinary approach to the development of software agents (see section 3.4.9) in the entertainment areas and explored the prospect of combining models developed in artificial life with existing AI approaches.

3.4.7 Fuzzy Logic

Attempts to apply the branch of AI known as Fuzzy Logic (FL) to IR are not new (Radecki, 1976; Hosono *et al.*, 1985) but have recently gained increased currency along with the growth of FL applications generally (Mansfield Jr & Fleischman, 1993; Yager & Larsen, 1993; Zadeh, 1993). Zadeh (1965) introduced the idea of fuzzy sets which challenges traditional reliance on two-valued logic and classical set theory as a basis of scientific endeavour. Unlike in classical set theory, each element is not fixed as either being or not being a member of the set. Rather, each member is given a 'strength' or 'grade' of membership, represented by a value between 0 and 1. The two boundary values represent the normal crisp grades of a classical set, i.e. 0 implying 'not a member', 1 'definitely a member' (Dubois & Prade, 1980). A general example of a fuzzy set is given in figure 3.3 (adapted from Klir & Folger (1988)).

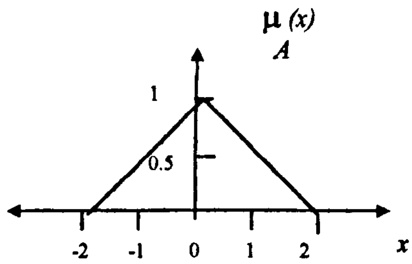


Figure 3.3 A Possible Membership Function of the Fuzzy Set of Real Numbers Close to Zero.

Consider that X denotes a universal set. Then the membership function by which a fuzzy set A is usually defined has the form: μ_A:$X \rightarrow [0,1]$ where [0,1] denotes the interval of real numbers from 0 to 1, inclusive. Figure 3.3 shows a graph of a possible membership function for the fuzzy set of real numbers close to zero, the function being:

$$\begin{aligned} \mu_A(x) &= 0 && x \leq -2, x \geq 2 \\ &= (x+2)/2 && -2 < x \leq 0 \\ &= (2-x)/2 && 0 \leq x < 2. \end{aligned}$$

This is known as the *triangular* type of membership function and is actually a special case of the type of function known as *trapezoidal*, which is used quite often in fuzzy logic research. Generally, the trapezoidal function can be described by the function that follows figure 3.4.

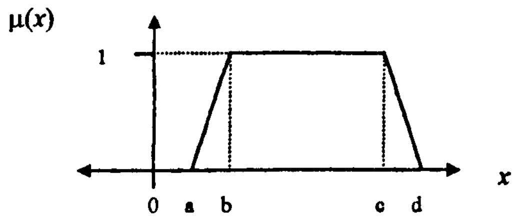


Figure 3.4 The General Form of the Trapezoidal Membership Function.

The function, $\mu(x)$, is:

$\mu(x)$	=	0	$x \leq a$
	=	$(x-a)/(b-a)$	$a < x \leq b$
	=	1	$b < x \leq c$
	=	$(d-x)/(d-c)$	$c < x \leq d$.

The concept of fuzzy logic has developed into a way to model human-centred systems (Zadeh, 1973) as 'fuzziness' seems to pervade most human perception and reasoning, including the process of IR (Nakamura & Iwai, 1982). Miyamoto (1989) developed an approach to IR using *fuzzy associations*. A fuzzy association is explained as a fuzzy relation defined on a set of indices in a database. This work was a generalisation of early work on developing a fuzzy thesaurus for IR (Miyamoto & Nakayama, 1986). An algorithm for fuzzy IR was developed but does not seem to have been tested on a real database.

Terano *et al.* (1992) devote a chapter of their book to the subject. In this they postulate using fuzzy sets to represent what they term *a priori* knowledge spaces, building on earlier work by Nakamura and Iwai (1982). An *a priori* knowledge space is defined as a structured keyword space, that is a structured space showing the affinity of keywords that describe the contents of one or many documents (or the *concept* of the document). In such a way, a vague query can be transformed into a request concept represented by a fuzzy set, and the system can infer the degree of inclusion in this set of another keyword by its distance in the *a priori* knowledge space. Terano *et al.* constructed a prototype document retrieval system to test their ideas, in which the knowledge space was defined from 812 keywords, with 5085 keyword and document pairs. However, the system was not tested in a real situation, rather the request concept was fixed as the union set of several documents. Recognition of the request concept, and hence the decision on which documents are retrieved, is achieved by finding the overlap between the request concept space and the knowledge space. Terano *et al.*'s approach would obviously run into problems with IR from an effectively unbounded (and definitely unknown) knowledge space like the WWW. Mansfield Jr and Fleischman (1993) describe a system which allows queries with fuzzy predicates to be generated for large but fixed databases.

Yager and Larsen's work has much more relevance for this research (Yager & Larsen, 1993). They developed the basic structure of an intelligent inquiry system in which a user's query is fuzzified and generalised into a fuzzified query. This fuzzified query is then given what they term a crisp envelope to allow its use in IR. To describe their approach they employ the example of a user consdering a house purchase by querying a database of information regarding houses for sale. In the query, desired house characteristics are expressed, e.g. price. These characteristics are then generalised with the aid of fuzzy sets. The concept involves considering a characteristic, such as price, not as a *crisp* or discrete value but as a range of values which are acceptable to the user to some degree - the range can then be fuzzified by representing it as a fuzzy set. Obviously some requirements cannot be generalised in this way, for example the desire for the house to have a fireplace. The fuzzy set is then used to extend the boundary of the desired characteristics by providing a *crisp* envelope of a possible range of values that will satisfy each characteristic, e.g. 120K to 150K say for a house price originally expressed as 140K. Thus, in this way the query itself can be extended to allow more database records initially to satisfy it. Also, it allows the use of an ordinary database querying language to be used to perform the search. The extended set of query results can also be ranked with the help of the fuzzy set. The authors discuss the use of mathematical operators for justifying the creation of the crisp envelope from the fuzzy representation.

Although they seem to have not tested their approach on a real database, the idea of capturing a user's information need and representing it by a fuzzy set is one approach that merited further evaluation. This is discussed further below.

3.4.8 AI and Hypertext

IRHIS (Zegler *et al.*, 1992) - Intelligent, adaptive information Retrieval system as Hospital Information System front end - was a prototype project, employing hypertext and KBS. Although never tested 'live', the system did show promising results and was well received by potential users. This interesting use of hypertext slightly predates the development of the World-Wide Web (WWW).

Lucarella and Zanzi (1993), in another approach to IR from hypertext, employ what they call 'plausible inference'. The use of fuzzy logic (FL) allows inference from uncertain imprecise premises, thus matching well the inherent imprecision in any IR process. Recently, Bell (1996) has proposed an adaptive hypermedia navigation system called HyBIS which relies on machine learning techniques to 'learn' what the user's information need is. The system relies heavily on the quality of the meta-information and thus its extension to a hypermedia of the proportions of the WWW seems unlikely, unless the recent research covering hypertextual meta-information (Heery, 1996) leads to the widespread adoption of a standard. Another recent development in hypermedia, XML (eXtensible Markup Language), promises to provide a way to produce documents on the WWW which are richer in terms of the types of information they contain (XML, 1999). It provides for the creation of hypertext documents with unique tags which can then be interpreted according to a document model. There is evidence, suggested by the number of papers and posters given at the recent 8th WWW Conference involving XML (WWW8, 1999), that XML is already being well used in new research.

The development of the hypermedia known as the WWW has led to researchers trying to find ways to improve IR from it. Some of these fall into the category of intelligent agents and are discussed below. Others cannot be called agent approaches. One is Rice's (1997) proposal to use thematics to allow concept and context filtering. This unique approach involves the use of profiles of selected attributes to filter WWW pages. The argument is that the use of Special Interest Group - oriented thematic vocabularies or 'pidgins' - will allow conceptual filtering during IR from the WWW. Basically these are mini-thesauri, tailored to meet the needs of a specific subject domain. The work is in progress and it remains to be seen if any concrete improvement over existing WWW search tools can be achieved with this approach.

3.4.9 DAI and the Notion of Agency

The research focus of this thesis is not concerned with the development of a *new* search engine or tool which would act 'intelligently'. Within the field of Distributed Artificial Intelligence (DAI) and computing in general there has been much research and development into the notion of *agency* and intelligence - the goal being to produce intelligent autonomous agents which will perform tasks proactively (Wooldridge & Jennings, 1995). The debate not only rages about what an agent is but also over agent formalisms, architectures and programming languages (Agentlist, 1997; Agentweb, 1998; Genesereth & Ketchpel, 1994; Huibers & Van Linder, 1996). Basically, a software agent is piece of code designed to act autonomously in order to achieve some task. This is one definition and the term is discussed in greater detail in Wooldridge and Jennings (1995). The debate has not prevented agent technology being applied in various areas, including intelligent document retrieval (Mukhophadyay *et al*, 1986), information filtering (Maes, 1994a) and co-operative IR from diverse information sources (Voorhees, 1994).

Kahn and Cerf (1988) proposed an architecture for a set of information retrieval agents called Knowbots (Knowledge Robots). Knowbots were agents designed to search a variety of information sources to satisfy a user's query - the goal being to let the knowbot do the navigation instead of the user.

Maes (1994a) has used this concept to develop prototype applications such as the NewT system designed to filter USENET news articles. She applies machine-learning techniques in which the agent is trained by being given examples which illustrate the types of articles the user would and would not read. The agent then begins to make suggestions to the user and is further trained by feedback on its choices from the user. One is bound to ask how flexible the agent is at adapting to a user changing their choice profile quickly, as may be the case when accessing less structured documents on the WWW.

Etzioni *et. al* (1994) think of agents as software robots or 'softbots'. Here, AI planning techniques are used to allow agents to satisfy user goals, for example discovering whether a person is logged on. The agent works in the background, taking a logical expression of the user's goal as input and using scripting programs to search networks to satisfy that goal. Building on this work, Shakes *et al.* (1996) have developed a technique termed Dynamic Reference Shifting in a system called Ahoy! which currently searches for a Web homepage given the name and institution of a person. This system is in fact an information filtering tool, filtering the output of several WWW indices to extract likely matches. If there are no matches then the system uses heuristics to offer a partial match. This heuristic filter is developed via a machine-learning technique whereby information is extracted from the URLs of successful previous searches. This technique differs from the machine-learning approaches in other systems which employ learning about users tastes and preferences to improve IR (Armstrong *et al.*, 1995; Lieberman, 1996; Pazzani *et al.*, 1995).

Evaluation of Ahoy! showed that it fares well against existing search tools such as Alta Vista (1998a) or Yahoo! (1998). The authors believe that Ahoy! is applicable to other subject domains but whether the approach is generalisable to the whole WWW remains to be seen.

IBM is developing WWW agents for IR as well. Intelligent Miner is an agent employed as part of an electronic shopping service that uses NN to perform pattern analysis on large amounts of data (Tkach, 1998). It employs the AI technique to allow a user to recall where they have been whilst browsing on the WWW. However, as yet, there seems to have been little evaluation of these tools to judge the effectiveness of their IR capabilities.

Finally, an interesting agent project is InfoSleuth (Bayardo *et al.*, 1997) which is examining the semantic integration of information from different sources on the WWW. The architecture consists of a number of agents working cooperatively, including user and ontology agents, which use the AI semantic languages KQML (Finin, 1995) and KIF (Genesereth & Fikes, 1994) to communicate. The user agent, which *de facto* is the user's intelligent interface to the system, assists the user in formulating queries and displays results tailored to the user's view by employing knowledge obtained from domain models. It can be persistent and autonomous. InfoSleuth has been implemented for one domain, health care management, and though the authors do not report any evaluation, they are attempting to expand the system to cover other domains.

3.5 AI Techniques and this Research

From the examination of previous research, discussed above, it is apparent that attempts to incorporate intelligence into the process of IR are numerous. They have met with limited success, and involve many facets of AI. Owing to the limited success of AI techniques based on machine-learning principles and the inapplicability of using techniques which rely on extensive information about the retrieval database domain, the present research changed its focus to applying an AI technique to the user's query. The argument for doing this is that by attempting to refine intelligently the query before it is presented to the WWW obviates the problems encountered through the distributed, unstructured, ever-changing nature of the network. In the light of this change, research into techniques of query refinement was needed and this research is reported in section 3.6. First, in this section, the evaluation of AI techniques in the present research is discussed.

The emphasis on intelligent query reformulation was a major factor during the evaluation of AI techniques. It may be thought that, from this emphasis, any technique could be considered. However, any approach or use of AI to reform queries which relies on extensive domain knowledge would still be unsuitable because of the ever-changing diverse nature of the WWW. This was another factor that was considered during the evaluation. Initially, this evaluation examined expert systems and semantic networks before concentrating on the two approaches which previous research had shown to be most promising, neural networks and fuzzy logic.

Expert systems work well with tightly defined and small domains (Cavanagh, 1989), and require knowledge of the domain to function. The user query may be vague and unstructured and thus there may be little 'knowledge' to provide the ES with. Although the user query might be extended and refined through a set of ES rules, and rules could be added and amended to reflect changes to the query concept, ES are 'brittle'. It would be difficult to build an ES of sufficient flexibility to adapt to a changing user query concept, unless some sort of automatic amendment of the rule base could take place.

Semantic networks can be developed to represent queries as conceptual relationships between words (Tan *et al.*, 1990). One problem with this would be how to limit the growth of the network in order to prevent a simple user query becoming too complex and difficult to process computationally and temporally. Another problem would be any radical changes to the user query concept would be slowly reflected in the network's links.

Neural networks offer an approach which showed promise with regard to this research. They work best when lots of data are used to train them and when trained they can be used to recognise patterns and links between patterns in data. Therefore, they can be trained to recognise words in a query, and conceptual relationships between words. Neural networks can be self-organising and self-learning, are easily extended and adaptable. As already noted it is possible to use NN for IR and information filtering from the Internet without extensive knowledge of the information store (Jennings *et al.*, 1993), and their self-organising nature and ability to undergo unsupervised training means that the IR process could be to some extent automated to reflect the changing user needs and WWW changes as well.

An initial evaluation of the use of NN was undertaken. Based on the *n*-gram approach used by Scholtes (1992) which was discussed earlier, the concept was to develop trigrams ($n = 3$) for a series of words (about or related to one topic) and then, using a simple algorithm, transform these trigram representations into a dataset suitable for use to train a Kohonen NN.

A Kohonen NN, or self-organising map, is based on unsupervised learning and was chosen also because it constructs topology-preserving mappings of the training data where the location of a node carries semantic information. In other words it is ideal for representing words as clusters of data. Not only does the Kohonen net produce clustering of data, but there is a spatial ordering of the map so similar input patterns (datasets) tend to produce a response in nodes that are close to each other. For IR purposes, this means that Kohonen NN could be trained to recognise words as belonging to a concept represented itself by some cluster of data on the map. In this way, the nodes activated by any input could represent the refined query, and provide the basis of turning a user's vague query into a more focused and semantically rich one.

A set of words related to the field of AI were chosen as a test set (see figure 3.5), trigrams created for each word and these then transformed into suitable datasets for use in developing a Kohonen NN. The network was developed using software created at the University of Stuttgart (Zell *et al.*, 1994). The software's main components - a NN simulator kernel and a sophisticated user interface - provide the ability to create, manipulate and visualise highly complex NN in various ways.

The transforming algorithm was a simple one, i.e. a=1, b=2, c=3,..., z=26, producing datasets of the kind shown in figure 3.6, which also shows the corresponding trigram.

Words	
Artificial	Neural
Intelligence	Fuzzy
Networks	Logic
Genetic	Connectionism
Systems	Knowledge –based
Algorithms	Expert

Figure 3.5: The Sample Words Used in the Trigram Approach.

Word	Trigram	Dataset
Neural	*- - n*	0 0 14
	- n e	0 14 5
	n e u	14 5 21
	e u r	5 21 18
	u r a	21 18 1
	r a l	18 1 12
	a l -	1 12 0
	l - -	12 0 0

Figure 3.6: A Word Transformed into a Trigram and then a Suitable Dataset.

Unfortunately, the results showed that the network failed to converge to a stable state and no clustering occurred. This may be due to a lack of data used in the training, as NN require large amounts of data to train well, or it may be due to the transformation algorithm producing datasets which contain elements that prevent successful training, or some other cause. Despite this setback, in principle the approach still has currency, but to revisit the approach there would be a need to generate large amounts of data quickly and accurately through the use of word transformation programs.

However, another approach with equal if not greater promise is Fuzzy Logic (FL). It is the ability to encapsulate the linguistic and semantic 'vagueness' of human reasoning and conceptualising that makes FL such an attractive proposition for solving the problems of IR. The idea of applying it to this research was ignited by work discussed earlier (Terano *et al.*

1992; Yager & Larsen, 1993). The notion here is to recognise and attempt to capture the user's information need and represent it as a *fuzzy set* as described earlier. Fuzzy sets adapt easily and thus reflect changes in what they represent. In this evaluation of fuzzy logic it was recognised that it could be employed to represent knowledge about the user's experience captured through an on-line interaction. This knowledge may be represented as flexible fuzzy sets, through the development of fuzzy rules in a fuzzy knowledge base. These rules are much more flexible than those in traditional knowledge bases as they allow for degrees of correctness and for multiple and simultaneous rule execution.

The strength of membership of each set and thus the shape and profile of the fuzzy set would be determined and influenced by the on-line interaction. The information held in the fuzzy sets can be used to adapt a user's query and thus refine it intelligently.

For the purposes of information search and retrieval, though, the fuzzy sets will have to be 'defuzzified' in order to provide suitable search terms for use with an existing WWW search engine. This is difficult to do effectively without losing some of the information contained within the sets. Despite this difficulty, the FL approach offered the best way to progress the aims of this research. A schematic diagram of how the FL approach could perform the required query refinement is shown in figure 3.7.

Obviously, figure 3.7 offers no details of how the query refinement step could be performed – that is left to chapter 5. However, before the query refinement technique could be chosen, an investigation of query refinement research was undertaken. This is reported next, in section 3.6.

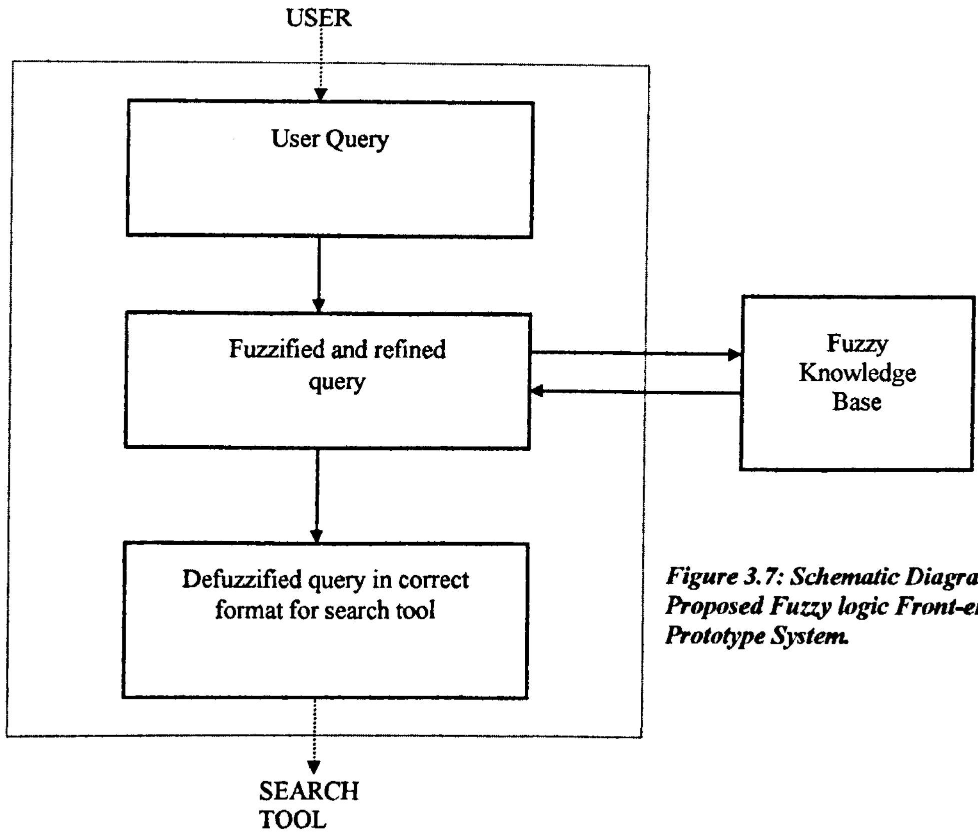


Figure 3.7: Schematic Diagram of a Proposed Fuzzy logic Front-end to the Prototype System.

3.6 Query Refinement in IR

Query refinement in IR systems to produce improved results is an active research area, in which different methods are being applied (Arens *et al.*, 1996; Bell, 1992; Cusack, 1991; Han, 1994; Litteck, 1992; Sun & Yu, 1994; Vielle, 1989). Arens *et al* (1996) describe a query reformulation approach within a project designed to provide intelligent access to heterogeneous databases. In the project, a domain model of an information domain is created which establishes a fixed vocabulary for describing datasets in the domain. Each information source or database is then described using this language. Queries to the system are posed using terms from the domain model and reformulation operators are employed to select dynamically an appropriate set of information sources and to determine their integration to satisfy the query. The system has shown some success in retrieving information from different databases, including those dealing with trauma care and transportation, but, as the approach relies on not only modelling the query domain but also the information source domain, it is inapplicable to the WWW which is semi-structured and constantly changing. The system also includes a component which employs semantics to optimize queries to the modelled databases (Hsu & Knoblock, 1995) but this also exploits knowledge about the contents of the databases and therefore is also inapplicable to the present research.

Research into the use of lexical-semantic relations. such as thesauri, to expand and refine queries has been performed for many years and is more relevant to the present research. Over twenty years ago, Salton and Lisk (1971) found that IR performance could be improved by using synonyms to refine queries. Wang and his colleagues found that a variety of lexical-semantic relations improved retrieval performance (Wang *et al*, 1985), but both of these conclusions were drawn from experiments on a very small collection of documents.

Vorhees (1994) reports on an examination of the utility of query expansion (a form of refinement) through lexical-semantic relations in a large collection that spans several information domains. The queries are expanded using relations encoded in WordNet (Miller, 1990), a large general-purpose lexical system built at Princeton University, and are run against the TREC (Text REtrieval Conference) collection (Harman, 1992). WordNet is a manually constructed lexical system whose basic object is a set of strict synonyms, called a *synset*. Synsets are organised by the lexical relations defined on them, which differ according to the relevant part of speech. The TREC collection is a large set of documents (over a million) which consist of English prose obtained from a variety of sources including newspapers and technical abstracts.

The text of a TREC query, known as a *topic statement*, is a complex natural-language statement of information need. Vorhees' approach is to add a new field to the topic statement which consists of a list of manually-selected WordNet synsets which contain nouns germane to the topic. To perform the IR process both topics and documents are indexed automatically using the SMART routines (Salton, 1971). In indexing the amended topic statements, the query is expanded when the synset addition is reached. The IR process becomes a matter of matching the expanded query terms to the document terms.

This indexing of both document and query means the detailed approach of Vorhees is inapplicable to the present research. This is because the focus here is on query refinement before submission to an existing WWW search engine and so there is no control or influence over the indexing method applied by the search engine. Also, it is doubtful that attempting a similar indexing strategy with regard to the WWW would work, as the TREC database, though large, is nowhere as near as extensive or as volatile as the WWW information corpus. Moreover, the TREC documents are much more structured than those prevalent on the WWW.

However, the approach of refining queries through recourse to lexical-semantic relations is promising. Vorhees notes from her experiments that, although little IR benefit is achieved when a user supplies a detailed query, the approach does have the potential to improve queries which are not detailed. Users, especially novice users, frequently do not supply a detailed query, and it has already been noted that it is on these kind of users that the research is focusing, as opposed to expert users who are much more likely to be able to form detailed queries. With this in mind, a query refinement approach which employs lexical-semantic relations was employed in the present research. The approach is discussed in detail in chapter 5 (page 77) but before that, this chapter finishes with conclusions and recommendations.

3.7 Conclusions and Recommendations

This section has presented an examination of intelligent IR, the use of AI to improve the results obtained during the process of IR. This examination showed that research in many areas of AI and IR has produced systems with notable, if limited, success. The main conclusion is that many of the AI approaches employed in general IR are inapplicable in this research owing to the nature of the WWW. Furthermore, it was concluded that to avoid to some extent the problems caused by this nature, and to effectively apply AI to IR from the WWW, the research should focus on query refinement. This conclusion led to an investigation of query refinement techniques which concluded that the best refinement approach would be one that used some type of lexical-semantic relations (Vorhees, 1994).

Two of the soft computing areas, fuzzy logic and neural networks, did exhibit promise as an AI technique to use, and both were further evaluated. This evaluation showed that fuzzy logic was the approach to adopt, and, more importantly, that it would be possible to use fuzzy logic to model the WWW user's experience and knowledge. The intention is to use this modelled information to refine the user's query intelligently.

This final conclusion led to recognition for a need, prior to development of any fuzzy modelling approach, to investigate user modelling and IR in general. This investigation is reported in the next section.

CHAPTER 4 - AI AND IR: THE ROLE OF USER MODELLING

4.1 Overview

This chapter reports an investigation of the application of user modelling in AI systems, specifically with regard to IR. This investigation was undertaken in order to identify which user modelling approach might be employed to improve IR from the WWW.

User modelling is a huge area of endeavour, with on-going research and development, not only in AI systems (Kok, 1991a; Lai, 1993; McTear, 1993), but also in on other areas of research , such as Human-Computer Interaction (HCI, 1998; Keates & Robinson, 1997). Before examining the application of user modelling to AI and IR, the underlying issues are discussed in the next section, 4.2. Section 4.3 looks briefly at user modelling and AI systems in general before section 4.4 focuses upon user modelling and intelligent IR. Bearing in mind the crux of the present research, section 4.5 narrows this focus to user modelling and IR from the WWW. Section 4.6 draws conclusions from this examination of current and previous research.

4.2 Underlying Issues

Achieving a proper overview of the user modelling field is difficult because of the increasing volume of work. Overviews do exist but suffer from being too specific, describing obsolete systems or lack of synthesis (Kass & Finin, 1988). Related to this is the fact that there have been various definitions of the term 'user model'. A user model can be considered to be (Coutaz *et al.*, 1992):

- the user's conceptual (or mental) model of a system, i.e. the picture the user has of a system and how it works
- a model of the user held within the system, i.e. a representation of a user's abilities, limitations, beliefs and goals
- a model or *definition* of a typical user referenced by system designer in aiding the building of a system.

Within the terms of this project, it is the last two definitions that apply. Coutaz *et al.*(1992) have suggested a general definition as *'an explicit and dynamic representation of all aspects of the user that are relevant to the system's adaptive behavior toward the user'* (p.419). For the purposes of information retrieval, behaviour can include the formation of user queries and the presentation of results to the user.

In intelligent information retrieval, the aim of modelling users is to improve the interaction between user and system. In other words, to improve the *performance* of the user in retrieving information from the database. For such an improvement, it is necessary that any retrieval system can somehow make *inferences* concerning the information the user might want. The system then can *aid* the user, for instance by giving suggestions or by adapting any query based on information furnished by the model. In order to do this the model should allow for the representation of user *interests* as they apply to the retrieval database, and for possible related interests according to the situation.

Several authors have attempted to classify user models according to a number of dimensions. These attempts highlight the underlying issues and are worth reviewing. Kass and Finin (1988) have proposed five dimensions: degree of specialisation, modifiability, temporal extent, method of use and acquisition procedure.

Degree of specialisation refers to a distinction between individual user models and models for classes of user (called generic models). Does the modeller create a model for each user, or models based on the assumption that different *types* of user can be identified?

Modifiability means deciding whether the user model is *static* or *dynamic*. A static model does not change during user interaction with the system, whereas a dynamic model is updated as soon as any data about the user becomes available. In order to track the goals of a user, models should be dynamic (Chang *et al.*, 1993).

Temporal extent refers to whether the user model is considered *short-term* or *long-term*. The model is long-term if the information contained therein is kept at the end of any interaction session, whereas if the information is lost then the model is short-term. An individual user model is generally long-term.

The fourth dimension of Kass and Finin (1988), *method of use*, addresses whether models are descriptive or predictive. Most models are descriptive, storing structured information about users, i.e. beliefs, goals and plans. According to Kass and Finin, a predictive model can also simulate the user for the system. Wahlster and Kobsa (1986) describe their use in 'anticipation feedback loops' in order to simulate the user's interpretation of a potential response of the system. Chang *et al.* (1993) have noted that in query systems, such as intelligent information retrieval, responses may be anticipated and presented to the user for further refinement. Ideally, any modelling component should use predictive models.

The *acquisition procedure* is the fifth dimension. That is, whether the model is acquired explicitly or implicitly. Explicit acquisition implies that the model is obtained by directly interrogating the user, so that a predefined model can be used or a new one constructed. Predefined models, such as User Modelling Front End (UMFE) (Sleeman, 1985), often involve stereotypes (see below). Implicit models are built during user interaction sessions by monitoring the user's behaviour. Often, a model is generated by combining both approaches (Chang *et al.*, 1993).

As already mentioned, a generic user model purports to represent a specific class of user. In some situations it might be appropriate to use a set of generic models rather that just one, in order to represent different classes. These classes are called *stereotypes*. Rich (1979) defines stereotypes as a *'collection of attributes that often co-occur in people'*, and proposes a stereotype-based approach for constructing user models that might be useful for this research. This approach basically involves building the model by dividing the users of a system into different classes. These classes or stereotypes are decided by clustering facts about the users into groups.

Others, such as Kok (1991b), have criticised the above scheme of classification for mixing two levels of analysis: technical and abstract. He proposes a classification based on four aspects - why, who, what, how - giving each aspect both a functional and a technical dimension. His classification is displayed in figure 4.1.

Two other issues to consider, for their effect on the user, are what Greef *et al.* (1988) call the change in *modality* of a system which employs user modelling and the issue of *privacy*. A change in modality means a change in the distribution of tasks between the user and system - usually the system takes on more to the detriment of the user - and a perceived shift of control over task execution in favour of the system. This might cause some consternation to users and needs to be addressed, perhaps by providing the user with the ability to bypass or override the user model during interaction with the system. The user also needs to be reassured that any information held about him or her will remain confidential.

To highlight how these issues have been tackled in other research, an examination of the application of user modelling to AI systems, specifically in regard to IR was conducted, and this is reported next.

ASPECT	DIMENSION	
	Functional	**Technical**
Why are users modelled?	What is the overall aim?	Which parts of the system need user information?
Who is modelled?	What is the user role in relation to the system?	How individual are the user models?
What is modelled?	What aspects of the user are represented in the model?	What is contained in the model and what are the interpretation methods?
How are users modelled?	What are the modelling methodologies and their sources?	Which user modelling techniques should be used?

Figure 4.1: One Classification of User Modelling Issues (adapted from Kok (1991b)).

4.3 User Modelling and AI Systems

Over the past few years the field of user modelling in AI systems has rapidly expanded (Kok, 1991a). In this field, user modelling attempts to deal with the characteristics of users working with these AI systems. The concept behind the use of user models is to improve user interaction with AI systems and to make it possible for the system to adapt to a particular user's individual preferences. McTear (1993) makes a distinction between systems containing adaptable models - those in which the user can make choices from a series of options and save these choices in a user profile - and systems with adaptive models - those in which user knowledge is acquired automatically, is updated over time and is employed to adapt the system to the user's requirements. In IR, the best models would be adaptive as the intention is to satisfy individual user's information needs which may change over time.

This adaptation may be one of refining the user model so that it increasingly contains more knowledge about the user and thus the IR system is adapted to reflect more closely a user who perhaps had generally fixed characteristics. Alternatively, the adaptation may be one of changing to cope with changes in the user characteristics, for example changes in preferences. Perhaps the best user models would be those which could cope with both adaptive processes.

User modelling has been applied to different types of systems, including intelligent tutoring (Kass, 1989), dialogue systems (Kobsa & Wahlster, 1989) and intelligent IR (Brajnik *et al.*, 1987). Kok (1991b) and McTear (1993) provide overviews of user modelling in different AI systems.

A distinction must be made here, for the purposes of the present research, between user modelling in AI systems and AI techniques employed in user modelling. The former situation arises in systems that apply AI techniques to achieve their objectives and which also include user modelling components. For example, an AI manufacturing system may employ captured domain knowledge using KBS to control intelligently the system, and also use a separate user modelling technique to represent knowledge gathered from the expert users of the system.

The latter or second situation is one in which an AI technique is used within the user modelling process. The technique is employed in order to lend some 'intelligence' to the process, to capture user knowledge and model it intelligently.

The present research has focused on refining a user's query before submission to an existing WWW search engine and the stated aim is to aid the IR process intelligently. One way to do this would be to employ knowledge about the user in the process of refinement. If the knowledge was modelled by application of an AI technique then, within the definitions imposed by the present research, it could be argued that the query is being refined intelligently. Therefore, it is this second merging of AI and user modelling that is more applicable to the present research.

In the light of the aims of this research, this examination of user modelling now concentrates on its application within the field of intelligent IR, as there are useful parallels to be drawn with previous research and developed systems.

4.4 User Modelling and Intelligent IR

User modelling has always found applications within the field of intelligent IR. Users of IR systems, in particular casual users, have great difficulty in exactly expressing their queries. A number of possibilities for alleviating this problem - allowing broad interpretation of queries, providing extra information, showing results in an informative manner - all need detailed information about the user (Kok, 1991a). A number of systems are described below which reflect the types of user modelling techniques used in intelligent IR. This discussion will shed light on the use of user modelling in this research.

One of the first IR systems to exploit individual user models was the GRUNDY system (Rich, 1979). GRUNDY recommends books according to its concept of user interests. Rich created the stereotypes used in GRUNDY on an *ad hoc* basis and although they can change gradually during user interaction, GRUNDY cannot automatically create new stereotypes. The user stereotypes are used to initialise and maintain the system's view of the user. The stereotype is considered to be a set of characteristics, or facets. Each facet is allocated a value between –5 and +5, thus producing in effect a stereotype which consists of facet-value combinations. The system uses *triggers* or events which activate a particular stereotype. A trigger is an object associated with a particular situation. A trigger has a name, the name of its associated stereotype and a rating (a measure of applicability to a situation). For example, in GRUNDY, if a user offers the self-description DOCTOR, this would trigger characteristics such as well-educated, caring, fairly affluent, and well respected in the community. Thus an individual user model is made by combining direct information from a user, inferences from the user's behaviour, and predictions based upon the stereotypes deemed appropriate to the user.

As this method builds up 'uncertain' knowledge, each item of knowledge needs to be tagged with a confidence rating. A rating is tagged to each facet-value combination. For each stereotype, the system maintains a list of triggers which activate it. The more triggers related to a stereotype, the higher the appropriateness rating of that stereotype. A list of facet-values must also be associated with each book and therefore the system does rely on domain knowledge.

GRUNDY achieved the goal of showing that for one system, in at least one domain, individual user models could improve retrieval performance. It also highlights the use of stereotypes but because it relies on extensive domain knowledge, the actual nature of the models would not be applicable to this research.

The Information Retrieval - Natural Language Interface (IR-NLI) (Brajnik *et al.*, 1988) has a user modelling component called UM-tool. Here the user models are used to interpret the search requests and improve on them if necessary. The model contents are divided into two parts: one containing specific information about the user, the other about the level of familiarity with the subject domains, the databases (e.g. INSPEC) and IR systems in general. Again the models are constructed by selecting one or more stereotypes. Subsequently, after every interaction, the selection is re-evaluated and those stereotypes that no longer apply are deleted from the user model. Other sources of information for the models are the current interaction and a database of former sessions.

The Intelligent Intermediary for Information Retrieval, I^3R (Thompson & Croft, 1989) is a document retrieval system that combines several methods of retrieval. One of the interesting aspects of this system is its use of a domain knowledge network for extending a query by means of spreading activation - a technique discussed in chapter 3. The user modelling function in I^3R consists of a simple classification of the user and bookkeeping of user-specific information. Unfortunately, I^3R relies heavily on domain knowledge, i.e. knowledge about the information being retrieved *not* the user's understanding of the knowledge, to function properly.

The IMPACT system (Botman *et al.*, 1987) is an implemented prototype of a user modelling interface to databases and is domain independent. The user models are constructed through the current user interaction and use of information from previous dialogues with this and other users. Patterns in the interactions are sought and stored as *profiles*, small sets of interests with grades (how interesting) and certainty factors (how sure is the system). There is some resemblance between profiles and small stereotypes, but the modelling in IMPACT is much more dynamic. This means the system can quickly adapt to changing user interests - a factor that could be important to the aims of this research. Obviously, a disadvantage is the increased use of resources, in terms of computation and time, in constructing dynamic models, but traded against this is the adaptive power of this method.

Logan *et al.* (1994) outline an approach which attempts to model the information intermediary - someone, such as a library and information science professional, charged with aiding a user to satisfy an information need - through an AI theory of belief revision. They apply the work of Galliers (1993) and Belkin *et al.* (1993). Part of the argument is that the information intermediary can be modelled as a collection of specialised functional experts, engaged with subtasks of the overall task of satisfying the user's information need. Included in the 'experts' is a part termed 'User Model'. Logan *et al.* incorporate a measure of cognitive intention which is dependent on beliefs in a measure they call cognitive attitudes. This is a knowledge-based approach using inference rules. The prototype system holds a dialogue with the user in order to establish their needs and beliefs. The authors themselves note that the approach produces computational complexities because many sets of attitudes can be produced. It is difficult to see how it could be applied to IR from the WWW.

The systems described above use a knowledge-based approach to build their models. An alternative approach is the connectionist approach described by Chen and Norcio (1992). They propose a research framework for building user models by using Neural Networks (NN). They view the information in a user model conceptually as an individual task-related 'user image'. Thus, pattern recognition techniques can be used to process this image. They argue that NN, with the ability to learn, can deduce the relationship between user knowledge input and model output regarding beliefs, goals and plans. Their approach is to use NN techniques such as pattern classification, pattern association and associative networks to classify, store, represent and adapt user information. The authors do not report any implementation of the approach so it is impossible to assess its usefulness for IR. However, as noted in chapter 3, Jennings *et al.* (1993) have employed NN to develop a user model for a personal news service. A user model of preferences is constructed based on articles read and rejected. The NN adapts over time to better represent the user's interests and once in place can rank articles according to this model. It therefore does provide a way to adapt to changing user preferences though the authors acknowledge that it is difficult to do. There is no way to take account of sudden and distinct changes in the user model as nodes and links can not be destroyed. They will only decay slowly as the user continually rejects articles that contain the words they represent. Thus, models would have to be recreated from scratch to account for those changes that could be expensive and time-consuming.

A criticism of the above approaches is that they have not been tested on large information stores in real systems and thus their use in commercial situations is in doubt. There have been a number of systems which have employed a restricted sort of user modelling which might ultimately prove more useful for real systems. Kass and Stadnyk (1992) employ a fairly simple model in an information filtering system. The model maintains knowledge and information needs of the members of an organisation and uses this model to determine the distribution of new product information.

Slightly predating the development of the WWW, Boyle and Encarnacion (1991;1993) and Kaplan *et al.* (1993) offer two different applications of user modelling to hypertext IR. Boyle and Encarnacion describe an adaptive hypertext reading system MetaDoc, which uses the technique of 'stretchtext' to vary the amount of information and detail presented to an individual user. The domain for MetaDoc is the management of a computing operating system and what each user sees is based on the system's model of the user's level of expertise with the operating system. Users indicate their level before the first session and are classified according to the information they provide as novice, beginner, intermediate or expert. After that the system monitors the user's actions in interacting with the system and can reclassify the user, changing their level of expertise. For example, lack of knowledge of a required concept leads to demotion whereas a request for more detail implies an understanding, in MetaDoc's concept, of the basic concept and leads to an upwards revision of the user's level of expertise.

Kaplan *et al.* use a different approach in the HYPERFLEX system, which supports hypertext navigation through text by recommending particular topics based on the user's informational preferences and goals. In this system, association matrices (Maybury, 1995) are used to link topics together in a document and to link topics with nodes representing particular user goals. Weighting of the nodes allows the system to learn by adjusting the weights through user feedback. In this way the system is adaptive, though it is obviously heavily dependent on information about document contents and it is difficult to see how it would function well with large changing document sets. This system is a precursor to the 'recommender' systems now becoming popular on the WWW. These are discussed as part of the final section of this chapter which examines how the user modelling has been applied to aiding IR from the WWW.

4.4 User Modelling and IR from the WWW

Chang *et al.* (1993) noted some time ago that:

'[A] user model could be particularly useful for interacting with a distributed database system, because these databases are queried by users with very different interests, either experts or non-experts, requiring information at various levels of abstraction depending on particular motivations.'

Even though the focus of their work was different - modelling users in an adaptive visual interface - their point is very relevant to the process of IR from the WWW. The WWW can be considered to be a globally distributed heterogeneous database and the users of the WWW differ greatly in terms of interests, experience and subject domain knowledge. Over the past few years, WWW researchers and developers have recognised the potential of employing information about users to improve IR, and this research is examined in this section.

An early attempt to apply a kind of modelling - user profiling - to IR from the WWW was the Borges project (Cullen, 1997). This was a project to develop a prototype information filtering system for Usenet and WWW information sources. The concept is that the user constructs personal profiles of their queries or information needs and these profiles are then matched against documents by means of a filter tool. The tool filtered all postings to a set of newsgroups and a number of specified WWW pages. These pages seem to have been chosen in a rather ad hoc way - on the advice of those taking part in an evaluation of Borges or of staff in the library where the project was developed. Each profile generates a digest of the top 50 ranked documents. Profiles can be refined or created anew and users can have as many as they wish. Cullen presents anecdotal evidence of a good response but there seems to have been no statistical evaluation of the user study.

Based on experience in developing the WWW tool the RBSE (Repository-Based Software Engineering) Spider (Eichmann, 1994), Eichmann and Wu (1996) are attempting to apply user modelling in a system called Sulla, which they term 'a user agent for the WWW'(p.1). This is a software agent with responsibility for tracking WWW information and relating it to the user in a form which reflects the user's interests. These interests are inferred from a user interest profile. The knowledge representation scheme required to do this is yet to be developed, so it is not currently possible to assess the usefulness of the user modelling component. Presently, Sulla stores the results of searches using existing WWW tools and uses the keywords they contain to seek out and offer to the user other URLs that might match the stored results.

Another project employing user profiles and being carried out under the auspices of the eLib program (UKOLN, 1998) is NewsAgent - an electronic library project designed to provide a personalised current awareness service for Library and Information Science (LIS) professionals (Kerr, 1997; Yeates, 1998). The concept is to select, collect and filter information before presenting it to the user and to do this independently. There are other projects involved in developing similar kinds of Internet alerting service (NewsPage Direct, 1997; Pointcast, 1997). The core of the NewsAgent system is a user profile, completed via a WWW form, which contains information regarding the user's information needs (chosen from a list of keywords or topic headings) and the frequency of information delivery (daily or weekly). Thus, the user preferences are modelled in order to provide as output an email or a WWW page which contains a set of references matching those preferences.

The references are taken from selected information sources which were decided as being relevant to the LIS community in a user needs assessment. The system also incorporates sources which have been marked-up using the Dublin Core metadata standard (Weibel & Miller, 1997).

As noted in chapter 3, Bell (1996) is developing a system called HyBiS which is designed to be an adaptive navigation system for hypermedia. It is included here as the adaptation is driven by the user. The essence of the system is that the user's navigation of the WWW drives an adaptation engine. In a sense the user is modelled since information from each point or *node* visited by the user whilst browsing the WWW is then employed to provide the user with new links. The system is, according to Bell, assessing the user's interests from the information provided by each node. The assessment technique employs metadata, the keywords associated with each WWW node visited by the user, to construct a model of the user's interests. This model is basically the navigated path of the user's previous interaction with the WWW. This path provides a guide to what the user is looking at via the metadata associated with each node. The system then uses the model to extract information from the WWW, dynamically generate new search terms and produce a new set of links by submitting these terms to a search engine. This is an example of the use of machine-learning to improve IR. The system is highly dependent on the quality of the metadata which in the WWW presently is quite poor, though the current initiatives on metadata might improve that situation (Heery, 1996). Also, it is difficult to see how the system would cope with rapid and marked changes in the user's interests.

Resnick and Varian (1997) make the point that often it is necessary to make choices without experiences of the alternatives. In other words, reliance and choice is made on the recommendations of others, either by word of mouth or via printed matter such as guides or reviews. There is a class of systems - termed recommender systems - which has been developed on the WWW to assist and augment this process for electronic information. Some of these systems are included here because, in terms of the way they function - attempting to model users' interests and preferences to other users or information - they display a modicum of user modelling (Balabanovic & Shoham, 1997; Rucker & Polanco, 1997). The developers of the first system, Tapestry (Goldberg *et al.*, 1992), coined the phrase 'collaborative filtering' but recommender systems is now preferred for two reasons. One, the system may not explicitly collaborate with recipients who may be unknown to each other. Two, recommenders may suggest particularly interesting documents, as well as indicating those to be filtered out.

Siteseer (Rucker & Polanco, 1997) uses individual WWW bookmark lists ('hotlists' or 'favourites') and the organisation of these lists into folders for predicting and recommending relevant WWW documents. User's interests are modelled by considering each bookmark to be a declaration of interest in the underlying content and the grouping of subjects in folders as an indication of semantic coherency of relevant groups between users. Siteseer examines each user's folders and bookmarks and measures the degree of overlap (common URLs). It uses this measure to create a virtual community of users with similar interests. To any user, it then provides relevant WWW documents which have been bookmarked by that user's virtual neighbours.

The developers themselves recognise the limited potential of such a purely collaborative approach. It is impossible to help a first-time user or one creating a new category for bookmarks because there is no collective experience to draw on. Also, the system relies on users organising and structuring their bookmarks coherently. Presently, the system has only 1000 users so it is difficult to assess if it would scale up to cope effectively with many users and the vast number of possible subject categories in the WWW information space.

The creators of Fab (Balabanovic & Shoham, 1997) make the distinction between *content-based* and *collaborative* recommendation. In the former, documents which are similar to those a user has shown interest in previously are recommended. In the latter, users whose interests are similar are identified and then documents which they found interesting are recommended. The content-based approach, which involves matching between a user profile and document sets has its roots in the IR field (Harman, 1994; Krulwich & Burkley, 1996; Lang, 1995). It suffers from a number of problems, including over reliance on user feedback and on knowledge of the document set. The collaborative approach also suffers from its own problems (Konstan *et al.*, 1997) such as how to deal with completely new documents that have not been rated by anyone. This is a crucial factor in the WWW, where the document corpus continues to expand rapidly. There is a danger that purely collaborative systems will suffer from poor rated to unrated document ratios.

Fab attempts to avoid these pitfalls by employing a hybrid approach. It maintains user profiles based on content analysis and directly compares these profiles to determine similar users for collaborative recommendations. It employs collection agents to retrieve information into a database and selection agents to collect and display information for a particular user from the database. The user profile is maintained within the selection agent and, as the authors note, the construction of an accurate profile is the key task. Effectively, the user is modelled within this profile and the profile is built from the keywords within previously user-rated webpages. This is updated every time the user provides feedback so the system is heavily reliant on user intervention. Also, the system may struggle to cope with sudden changes in the user's preferences. The system has been evaluated with a small number of users and has shown promising results but again it remains to be seen whether it will scale to make it an effective and realistic IR tool.

Recommender systems are another example of machine learning techniques being applied to retrieving information for the WWW user. The employment of user modelling is limited and sometimes not very sophisticated but the results highlight the advantage of applying user modelling to IR from the WWW.

4.5 Conclusions

The above discussion shows that user modelling is a very active and on-going area of research and development within AI systems. User modelling has long been applied to the problems of IR and the development of the WWW has given a boost to this aspect of user modelling research. It has also illuminated the advantages and disadvantages of employing user modelling in this way.

One advantage is the shift in focus away from the information corpus itself towards the user trying to access the information. This is crucial in trying to improve IR results from such a vast information store as the WWW. The limited success of those systems which employ user profiles as their modelling component highlights what might be achievable, in terms of improving IR results, by modelling the WWW user. User profiles can be considered to be similar to the stereotype construct employed by Rich (1979) and both these constructs have been used extensively in user modelling (Beaumont, 1996; Botman *et al.*, 1987; Brajnik *et al.*, 1988; Cullen, 1997; Yeates, 1998). The actual implementation of some of these profiles is not that sophisticated and it is in doubt how well they model the user's information needs. The issue is to make the stereotype or profile as adaptive and flexible as possible whilst effectively capturing and representing the required aspect of the user being modelled, whether that be a preference, an information need, or information about the user's knowledge or experience.

A disadvantage of the machine-learning approaches to IR from the WWW and the earlier knowledge-based approaches to intelligent IR is their inflexibility and inability to adapt to changing user characteristics. Sheth and Maes (1993) have suggested applying genetic algorithms to introduce randomness and spontaneous evolution into the process but there is doubt over preventing inconsistencies appearing in the user models. The NN approach has shown promise (Jennings *et al.*, 1993; Sholtes, 1992) but there is high computational cost in training the networks and difficulties in ensuring convergence without spurious nodes appearing.

The main conclusion to be drawn from the investigation reported in this chapter is that user modelling can provide a way to capture and represent user characteristics which can then be used to aid IR from information sources such as the WWW. Those models which are adaptive and flexible seem to achieve the most. There have been different techniques applied to represent user characteristics and they have had limited success.

Drawing on the conclusions of chapter 3, which indicated that fuzzy logic would be a valid technique to apply to intelligent IR, and noting that this technique has not been applied to user modelling, it was decided to use fuzzy logic to create the user models in this research. Fuzzy Logic (FL) is a suitable AI technique to employ in the creation of the user models for a number of reasons.

One, it is possible to use FL to capture and represent degrees of knowledge. This knowledge could, for example, be a user preference or indication of experience. Fuzzy logic is a technique which is flexible enough to represent vague concepts such as 'mainly prefer' and 'quite experienced'.

Two, the constructs of the technique are such (for example the fuzzy knowledge base) that degrees of knowledge and information captured can be propagated through the intelligent IR system. Hence, less knowledge may be lost through the use of FL than might be through the use of more rigid techniques, for example a traditional expert system.

Three, its inherent flexibility will allow the process of capturing user knowledge to be more realistic. Fuzzy logic is able to handle the answers to open questions such as those which employ a sliding scale to elicit the user's strength of agreement with two extremes. All the degrees of agreement in such a situation can be readily represented in FL.

Four, the use of FL would allow the models to be adaptive, in the sense of reflecting changes in the user information, as fuzzy sets can adapt to changes in the knowledge they represent.

Therefore, FL seems eminently suitable to providing the adaptability needed to provide effective user modelling. This fuzzy user modelling approach to intelligently aiding IR from the WWW is novel and is discussed in greater detail in the next chapter.

CHAPTER 5 - INTELLIGENT INFORMATION RETRIEVAL: A NEW APPROACH EMPLOYING FUZZY USER MODELLING

5.1 Overview

The three previous chapters examined, in turn, IR from electronic sources, attempts to produce 'intelligent' IR by the application of AI techniques, and the role of user modelling in IR. The outcomes of this examination were drawn together and informed the development of a novel approach to IR - *Fuzzy User Modelling* - which is reported in this chapter.

The research covered in chapter 2 highlighted the chronic need for tools to assist the IR process. The growth of networked and electronic sources of information continues with unabashed pace. The advent of digital libraries, hypermedia databases and autonomous agents points to increasing interest in and continuous development of electronic sources of information (Agentlist, 1997; *Ariadne*, 1998; Collier, 1993; DMU, 1997; Duval, 1998; Maes, 1994a).

Attempts to apply AI to the process of IR are not new and have been on-going for many years, from the early expert systems in libraries, such as POINTER to the current cutting-edge research of intelligent agents (Agentweb, 1998; Cavanagh, 1989). These attempts, reported in chapter 3, though many are limited in their success, have shown it is possible to improve IR results through the application of AI techniques. However, many of these applications have relied on extensive knowledge of the information corpus. In the present research, the interest is in improving IR from the WWW. The distributed heterogeneous nature of the WWW precludes the use of techniques which rely on domain knowledge, and thus two major factors influencing the present research were gleaned from the investigation reported in chapter 3. One, the focus of the research should concentrate on applying intelligence to an area of which a lot could be known - the users' queries. Two, that AI techniques which are adaptable and flexible, such as neural networks and fuzzy logic, should be further evaluated. This evaluation, also reported in chapter 3, showed that fuzzy logic, with its inherent ability to capture and represent partial information, would be suitable to employ in research which was now focusing on refining user queries.

It is the contention of this thesis that one of the major problems with IR lies in the prospective searcher, whilst perhaps having a well-defined information need, have only a vague idea of how to formulate an IR strategy in which to express and therefore satisfy that need (Cavanagh, 1989; Chen & Dhar, 1990; Gauch & Smith, 1991; Gauch *et al.*, 1998). This vagueness leads to the classic IR problems, such as information overload, and these are exacerbated by the nature of the WWW. For example, a search with the tool Alta Vista using the search string 'information retrieval' produced 62 ,000+ hits (Alta Vista, 1999a).

However, by applying user modelling techniques to gather and represent information about users' knowledge and experience, an information need or query can be refined and hopefully the IR results obtained from any search improved (Eichmann, 1996). This premise led to an investigation of the role of user modelling in IR which was reported chapter 4. This investigation showed that the use of *stereotypes* - default categories of user determined by the values stated for a number of criteria - and *profiles* - user information, such as their goals, aims, previous interests and information search histories, used to classify users - have been employed to aid the process of IR with some success (Botman *et al.*, 1987; Rich, 1979). It became apparent that such techniques could be, and recently have been, applied to IR from the WWW (Cullen, 1997; Eichmann & Wu, 1996).

Thus, the focus of the present research became one of applying user modelling techniques to assist intelligently the refinement of user queries before they are employed to retrieve information from the WWW. As fuzzy logic had been the technique chosen to imbue this intelligence, the whole concept is termed *Fuzzy User Modelling*. This combination of fuzzy logic and user modelling techniques was developed with the research issue stated in chapter 1 in mind. This issue can be restated, taking account of the new concept as:

> Whether the application of a prototype system, developed using Fuzzy User Modelling (FUM), can improve the IR results from the WWW, for novice and intermediate users, employing queries in one information domain?

Note that the research focuses on what are termed 'novice' and 'intermediate' users. In conducting this research it became clear that any system developed to assist IR would likely most benefit those users whose experience and knowledge of the WWW and the information domain was not substantial. Expert users are more able to satisfy their information needs and thus any benefit to them is likely to be slight. The benefits to less able users are probably likely to be greater and thus easier to detect and assess. This issue begs a number of questions, such as how improvement is defined and measured, how novice and intermediate users are defined and represented, and how the application of FUM is implemented and assessed. The implementation, through the development of the prototype system, is discussed in depth in chapter 6. Chapter 7 explores the assessment of the prototype via means of a user study , and examines issues of performance and 'improvement'.

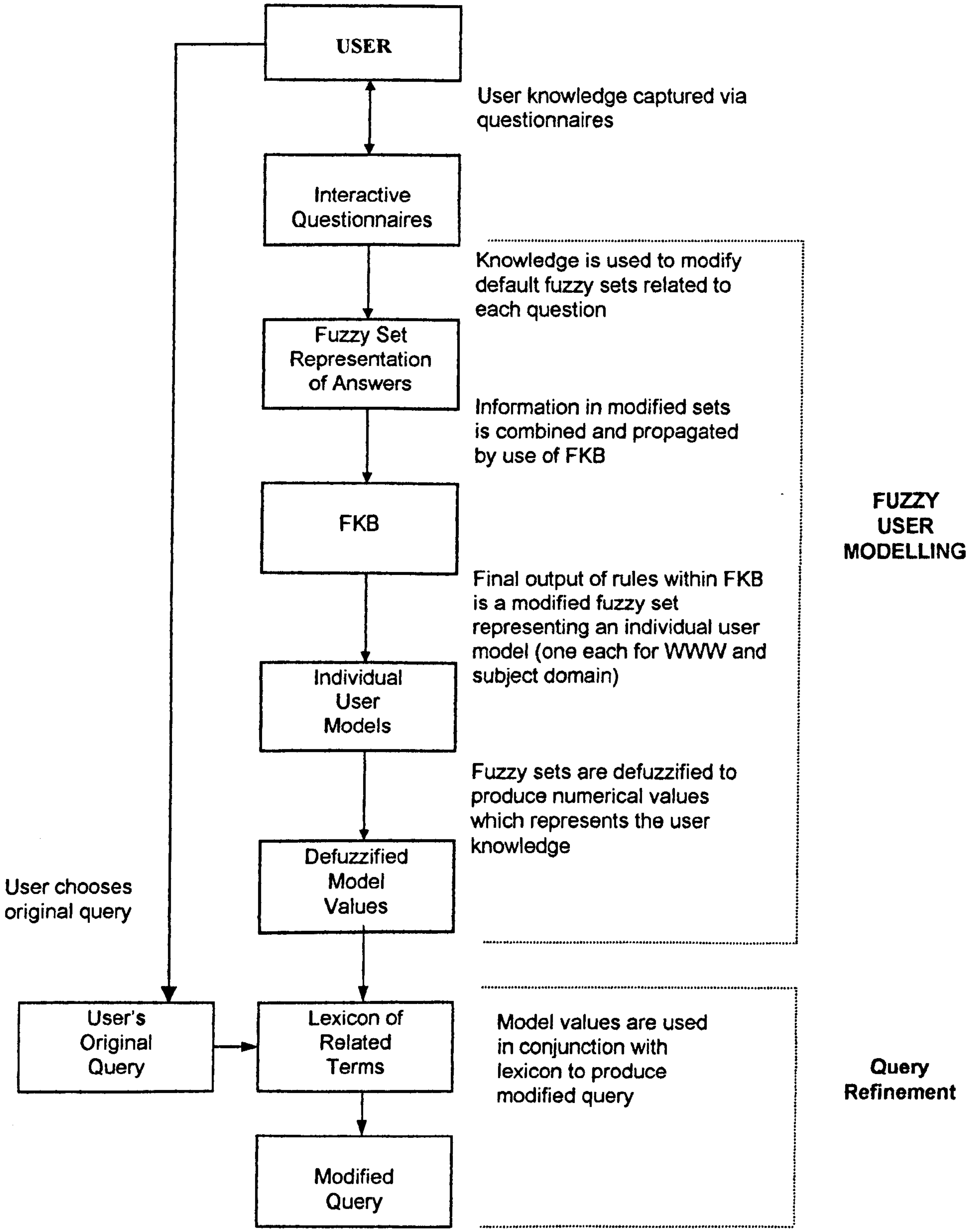


Figure 5.1 The Stages Involved in Fuzzy User Modelling and the Subsequent Query Refinement

The different sections of this chapter explain the methodology of FUM in detail. However, it is useful to have an overview of the approach – this is shown in figure 5.1. Briefly, FUM consists of the following stages. Knowledge gathered from the user interaction with questionnaires is used to modify default fuzzy set regions associated with each question. The information contained within these sets is combined through the use of a Fuzzy Knowledge Base (FKB), the output of which is two modified fuzzy membership functions which represent two individual user models – one each for the user's WWW and information domain knowledge and experience. These two functions are defuzzified to produce two numerical values. The last stage is not strictly part of the FUM process but is rather part of the query refinement process. The two numbers are used, in conjunction with the user's original query, and a lexicon of semantically related terms, to modify the original query.

The rest of this chapter focuses in detail on the approach just outlined. Section 5.2 looks at the application of user modelling and how it relates to issues discussed in chapter 4. Section 5.3 examines the roles of fuzzy logic in the present research, focusing on the fuzzy sets and fuzzy knowledge bases (FKB) employed to develop a coherent representation of the user. In section 5.4, the whole concept is examined, looking in particular at how it can be used to refine information queries. Finally, conclusions are drawn in section 5.5, in which the strengths and weaknesses of FUM are discussed.

5.2 Modelling the WWW User

It has already been established that, by modelling a user, it is possible to affect and refine the interaction with the WWW in order to improve the results of any search for information (Eichmann, 1996). In the context of the present research, modelling the WWW user means acquiring and representing information regarding the user's knowledge and experience of both the WWW and an information domain.

For the purposes of constructing models quickly and testing the approach, only one information domain was considered - AI topics in computing - but, as will become apparent, the approach is generalisable to other domains. Background information about the user, such as gender, occupation, age range was also collected, for the purposes of analysing the results of the user study (see chapters 7 and 8).

In user modelling, the use of stereotypes is well documented (Kobsa & Wahlster, 1989; Rich 1979). Stereotypes are constructs for placing the user in a default category by means of the values of a number of criteria. The concept is that the IR system can then make assumptions about the user and the information need, depending on the default category or model in which they are placed.

There should be a way, for example through user feedback, for the default models to be refined to represent the individual user more closely - stereotypes that can do this are termed *adaptive*. Figure 5.2 schematically reflects this process of building individual models through the use of adaptive stereotypes (Rich, 1979).

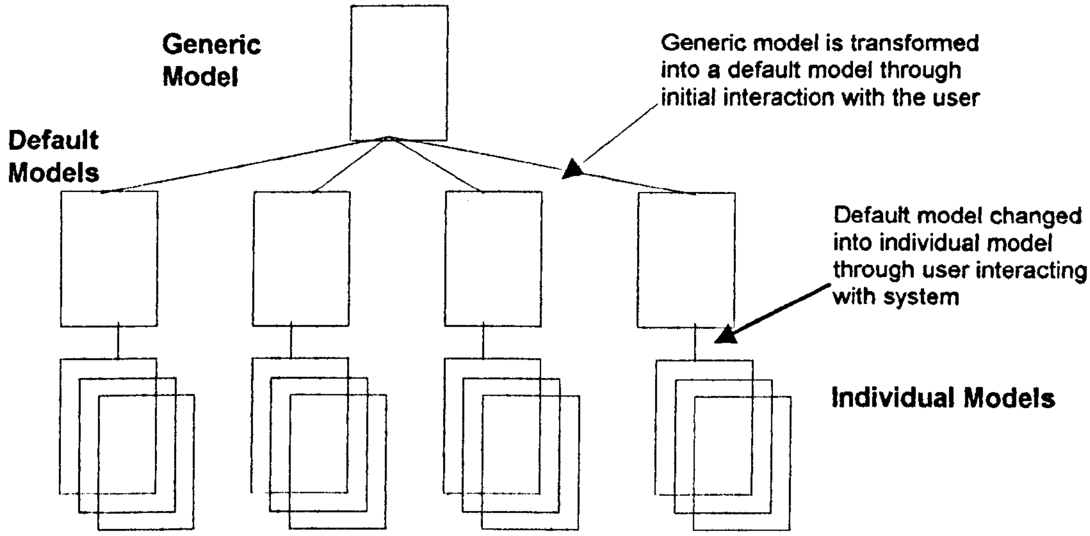


Figure 5.2 A possible model architecture using adaptive stereotypes.

Note from the figure that it is possible for the user to have more than one individual model. This might be useful if the user wants to retain the instances of their models as they develop in time, or if they want to pursue more than one different and distinct information need.

In the present research, individual user models are developed by modification and combination of default regions or categories representing 'novice', 'intermediate' or 'expert' users with regard to WWW knowledge and experience and to knowledge of the information domain. In this sense, these categories can be considered to be stereotypes into which the user fits. Intuitively, however, and in reality a user would move through these categories in a continuous way as their knowledge and experience grew (Zadeh, 1994). Also, at one instance, a particular user may be considered to be novice by some and at an intermediate level by others and, therefore, in one sense can be thought to have some degree of membership of both categories. As will be shown in section 5.3, which details the FL representation of the user models, the constructs of FL can easily be used, not only to represent categories that are not discrete, i.e. continuous categories, but also to account for simultaneous degrees of membership of more than one category, i.e. overlapping categories (Zadeh, 1994). Thus, FL can be used to build models that reflect the real-world situation.

So, the use of FL to create models differs from the building of user models by employing stereotypes. Stereotypes, even if they are adaptive over the time of their use, usually represent discrete types or category of user and the modelling process places the user in one stereotype. Using FL takes account of the realistic linguistic notion that a user could co-exist in two categories simultaneously, to a certain degree.

The use of question-answer dialogues and questionnaires to gather information about users in order to build models is well documented (Kobsa & Wahlster, 1989; Nessen, 1987). In the present research the user is not placed wholly in one category by answering questions which allow for only a few distinct answers, as is the case in some previous approaches involving stereotypes (Rich, 1979). Rather, the user answers a series of sliding-scale questions , indicating how well they agree with the two possible answers on either end (see figure 5.3, for an example). Each question within the questionnaires has default categories of novice, intermediate and expert related to it. The answer to a question thus provides knowledge about the user which is used to indicate membership of a category *to a certain degree* for that particular question. This process is repeated for each question, separately for the WWW and information domain, and thus the user knowledge is modelled and retained within the modified category representation. Section 5.3 details how this done and how this information is combined and propagated through a fuzzy modelling process until two individual user models (one for the WWW and one for the information domain) remain.

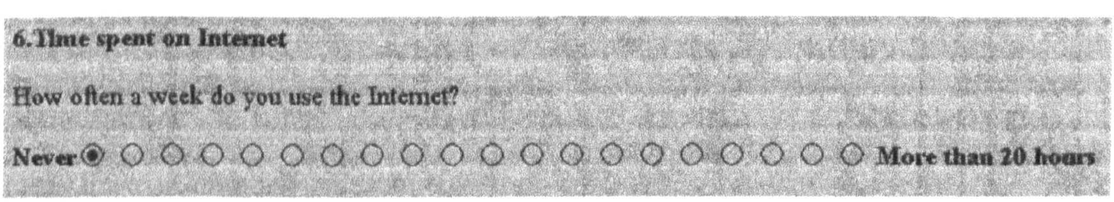


Figure 5.3 An Example Question from the WWW Questionnaire.

This approach is one of direct or explicit knowledge acquisition in order to build and adapt predefined models, an approach which the analysis of chapter 4 showed to be prevalent in previous research (Chang *et al.*, 1993; Sleeman, 1985). The questions are designed to elicit and acquire sufficient knowledge to build useful models.

Modelling of the users in this way is also related to the employment of *user profiles* in IR and information filtering, an approach which, as discussed in chapter 4 (page 54) has been used with regard to the WWW (Bell, 1996; Etzioni *et al*, 1997). Profiles - which can represent such aspects of a user as goals, aims, interests, previous search histories - are employed to modify IR searches and results. In the present research the user's questionnaire answers modify default overlapping regions to produce individual models. This individual model can be considered to be a profile of the user's experience and knowledge. However, though it can be considered to be a profile, the model built via FUM is more robust in the sense that it is based on knowledge directly acquired from interaction with the user via the questionnaires. Profiles often rely on past information, such as previous search histories, to obtain a picture of the user – information which could rapidly lose significance for the user.

The user models are used to modify and refine a query in the information domain. Section 5.4 discusses in detail how FUM is used to refine queries but, before that, section 5.3 examines how fuzzy logic has been employed to implement the user modelling technique.

5.3 The Role of Fuzzy Logic

Zadeh (1996) has called fuzzy logic "Computing with words (p.101)". By this he meant that fuzzy logic offers a better methodology for computing with words than other AI techniques such as neural networks, Bayesian networks and predicate logic. Chapter 4 (page 59) provided the argument for using FL as the user modelling tool. Recall that it is its abilities – to capture, represent and propagate degrees of knowledge, to deal with vague linguistic concepts, and to exhibit flexibility and adaptivity in representing such concepts – which form the crux of the argument.

Zadeh (1996) notes that humans employ mostly words in computing and reasoning, and that the graded meanings these words may have can be captured and represented by fuzzy logic. Fuzzy logic has also been noted for it ability to model linguistic categories by others (Brimicombe, 1997; Koo, 1996; Oliver, 1997). Dubois *et al* (1997) point out that Zadeh's view highlights:

"the importance of the interface between data emanating from the physical world and the categories with which human beings are most comfortable in comprehending and using information" (p.4).

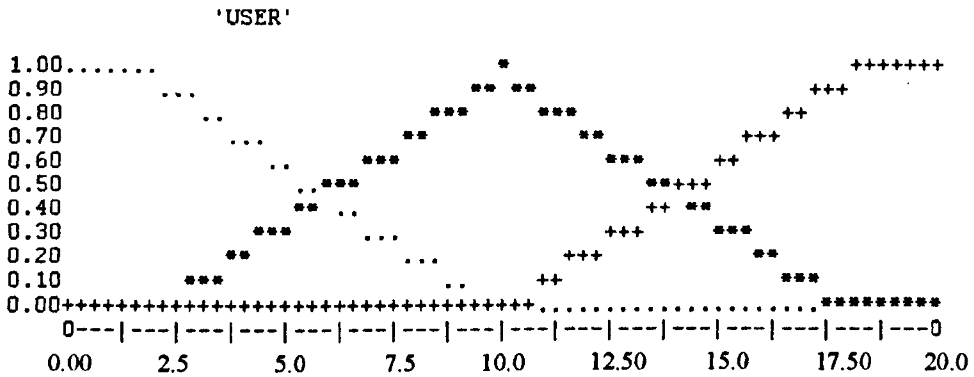
Thus, in respect of IR, a process which involves the expression of an information need in words, fuzzy logic seems an eminently suitable technique to employ in any approach designed to assist this process. A number of different semantics can be associated with the use of fuzzy logic and sets. Within the context of this present research and the modelling of WWW users, two different semantics are related. One is the use of fuzzy sets to represent incomplete or vague states of information (Zadeh, 1976). The use of open-ended questions to gather information about the user allows a certain amount of vagueness in the answers. A fuzzy set representation is very able to deal with this vagueness.

The second semantics, which is more applicable to the present research, is useful when modelling a flexible constraint, specification or goal. In this semantics, preferences are expressed between more or less acceptable solutions with respect to the constraint (Bellman & Zadeh, 1970). The gradeness introduced by fuzzy sets refines the simple binary distinction made by ordinary constraints between completely acceptable and completely forbidden values. In chapter 3 (page 34), it was argued that FL, with its ability to capture and represent partial knowledge, was a valid AI technique to use in the present research. This ability means that it is well suited as a tool to use to represent knowledge about users' preferences, opinions and experiences. Also, the property of gradeness allows this knowledge to be successfully represented even if it is vague and imprecise. Finally, the constructs of FL allow such knowledge to be retained and propagated (see below). Recently, support for applying FL to IR from the WWW has been expressed by Chang and Chen (1998), who have developed a document retrieval model for WWW documents based upon using FL to capture and represent the uncertainty in, what they assert are, the three main components in any IR system. These are: document representation, user queries, and document evaluation. Their work differs from the present research in that they use a fuzzy-knowledge base approach to develop a concept network in which queries and documents are uniformly represented by a set of concepts. The present research focuses the fuzzy knowledge base on refining queries.

Thus, in terms of representing information about WWW users as an overlapping series of default regions which are then rendered as an individual model via graded answers to questions, FL seems to be the ideal technique to employ.

Dubois *et al* (1997) have noted that, in exploiting information, fuzzy sets may appear, amongst other aspects, in the representation of uncertainty and in the expression of preferences. In employing fuzzy sets to represent user models, this is exactly what the present research is trying to achieve.

The user is modelled, then, by two default models - one for WWW knowledge and experience and one for domain knowledge and experience. Each model is represented by a series of three overlapping fuzzy sets (see figure 5.4 for an example and Appendix A for the complete set) and each set represents the region or category in which a user can be 'novice', 'intermediate' or 'expert'. Such a fuzzy set representation is related to each question within both questionnaires and how these representations are modified and combined to produce the individual models is discussed below. By employing this representation, it can be seen from the figure that it is possible for a user to have a grade of membership of more than one area simultaneously. Thus, this representation takes account of the real-world notion that a user moves through these stages continuously as their experience and knowledge accrues. To some extent the subjectivity of the terms novice, intermediate and expert are accounted for by the fact that each set covers a region of values on the x-axis and that a particular value on the x-axis can often map to more than one area. The terms themselves are in reality fuzzy concepts and the sets reflect this reality.



. Fuzzy Set - Novice
Domained 0.00 to 20.0
* Fuzzy Set - Intermediate
Domained 0.00 to 20.0
\+ Fuzzy Set - Expert
Domained 0.00 to 20.0

Figure 5.4 Fuzzy Set Representation of Default User Model.

The range of values or domain - shown along the x-axis of figure 5.4 - to represent the fuzzy concepts of novice, intermediate and expert is 0-20. This range reflected the sliding-scale answers allowable for each question in the questionnaires, which ranges from 1-20. Although a different scale could have been used, say 0-10, the domain scale was chosen to make it easier to use the information gathered in the answers to adapt the default user models, and to create an individual fuzzy user model. A scale of 20 was used in the questionnaires so as to allow for a finer graduation in the respondees' answers than would be produced by using a scale of say 3 to 5 possible responses to each question. Any finer graduation of possible answers (say 50 or 100) was precluded by considerations of the WWW interface and the users experience. Too many possible circles or squares to check would have been unwieldly and difficult to achieve in terms of display and may have confused the users.

The y-values range from 0 to 1 and represent the strength of membership of a region. Thus, any value on the x-axis between 0 and 2 would be considered to have a membership value of 1 in the novice region, or more plainly, to be wholly within the novice region. The shapes or *membership functions* of the sets shown in figure 5.4 are known as trapezoidal. The membership function for the intermediate region is triangular which, as discussed in chapter 3 (page 34) where membership functions were first described, can be considered to be a special case of trapezoidal in which two of the main co-ordinates are the same.

Choosing this membership function for the sets was influenced by research into fuzzy logic and control processes. This has indicated that trapezoidal type functions are very effective for the representation of fuzzy concepts which have typical values (Kosko, 1992; Lee, 1990). In the present research, the sets are representing the fuzzy concepts of novice, intermediate and expert and thus trapezoidal sets or functions are a valid approach to adopt, even though the numerical values given to the concepts are somewhat arbitrary. Arbitrary, in the sense that, whereas it is fairly easy to place a typical range of values on a fuzzy concept such as 'about 5' - this could be represented by a fuzzy set covering the range of numbers 4.5 - 5.5, say - it is harder to interpret the fuzzy concept 'novice' with a range of values. However, this is not an issue for the present research as it is concerned with the veracity of the membership functions only in the sense of how they perform in aiding the IR process, and not with showing that certain functions are more empirically correct than others.

Zadeh (1994) has argued that though a numerical value such as 5 is simpler than the function 'novice', in the present case novice is a choice of one out of three possible states whereas the number 5 is a choice out of, say, 20 values, in this case. The point is that, instead of placing ranges of values upon each linguistic variable, words such as 'completely novice' and 'extremely expert' could have been used but the outcome would be the same. The use of words or *linguistic values* would have, in this case, the same effect as using numerical ranges. That is, one of data compression which can be thought of as fuzzy granulation (Zadeh, 1994). This granulation, as can be seen in figure 5.4, results in a gradual rather than abrupt transition from one region to another, thus mimicking the way in which humans perceive linguistic concepts. The use of numerical ranges rather than words made the interpretation of the knowledge captured by the fuzzy modelling process easier in this case.

Similarly, the actual membership function of each fuzzy set was decided heuristically. To ascertain the functions of these sets from empirical evidence would have involved a large and time-consuming study of the users, their abilities, perceptions and experiences. In fact, the determination of membership functions is an important research issue itself (Chen & Otto, 1995; Kantrowitz *et al.*, 1997) and is outside the terms of reference and scope of the present research. However, the membership functions of the sets were not chosen arbitrarily, and so are not without meaning or import. Figure 5.4 reveals that the shapes imply that, at the lower end of the domain, the user is totally novice, whereas at the higher end they are totally expert. Between these two extremes, the user can be a mixture of novice and intermediate or of intermediate and expert, except in the domain middle when the user is totally intermediate. Thus, the real-world continuity of the possible states the user can have as their experience and knowledge develops can be mapped by using a fuzzy representation with these sets.

As already mooted, the validation of the membership function of the sets lies ultimately in whether, by their use, a user can be modelled effectively and the model can be used to improve IR from the WWW. Chapters 7 and 8, which deal with the user study and its results respectively, address this issue of effectiveness.

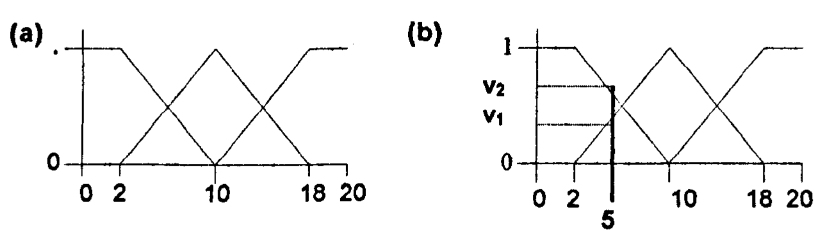


Figure 5.5 A Domain Value of 5 Produces Membership Values for Two Regions, v_1 and v_2.

In order to modify the default representations for WWW and domain knowledge and experience to produce an individual representation, the knowledge acquired via the questions has to be quantified and used to modify the fuzzy sets which constitute the default user model. This is done by relating to each question a default fuzzy representation which is the same as that shown in figure 5.4. An answer to a question, given by highlighting a circle along the sliding scale (see figure 5.3), is quantified by assigning a number to it equivalent to the count of circles from the left-hand edge of the scale. This answer is then used as the domain value of the fuzzy sets so that membership function values in the regions covered by that domain value can be calculated. That is, the answer is transformed into an indication of the degree of membership of the novice, intermediate and expert regions. Figure 5.5 displays an example of this process for a question from the WWW experience questionnaire - the question shown in figure 5.3. In figure 5.5(a) the default sets can be seen. An answer of '5' will indicate that, for that question, there is a membership of the novice region to the degree of v_2 (0.58) and of the intermediate region to the degree of v_1 (0.34) (figure 5.5(b)).

In such a way, all the knowledge acquired from the answers given in the questionnaires can be transformed into fuzzy statements of a similar form. For example, given a value of 5 as an answer to question 5 the following fuzzy statements are *both* true:

The user is novice to the degree of 0.58
The user is intermediate to the degree of 0.34.

There will be similar fuzzy statements arising from the answers to every question posed in the questionnaires. They represent and contain knowledge about the user and need to be interpreted and combined so that eventually they can be employed to modify the fuzzy sets representing the default user model. In combining them, the information they contain needs to be retained so that a method of combination is needed which will allow statements that appear contradictory to exist and be true simultaneously. In the example above, the answer given to question 5 allows two apparently contradictory statements to made: the user is novice and the user is intermediate. It is the ability of FL to represent grades of truth that allow these statements to co-exist. Another way to phrase the two fuzzy statements above would be:

ans(5) is low to the degree of 0.58
ans(5) is medium to the degree of 0.34

where 'low' corresponds to the novice region, 'medium' to the intermediate region (and therefore 'high' corresponds to the expert region).

One way to represent, maintain and propagate the knowledge contained in such fuzzy statements is to define and create rules for their combination and interpretation. This is performed by building and employing a Fuzzy Knowledge Base (FKB). A FKB, unlike a conventional knowledge base, contains fuzzy rules which are activated simultaneously and in parallel. The rules have premises (antecedents) and consequents like conventional rules but they represent partial truths. So, the following two rules:

If X is A then Y is B

If X is C then Y is D

(where A, B, C and D are fuzzy sets)

could be satisfied at the same time if X simultaneously had a degree of membership of the fuzzy sets or regions A *and* C. The consequents in the two rules, B and D, are also defined as fuzzy sets. This allows seemingly contradictory consequents such as 'Y is B' and 'Y is D' to be simultaneously true and gives meaning to the consequents. The membership values of the antecedent can be mapped to the consequent fuzzy set to give a degree to which the consequent is true. Hence, using a FKB allows the knowledge contained in simultaneously true fuzzy statements to be combined and propagated.

Example rules from the present research are shown in figure 5.6. In the developed rules, 'low' corresponds to novice, 'medium' to intermediate, and 'high' to expert.

```
R1    if ans(exp) is low and ans(time) is low
      then exp/time is very low
R2    if ans(exp) is low and ans(time) is medium
      then exp/time is medium
R3    if ans(exp) is low and ans(time) is high
      then exp/time is medium.
```

Figure 5.6 Example Rules from the FKB.

ans5 \ ans 6	L	M	H
L	VL	M	M
M	L	M	H
H	M	M	VH

Figure 5.7 Fuzzy Associative Memory (FAM) matrix for questions 5 and 6 of the WWW questionnaire.

The second of the rules in figure 5.6 is implying that if the answer to question 5 is low and the answer to question 6 is medium then the consequent 'exp/time' is medium. Note that if the answers given by the users to the appropriate questions imply values of *ans(exp)* and *ans(time)* which in turn mean a degree of membership of the low and medium fuzzy sets, respectively, then R2 will be activated and the consequent 'exp/time is medium' will be held to be true. How the degree of truth of the consequent is calculated involves a process of fuzzification and defuzzification and is discussed below.

The consequents for each rule have been developed by employing the well-accepted technique of Fuzzy Associative Memories (FAM) (Kosko, 1992). This technique, used widely in fuzzy control applications, allows the mapping of fuzzy sets to other fuzzy sets. By their use, different premises or antecedents can be associated and the effect of the states of the antecedents on a consequent indicated. Thus, by using a FAM matrix (see figure 5.7 for an example), each possible combination of the allowable antecedent states can be mapped to a number of consequent states.

The element of the matrix pertaining to the second rule in figure 5.6 is highlighted. The actual consequent states (very low, low, medium, high, or very high) for each combination of antecedents was decided intuitively from consideration of the questions. The consequent states 'very low' and 'very high' are the states 'low' and 'high' modified by the qualifier 'very. These two states can be thought of representing the user categories of 'very novice' and 'very expert'. Most software for creating fuzzy membership functions allow basic sets to be modified automatically by the application of some transformation or qualifier (Cox, 1996). This is so concepts such as 'very' for example can be represented easily. The modification in this case involves changes to the trapezoidal membership functions so that the slope is steeper.

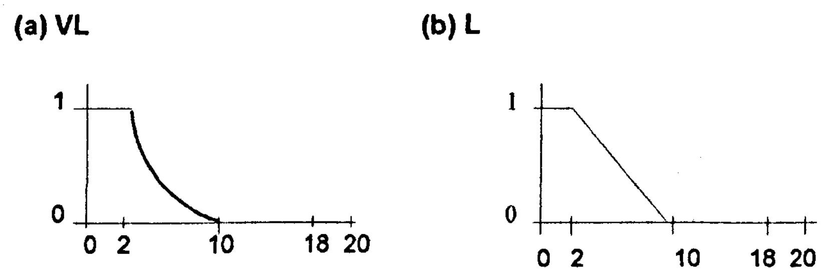


Figure 5.8 Diagram Comparing the shape of the membership functions Low (L) and Very Low (VL) – the difference has been exaggerated to show it clearly.

Figure 5.8 illustrates the effect of the qualifier 'very' on the state 'low'. The effect is to reduce the extent or truth of the membership value for the range of domain values representing the novice region except at the extremes of the domain. In effect the 'very' qualifier or 'hedge' means that, for equivalent membership truth values, a domain element must occur further to the left in the very low case.

Within the matrix in figure 5.7 there is some bias towards the antecedent state pertaining to answer 5 (experience of the WWW), as it is considered to be more important than answer 6 (time spent on the WWW). In combining other answers, heuristic decisions were made about the consequent states as well. Though, the construction of the FAM matrices and the rules has been performed heuristically, there is some support for this approach in previous fuzzy logic research (Baaklini & Mamdani, 1975; Brae & Rutherford, 1979; King & Mamdani, 1975; Van den Berg & Van Dijk, 1997).

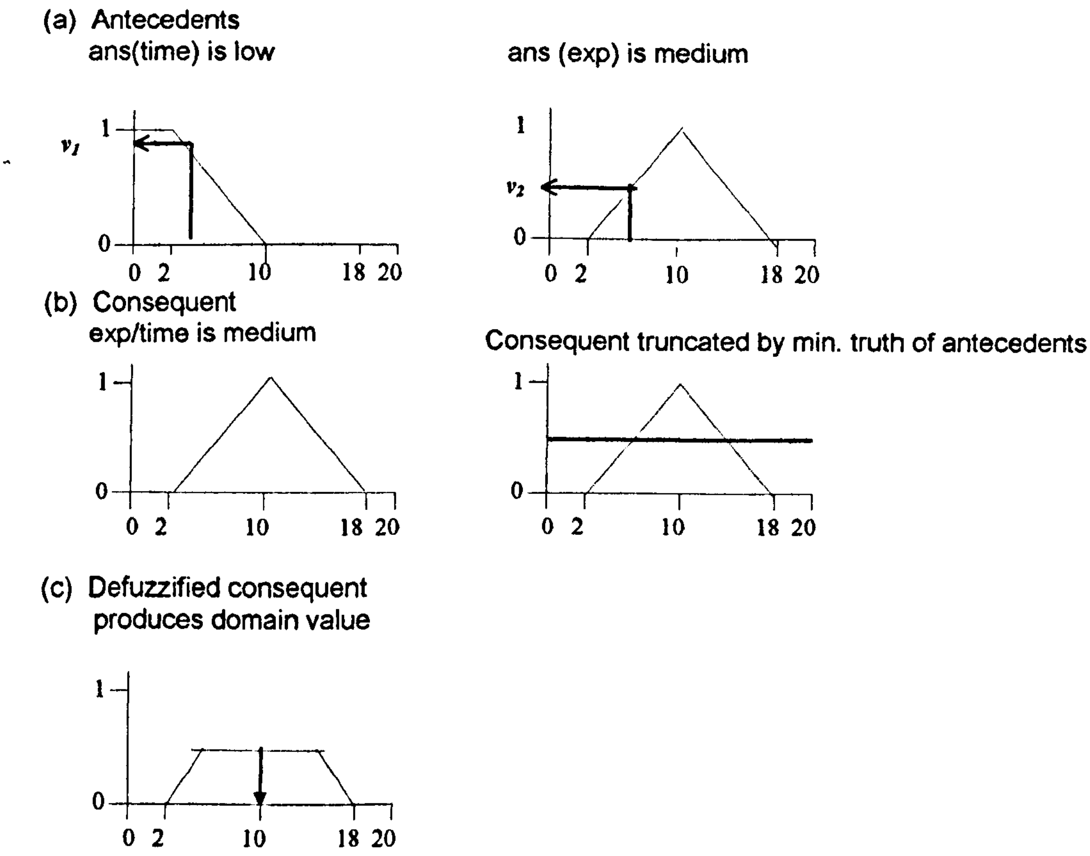


Figure 5.9 Diagrammatic Representation of the Activation of a Fuzzy Rule (Rule 2 in figure 5.6).

As already noted, the membership function values associated with the antecedents can be mapped to the consequent fuzzy set to give a degree to which the consequent is true. Figure 5.9 shows diagramatically how this is performed for the second rule given in figure 5.6. The lowest membership value of the two antecedents, v_2 (see figure 5.9(a)) truncates the consequent fuzzy subset pertaining to the region under consideration (low, medium or high) (figure 5.9(b)). This is known as the min-max inference method (Zadeh, 1965) which derives its name from the method involved. The consequent fuzzy region is restricted to the *minimum* of the predicate truths as described above. In the method, the final output fuzzy region would be updated by taking the *maximum* of any minimized fuzzy sets.

However, in this case to provide a measure to be used in the appropriate fuzzy rules, the truncated set in figure 5.9(b) is then defuzzified using the standard *composite moments* or *centroid* technique (Cox, 1996) to produce a single value (figure 5.9(c)). (See section 5.4 for discussion of the defuzzification technique used.) This value can be thought of as a measure of the consequent state 'exp/time'. In fuzzy logic, exp/time is known as a solution variable. The defuzzified value shown in figure 5.9(c) represents a possible value for this variable.

In this manner answers to certain questions are combined and then other FAM matrices are used which employ the consequent variables in their premises (see figure 5.10 for an example FAM matrix and rule in which exp/time appears as a premise).

ans4 \ exp/time	L	M	H
L	L	M	M
M	M	<u>M</u>	M
H	M	H	H

```
R5    if ans 4 is medium and exp/time is medium
      then A is medium
```

Figure 5.10 Possible FAM Matrix and Rule 5 (Highlighted in Matrix) Using exp/time as a Premise.

In these FAM matrices, the values originally produced as solutions to the consequent state are used, in a similar manner as shown in figure 5.9(a), to produce a degree of membership for the new premises. Thus, rules can be developed in this way so that the answers to all the questions can be combined, with the user knowledge they represent being retained and propagated.

Eventually, a set of rules is developed which can be used to modify the fuzzy sets representing the default user model. The whole set of rules constitutes the FKB. Thus, the FKB can be used to represent and propagate the knowledge acquired via the questionnaires, the end result being a modified fuzzy set representation of an individual user model. An example for WWW experience is given in figure 5.11.

```
  FuzzySet:      WWW experience
  Description:
1.00
0.90
0.80
0.70
0.60
0.50                                                                 ........
0.40                                                             .....
0.30                                                        .....
0.20                                                   .....
0.10                                              .....
0.00....................................
   0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
  0.00   2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.0
  Domained:          0.00 to        20.00
```

Figure 5.11 Final Fuzzy Set Representation of an Individual User Model for WWW Experience

The membership function of this final set has been determined by the answers given by the user. This can be asserted because the answers given by the users to the questionnaires were captured as values which were then used to modify fuzzy set representations. These modified fuzzy set representations were combined via the FKB to produce the final modified fuzzy set representation. In effect, the information volunteered by the user has been retained and propagated through the FMQA to the point of this final representation. So, it can be considered that this set does model the user insofar as the initial questions capture the user's knowledge and experience of the WWW. The next section considers how the information contained within the final models can be used to aid IR from the WWW - how FUM can be employed to perform query refinement.

5.4 Fuzzy User Modelling - Its Use in Query Refinement

The novel process described above produces two individual fuzzy sets, which model and represent the WWW and information domain knowledge and experience of a user. In order to use this modelled knowledge to aid IR, the sets have to be defuzzified, so that the knowledge can be interpreted and employed in query refinement.

Defuzzification or decomposition of a fuzzy set is the process of producing a scalar value from the vector space represented by the fuzzy region. There have been several methods of defuzzification developed within fuzzy logic research and which is the best to use continues to be a research issue (Cox, 1996; Ross, 1995). Hellendoorn and Thomas (1993) specified five criteria against which to measure the methods: continuity, disambiguity, plausibility, computational simplicity, weighting method.

Continuity means that for a small change in the input of a fuzzy process - i.e. in the context of this research, a small change in the answers producing a change in the shape of the model - should not produce a large change in the output - i.e., there should not be a large effect on the defuzzified scalar value. *Disambiguity* means that the defuzzified value should be unique. *Plausibility* means that the value should lie approximately in the middle of the fuzzy region bounded by the set, and should have a high degree of membership of the region. The fourth criterion, *computational simplicity*, suggests that the more time-consuming a method is, the less value it will have in a system. The fifth, *weighting method*, pertains to how the final fuzzy sets are weighted.

This final criterion does not apply in the present research, as the sets are not weighted. Of the other four, the method chosen – the *composite moments* or *centroid* method (Cox, 1996) - is at least as good as the others. It is disambiguous and plausible, and, in the context of this research, was shown to have continuity and to be computationally simple. The centroid technique finds the 'balance' point of the output fuzzy region or set by calculating the weighted mean of the fuzzy region. In other words, the technique finds a point representing the fuzzy set's centre of gravity.

It is the most widely used technique, owing to several desirable properties:

- the defuzzified values tend to move smoothly around the output fuzzy region, i.e. changes to the fuzzy set topology result in smooth changes to the defuzzified values
- it is relatively easy to calculate
- it can be applied to both fuzzy and singleton output set geometries.

Additionally, as Ross (1995) has pointed out, the method of defuzzification should be assessed in terms of the goodness of the scalar value in the context of the data available. During the development of the FUM approach, it was found that using the centroid method did produce 'good' answers. However, as Cox (1996) has noted, the current understanding of defuzzification relies more on heuristics than rigorous mathematical algorithms. This may be the best that can be achieved in a process which, in effect, is trying to represent a complex, multi-dimensional space with a single number. There is, of course, the inevitable loss of information caused by performing this process. Information which has been represented by the space is now , in effect, given a numerical value. However, currently, in order to interpret the information contained within a fuzzy set, some sort of defuzzification must be performed.

Perhaps the true test of a defuzzification method is in the interpretation of the scalar values produced. In the present research, the two final fuzzy sets produce two numbers. These numbers then become, in effect, the model of the user's knowledge and experience of the WWW and the information domain. Figure 5.12 shows the set displayed in figure 5.11 defuzzified to produce the number, in this case the value is 16.56.

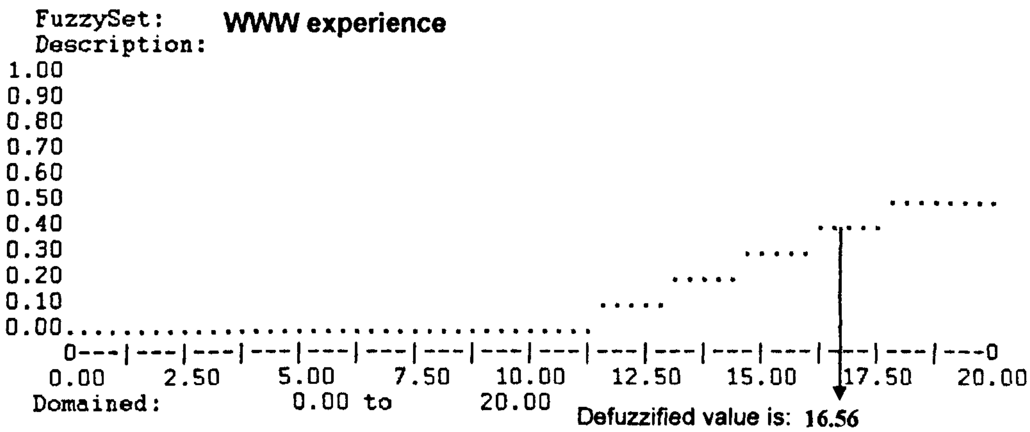


Figure 5.12 Fuzzy Set Representation showing Defuzzified Value.

The interpretation of the numbers is that the higher the number, the more knowledgeable and experienced about the concept the user is (whether it is the WWW or the information domain). This interpretation is intuitively correct in that higher defuzzified numbers are produced when the answers to the questions lie at the upper end of the sliding scale. With this interpretation in place, the numbers can then be used to develop a query refinement approach. Before this, though, the representation of the user's query must be examined and explained.

As noted in chapter 1, the general research issue was narrowed so that the focus of the present research could concentrate on one information domain - topics in AI. This was done, not only because of resource constraints on the research, but also so that the point of the research remained the application of the new approach to IR from the WWW. In developing FUM, only two models were developed, one for the WWW and one for the information domain of AI. This is not a limitation of FUM, as models for other information domains could easily be developed in the same way. However, as there is only a model for the information domain of AI, then the user's query is limited to a topic within this area.

Additionally, and again for the reasons of resource constraints and so the research should retain its focus, the user query is limited to a choice of keyword phrases representing major AI topics. This limitation was also applied so as to avoid dealing with badly formed and misspelled queries and to provide some assistance to what were thought to be AI-domain novices. Constraining the query in this way does not effect the testing of the research issue (see chapters 7 and 8) and, as will become apparent, FUM could be generally applied to natural language queries formed by the user.

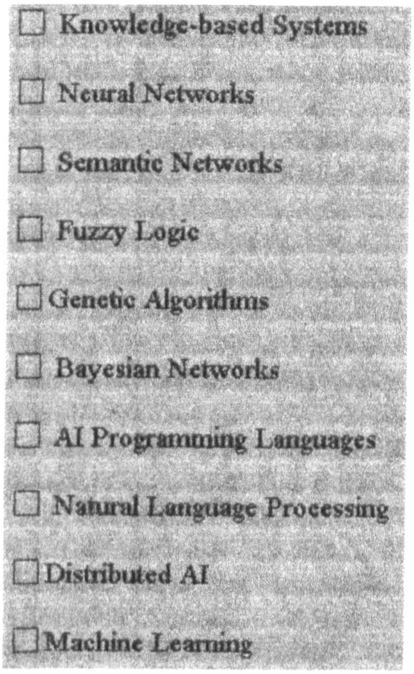


Figure 5.13 Possible Keyword Phrases that can form the Original Query.

Figure 5.13 lists the possible keyword phrases that can be chosen as a query. Thus, a keyword phrase from this list becomes a user's original query and the process of query refinement becomes a matter of using the numbers produced by the FUM to adapt this original query. As noted in chapter 3 (page 43), which reported the investigation into query refinement in IR systems, this is an active research area in which different methods are being applied (Arens *et al.*, 1996; Bell, 1992; Cusack, 1991; Han, 1994; Litteck, 1992; Sun & Yu, 1994; Vielle, 1989).

Research into the use of lexical-semantic relations in query refinement is not new and over the years researchers have found that IR performance can be improved by using this refinement approach (Salton & Lisk, 1971; Vorhees, 1994; Wang *et al*, 1985). Vorhees (1994) has noted the benefit to be achieved from applying the approach to queries that are not that detailed. These are queries that are likely to come from novice and intermediate users, the type of users the FUM approach is likely to benefit most and therefore it was decided a query refinement approach that involved lexical-semantic relations was the best type to employ.

Keyword Phrase	***Related Terms***
Neural Networks	Parallelism
	Connectionism
	Feedforward
	Kohonen
	Neural Computer

Figure 5.14 Related Terms for Keyword Phrase 'Neural Networks'.

The approach adopted in the present research is one of refining the query by expansion or 'contraction' (explained below) through recourse to a small lexicon of semantically related terms. For each keyword phrase that acts as an original query, a series of related terms have been manually created. The terms were chosen by means of an investigation of academic journals and books which cover the AI topics that constitute the original query list. Terms covering major areas related to the keyword phrases were chosen. Again, this was done with the research focus in mind, it is not to develop a semantically rich vocabulary of AI terms or to examine the issues surrounding the use of synonyms and other linguistic concepts. The approach does have similarities with the use of a thesaurus to provide related terms but in this research the chosen terms are related to the keyword in different ways -there are not just synonyms in each list. The focus is applying FUM to aid IR and the crux, here, is how the information gathered via FUM is applied. Figure 5.14 shows the keyword phrase 'Neural Networks' and the terms chosen to be employed in query refinement.

There is not a uniform number of related terms for each keyword phrase. Rather, each keyword phrase was considered separately and the number of related terms decided upon their suitability for use in refining the original query. The aspects of computational simplicity and resource constraints were also considered when the total size of the lexicon was decided, but it is of sufficient size to allow the research focus to be applied and tested.

In order to refine the query by applying the user knowledge and experience modelled via the process of FUM, the interpretation of the defuzzified numbers already mooted was used. The higher the number, the more knowledgeable and experienced the user is considered to be with regard to that aspect (WWW or information domain). The number associated with the information domain experience model is regarded to be more important, as the refinement involves changing the query by the use of related terms *within* that information domain.

If the user model values are at the low end of the possible range (0-20), i.e. less than 6, then the original query is left unchanged or 'contracted'. By contracted, the meaning here is that the query is replaced with a new phrase which is related to the original but is likely to produce general information regarding the query domain. The argument for this is that a user modelled by these low values can be considered to be novice to the greatest degree (even though they may still have a grade of membership in another fuzzy region). Therefore, the assumption is that, in the original query choice, they expect to retrieve information which generally covered the domain represented by the keyword phrase. The query refinement approach attempts to ensure this by retaining the original or by using a replacement with a more general focus.

For all other values, the original query is expanded by the addition of terms from the lexicon. The higher the values, the more terms are added. The argument this time is the converse of the above. That is, the less novice a user is or, in other words, the more knowledgeable and experienced they are, the more likely they are looking for documents which cover the AI topic in more specific and detailed ways. Adding more terms enriches the query in the sense that, when it is submitted to the existing search engine, results or 'hits' which satisfy the most terms are displayed first. The documents to which these results pertain are more likely to cover the AI topic in detail and are therefore will most probably satisfy the information need of the more experienced user. There are three levels of 'expansion' of the original query – *Slight (S), Extra (E) and Great (G)*. This point and the effect the search engine had on the implementation of the approach are discussed further in chapter 6, which focuses on the development of the prototype system used to test the approach. Before that, this chapter is concluded with a discussion of the strengths and weaknesses of FUM.

5.5 Conclusions

This chapter has examined the development of a novel combination of user modelling and fuzzy logic, designed to assist IR from the WWW. The combination, termed Fuzzy User Modelling (FUM), achieves this assistance through the refinement of user queries.

Most of the weaknesses displayed by FUM are found in the implementation of the approach. The models are fixed and predefined and are not able to be adapted after the user has answered the questionnaires. There is no use of user feedback or other techniques to adapt the models after a query is submitted and results have been returned. However, as will be explained further in chapter 6 which deals with the prototype system, the models are rebuilt each time the user accesses the system, so they remain very current. Each time users access the prototype they are asked to fill in the questionnaires again and therefore new models are constructed each time the user interacts with the prototype. This direct knowledge acquisition via the questionnaires is a strength of the approach. This is only a feature of the prototype, though, and there is no reason why the models from the first interaction could not be stored in a future implementation of FUM. Then, knowledge gained from user feedback could be used to adapt them. This feedback could be, for example, user reaction to the IR results or new user information supplied by the user.

Although there is no empirical justification of the questions asked (see Appendix B for the full questionnaires), they are designed to be open-ended, without fixed categories of answer, and are focused on aspects of the user's experience that will provide the required knowledge to build user models. Using sliding-scale questions means also that, unlike some other modelling approaches (Kok, 1991a; Rich, 1979), there is more flexibility in the models created - the user is not being forced into a distinct category through binary or fixed value answers.

Another advantage is that the models dynamically represent the current knowledge and experience of the user, and are not trying to deduce this from past actions or statements (Chang *et al.*, 1993; Sleeman, 1985).

A drawback of the approach is the intuitive and heuristic way in which the fuzzy set representations and FKB have been developed. The membership functions of the sets, the combination and content of the rules have all been heuristically developed, because the resources and focus of the research precluded the efforts required to produce evidence that might suggest certain fuzzy set functions and fuzzy rules. However, this does not mean that the heuristic approach adopted is unjustified, for there is support for this sort of approach in previous fuzzy logic research (Yager & Larsen, 1993; Mansfield & Fleischmann, 1993).

The main strength of the fuzzy logic approach is its inherent ability to capture and represent partial knowledge which, in effect, the user is providing when answering the questionnaires. The overlapping membership functions are an effective way to represent fuzzy linguistic concepts such as novice and expert, and are particularly good in encapsulating the real-world continuity between these concepts. Fuzzy rules and the FKB exhibit another advantage. They are much more flexible than traditional expert system rules, and thus can retain and propagate effectively all the knowledge gathered by FUM. This ability to propagate information is an advantage gained by using FL to represent the answers given by the users during interaction with the questionnaires. The answers could just be represented numerically and then aggregated to produce some overall value which could then be employed in the query refinement process. However, it would be difficult to interpret this value with any meaning and thus to use it effectively in modifying queries. Employing default fuzzy representations of each question and then modifying the membership functions according to the user's answer, allows that answer to be interpreted and captured as a piece of knowledge. Using the FKB to combine the modified functions allows that knowledge to be propagated, so that it can be said with confidence that the final defuzzified number, though not as detailed as the final individual fuzzy user model, does actually carry some meaning.

The query refinement process detailed in the last section is weak in that only fixed original queries are allowed and in the fact that the lexicon is by no means sophisticated or exhaustive. Allowing user-generated queries and improving the lexicon could be achieved by incorporating natural language processing and semantic network techniques, respectively. However, these two aspects are not directly related to the research issue and the present query refinement process is sufficiently developed to test the issue. Also, it can be argued that a fixed original query means that the process does not have to deal with the problems of natural language processing, such as misspellings, redundant and unused words, and parsing.

In fact, a strength of the fixed-query approach is that it allows FUM to be applied, in this instance, to modelling knowledge about the user's experience of that particular AI topic, before the query is refined. Another advantage is that the development of the lexicon can be focused upon providing terms that will be usefully employed, rather than trying to build a large and unwieldy semantic net that attempts to pre-empt any user-generated query.

Thus, despite the weaknesses discussed above , FUM does present a novel and interesting approach with some advantages. The actual models, once developed, could be interpreted and applied in different ways. In this instance, FUM is designed to assist IR from the WWW by refining queries before they are submitted to an existing search engine. This query refinement is achieved by interpreting the user model values so that they can be used to adapt the original query - taken, currently, from a fixed list of keyword phrases - with terms taken from a lexicon of semantically related phrases.

To test and validate this instance of the approach, and thus test the research issue, a prototype system was designed and implemented. This system was then employed in a user study - this study and its results are reported in chapters 7 and 8. Firstly, in chapter 6, the development of the prototype system is examined.

CHAPTER 6 - THE FUZZY MODELLING QUERY ASSISTANT: A PROTOTYPE SYSTEM

6.1 Overview

The previous chapter reported on the development of a novel approach to aiding IR called Fuzzy User Modelling (FUM). This involved the modelling of captured user knowledge and the employment of information contained within the resulting user models to refine user queries. In order to test this approach with a view to addressing the main issue of this research, a prototype system has been developed which incorporates FUM. The prototype is called the *Fuzzy Modelling Query Assistant* (FMQA) and its development and implementation are examined in this chapter.

The prototype is considered to be an assistant because, in encapsulating FUM, it attempts to assist the user in IR from the WWW by refining original queries. It is not an *agent* in the sense meant primarily within the field of distributed artificial intelligence, in which 'intelligent' autonomous agents are increasingly being researched and developed (Agentlist, 1997; Agentweb, 1998; Maes, 1994a). However, as is discussed below, it is close to Maes' (1994b) interpretation of an *interface agent*:

> '*[C]omputing programs that employ artificial intelligence techniques to provide assistance to a user dealing with a particular application... . The metaphor is that of a personal assistant who is collaborating with the user in the same work environment.*' (p.71).

In the FMQA the intention is to work with and *assist* the user in retrieving information from the WWW, through query refinement. As already discussed, the query refinement process involves amending queries *before* they are submitted to an existing search engine. Hence, the FMQA is, in a sense, sitting between the user and the existing search engine, and therefore acts as a kind of interface to the search tool. Thus, in Maes' terms it can be seen as collaborating with the user to help them deal with the search engine. This interface notion is a little undermined by the hypertextual nature of the WWW, in that any user could bypass the FMQA interface and take advantage of the web-like nature of the WWW to access the search engine directly. However, from the point of view of a user of the prototype the notion is valid, as is discussed below in section 6.3.

A high level functional diagram of the prototype is shown in figure 6.1

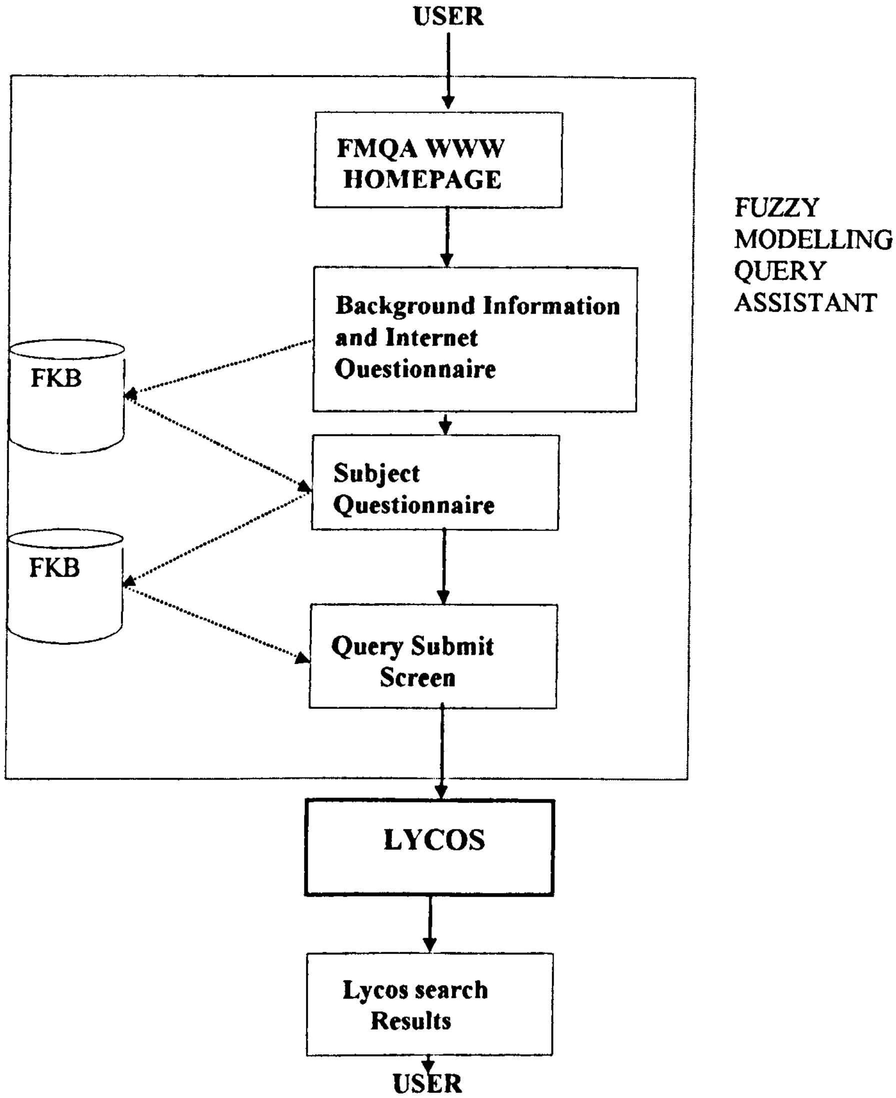


Figure 6.1 High-level functional diagram of the prototype FMQA.

As can be seen from the diagram, the user interacts with the prototype via the FMQA WWW homepage - the user interface of the FMQA consists entirely of HTML screens. The interface is discussed in greater detail in section 6.3. Two of the screens involve user interaction in answering the questionnaires, so that the user models can be built and stored. The heart of the FMQA, the fuzzy rules and models, is formed by two *Fuzzy Knowledge Bases* (FKB), indicated by the cylinders on the diagram. The implementation of the FUM approach is examined in section 6.2.

The user interacts with the prototype, not only to provide information about their knowledge and experience, but also to submit a query. The box in figure 6.1 marked 'Query Submit Screen' is where both the original and modified queries may be submitted to the existing

WWW search engine, Lycos™ (Lycos, 1996; Mauldin, 1995). This aspect is discussed further in section 6.4, which examines user interaction and query refinement within the context of the FMQA.

Lycos™ was chosen as the search engine to use as, at the time of the implementation of the prototype, it was the most complete indexed catalogue of WWW resources (Lycos, 1996). It can be argued that other catalogues, such as Alta Vista (1998a), are presently more complete. Once the query is submitted to Lycos™, control of the IR process is outside the boundaries of the FMQA (indicated by the large dotted-line box in figure 6.1). Lycos™ then returns a list of search results, in the form of a WWW page, to the user (last box in figure 6.1). This aspect of the prototype's functioning is also considered further in section 6.4. In section 6.5, the disadvantages and advantages of the FMQA are discussed, and conclusions drawn.

6.2 Implementation of the FUM Approach

To provide the functionality of the FMQA outlined above, software is needed to perform the following tasks.

- the creation and representation of the fuzzy set models and the FKB
- the capture and use of the questionnaire data
- the capture and representation of the user query
- the display of any information to the user during interaction with the prototype
- the submission of the query to Lycos™.

Section 6.3, which examines the user interface in detail, covers the capturing of user knowledge and the displaying of information to the user. Section 6.4 includes discussion of the software involved in the query refinement and submission process. This section, section 6.2, concentrates on the fuzzy modelling software employed to implement the FUM approach.

There are a number of software packages which have been especially developed to allow the building and use of fuzzy logic applications. Welstead (1994) has developed fuzzy logic systems using C/C++ and others have developed GUI-based packages to develop FKB and fuzzy engines to execute the rules (MATLAB, 1998; McNeill & O'Hagan, 1998). The software employed and adapted to suit the present research is based on the code developed by Earl Cox in his book *The Fuzzy Systems Handbook* (Cox, 1994).

The main reasons for using this software are that it provides a complete set of tools for developing fuzzy logic applications and that, as noted by Zadeh (1994), the emphasis is on dealing with linguistic variables rather than the more mathematical and traditional approach of difference and differential equations. There is a clear focus on the construction of fuzzy IF-THEN rules in Cox's development. This is ideal for the present research which employs many of these in the FKBs.

Cox's software is written in C/C++ and the kernel of the approach is that each fuzzy set is represented by a block of code called a fuzzy set descriptor block or FDB. The FDB, as shown in figure 6.2, describes the control parameters of the fuzzy set as well as a fixed-length floating-point vector containing the distribution of the membership function.

```
#include "mtypes.hpp"
struct FDB
  {
    char         FDBid[IDENLEN+1],     /* Identifier name of Fuzzyset */
                 FDBdesc[DESCLEN+1]; /* Description of Fuzzyset     */
    Ctlswitch    FDBgentype;          /* Generator Set type          */
    bool         FDBempty;          /* Is this a populated Fuzzyset? */
    domainval    FDBdomain[2],         /* Lo and Hi edges of the set */
                 FDBparms[4];          /* Generation parameters      */
    float        FDBalfacut;           /* AlfaCut for this fuzzy set */
    truthfunc    FDBvector[VECMAX];    /* The fuzzy set truth vector */
    FDB         *FDBnext;              /* Pointer to next fuzzyset   */
  };
```

Figure 6.2 The Fuzzy Set Descriptor Block.

The membership function is maintained in the floating point array *FDBvector[VECMAX]*. The domain of the set is the total allowable universe of values (represented along the x-axis in the figures in chapter 5) and the values contained in the array *FDBdomain[2]*, called *LoDomain* and *HiDomain*, are the low and high extents of the domain. In the present research the domain is 0-20. The difference between the two is known as the *Range* and for any value, *Scalar*, in the domain, its place in the membership function array, *TVcell*, is defined by means of the following expression:

$$TVcell = \frac{Scalar - LoDomain}{Range} * VECMAX$$

On the other hand, finding a domain value that is associated with a particular membership function value involves a two stage process. Firstly, a search of the membership function for the cell containing the value and then using this value to derive the domain value. This process is used during inferencing and defuzzification and is handled well in the software.

In figure 6.2, *FDBid* and *FDBdesc* are two character arrays which allow the set to be uniquely described. *FDBgentype* indicates which type of curve the membership function belongs to - in the present research these are trapezoidal and triangular, but other curves, such as Gaussian and bell-shaped, are allowed. *FDBempty* is an integer flag to indicate whether the set is empty or not. The elements in the array *FDBparms* are reserved for values that can be used to generate the different sorts of curves. **FDBnext* is a C++ pointer that allows sets to be linked.

FDBalfacut is a floating-point number reserved for the alpha-cut value. It is possible that, unlike the sets used in the present research, the non-zero part of the fuzzy set does not extend across the entire domain. The non-zero part is then known as the *Support Set* of the fuzzy set. An alpha-cut is a threshold restriction on the domain based on the membership grade of each domain value. The alpha-cut restricts the set to contain all the domain values that are part of a fuzzy set at a minimum membership value equal to alpha-cut value. An alpha-cut value set to zero defines the support set for a fuzzy set. In the present research, the alpha-cut is set to zero. An example of a non-zero alpha-cut application is Laffan's use of fuzzy sets to model groundwater discharge – a fuzzy GIS (Geographical Information System) combines topographical data with fuzzy models of factors such as lowness and curvature to predict where discharge will and will not take place. A threshold level represented by the alpha cut is used to divide the final model domain values between non-discharge and discharge values so as to make the data easier to interpret (Laffan, 1996).

The software also includes programs for the graphical display of the membership functions. These have been used for appropriate figures throughout the thesis, e.g. figure 5.4 on page 70. This allows the development of fuzzy sets and the effects of the fuzzy rules on these sets to be monitored graphically.

The implementation of the FUM approach crucially requires the development of FKB containing a number of rules. Consider the rule shown in figure 6.3.

```
R7   if ans1 is high and ans2 is low
           then aiexp is medium
```

Figure 6.3 Rule 7 taken from the FKB for the Subject Domain Model

This rule is taken from the FKB developed to handle the information delivered to the prototype via the subject domain questionnaire. To encode this requires the creation of 'Low', 'Medium' and 'High' fuzzy sets to represent the two premises' and the consequent's fuzzy regions. Variables are also needed to represent *ans1, ans2* and the consequent, *aiexp*. The latter variable is known as the solution variable. As discussed in chapter 5, the answers to the questions are converted into numerical scores - these would be stored in the variables *ans1* and *ans2*. The min-max inferencing method is then used to find the minimum membership function value of the two premise fuzzy sets. This value is used to truncate the consequent fuzzy set and then, the truncated consequent is defuzzified using the centroid method to find a possible value of the solution variable, *aiexp*. The software has the required functionality to perform these tasks, including the inferencing and defuzzification. Using the software functions and programs allows the rules of the FKB to be encoded and processed simultaneously during a user interaction with the prototype.

However, this particular implementation considerably extends and builds upon the use of the software in the examples provided (Cox, 1994). Although Cox discusses the use of the FAM matrix technique to develop fuzzy systems, he does not cover an implementation in any detail and does not provide an example. Also, in the examples he does provide, the maximum number of rules developed in any system is 5, whereas the prototype employs over 100. Additionally, the use in the FMQA involved considerable re-engineering of the code and implementation upon a different platform. Thus, the FMQA development represents a significant extension of the initial software. The software development is discussed further in Appendix I.

In implementing the prototype, a logfile approach was adopted. As an interaction with the user takes place, the fuzzy rules are written and the fuzzy sets graphically displayed in the logfile so that an inspection will reveal how the final fuzzy set representations of the user models were arrived at. An example logfile is reproduced in Appendix C.

To perform the interaction and build the fuzzy models, the information provided by the user through the questionnaires needs to be captured and made available to the FUM part. This is done by embedding the FUM software in other programs called CGI-scripts (Common Gateway Interface). CGI-scripts are programs which can be designed to accept information entered via a user interface and then take this information and process it in some way. Often, the script resides on a computer remotely connected to that with which the user is interacting. In that way, a script program can be made available to any instance of the user interface at any time. This is the situation with the FMQA so that multiple users of the prototype can access it simultaneously. Also, the script program can, as part of the process, return information to the user's computer, and this can be information which is dependent upon the information originally captured by the script (Perl, 1998).

Used in conjunction with the WWW, these scripts are designed to capture information entered into special HTML pages called *Forms*. Forms are used in the FMQA for this reason and are thus part of its user interface which is discussed next.

6.3 The User Interface

The user interface of the FMQA consists of a number of WWW pages or screens designed to provide straight forward and easy access to the prototype. HTML is the language used to create and encode pages for the WWW. It is a subset of the more general SGML (Standard Generalised Markup Language) which is a very comprehensive language designed for producing full formatted text and graphics and to be used by typesetters and printers in the publication process. HTML is not as rich syntactically as SGML but can be used very successfully to produce very complex formatted text and graphics. The developers of the WWW (Berners-Lee *et al.*, 1992) chose HTML because it is a lot simpler to understand and use and it allows pages to be produced and linked together as hypertext easily. HTML works by the use of *tags* which surround particular parts of a page and indicate how that part should be represented.

Observing figure 6.4, which reproduces as an example a piece of HTML code, along with figure 6.5 which displays how the code would look when viewed through a WWW browser, it can be noted that the tags *<I>* and *</I>* encapsulate italic text.

```
<HEAD>
<TITLE> IIFQA - Accessing the System</TITLE>
</HEAD>
<body>
<a name="Top"></a><H1><B>Fuzzy Modelling Query Assistant</B> </H1><P>
<H2><I><B>Accessing the System</B></I></H2><P>
<HR><P>
<H3><I>Introduction</I></H3>
Thank you for accessing the Fuzzy Modelling Query Assistant.  This tool is
designed to aid and refine queries to the WWW search tool
Lycos<SUP>TM</SUP>.  The tool aims to improve information retrieval
performance by employing fuzzy logic and user modelling techniques.<P>
...
Please make a choice by clicking on the highlighted text.<P>
<UL>
<LI><A HREF="http://146.227.155.136/~gjmooney/preques2.htm">Use the
Assistant</A>
<LI><A HREF="http://146.227.155.136/~gjmooney/iiqahome.htm">Return to
Assistant homepage</A>
</UL>
```

Figure 6.4 A Piece of HTML Code.

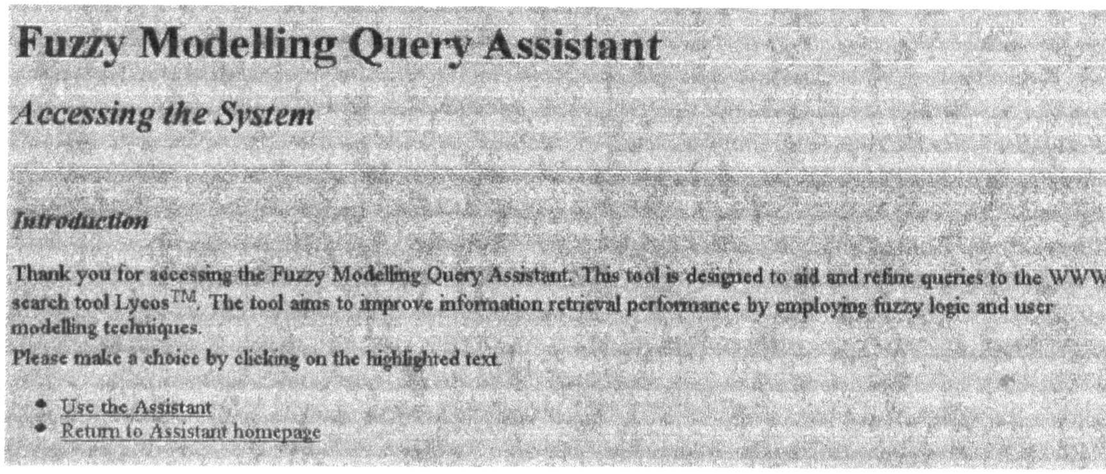


Figure 6.5 The Code in Figure 6.4 Displayed with Netscape.

In this way text can be displayed as italic, bold, underlined, and quite elegant and complex pages can be produced. Also note the tags *<A HREF =...>* and *</A>* surrounding the phrase 'Use the Assistant'. This is a hypertext link in HTML, a pointer to another WWW page. The concept is that a WWW user, clicking on the phrase, would be presented with the page that has the address given to the right of the equality sign. Thus, HTML pages can be quite easily linked together. Rather than hierarchical links, the nature of the WWW is a web-like structure which means that a page can have links to a number of pages simultaneously.

In this manner, the FMQA interface is a WWW interface consisting of a number of HTML pages linked together. The main reason for using a WWW interface for the prototype is not because of the ease with which a usable interface can be constructed, it is owing to the nature of the research issue and how it has been focused.

At the heart of the research issue is the notion that application of the prototype FMQA to the problem of IR from the WWW will improve upon the results obtained in comparison with those from the use of existing WWW search engines. Constructing the FMQA with a WWW interface avoids having to account for any perceived or real differences and biases between types of interface, which may affect IR results. Also, as previously discussed, the focus of the research is on modifying queries before they are submitted to an existing search engine. So, it makes sense to build a WWW interface to the FMQA to ensure a seamless and smooth link between it and the existing search engine, Lycos^{TM}.

The WWW screens or pages that form the FMQA interface are reproduced in Appendix D in detail. There are four main screens. The first of these, represented by the box marked 'FMQA WWW Homepage' in figure 6.1, is shown partially in figure 6.5.

Information Sought	**Possible Answer**
Name	User decision
Email	User's email address (if have one)
Age	18-21, 22-25, or 25+
Course	Course attended if DMU undergraduate
Background	UK, EU, Overseas or Other (asked to explain other)
Position	Undergraduate, Postgraduate, DMU Staff, Other (asked to explain other)
Using the Internet	Indicate a measure on 20-point sliding scale between end points *Novice* and *Expert*
Internet Experience	Indicate a measure on a 20-point sliding scale between the end points *Never* and *More than 12 months*
Time using the Internet per week	Indicate a measure on a 20-point sliding scale between the end points *Never* and *More than 20 hours*
Opinion of the Internet in terms of difficulty	Indicate a measure on a 20-point sliding scale between the end points *Difficult* and *Simple* (Can also indicate *Not Applicable (NA)*)
Opinion of the quality of information on the Internet	Indicate a measure on a 20-point sliding scale between the end points *Very Poor* and *Very Good* (Can also indicate *NA*)
Previous use of a WWW search engine	Yes or No
Opinion of WWW search engine experience	Indicate a measure on a 20-point sliding scale between the end points *Novice* and *Expert*
Opinion of the quality of WWW search engine IR	Indicate a measure on a 20-point sliding scale between the end points *Very Poor* and *Very Good* (Can also indicate *NA*)

Table 6.1 Summary of the Information Sought in the first FMQA Questionnaire.

As can be noted from figure 6.5, the screen basically consists of a few lines of text which describe briefly the prototype and two hypertext links. To move to the first questionnaire and thus begin interacting with the FMQA, a user must choose the link 'Use the Assistant'. Doing so will mean the screen represented by the next box in figure 6.1 'Background Info. and Internet Questionnaire' is displayed.

In this screen, the user must answer background questions about their gender, age-range, status and they also encounter the first set of sliding scale questions designed to elicit information about their WWW experience. This is the first screen to use an HTML form. Table 6.1 summarises the information sought in the form.

HTML forms, as may be expected, are similar to real, everyday forms, in that they can be constructed so that the user can enter information into large boxes, indicate preferences by highlighting checkboxes or circles and answer questions in a similar manner. The information on a completed form is then submitted for transcription to a CGI-script (discussed in section 6.2).

The submission of a form is usually accomplished by the clicking on a button at the bottom of the form. In the two forms which are used to function as the questionnaires in the FMQA, the users submit their answers by clicking on a button marked 'submit questionnaire' at the bottom of each. There is also another button marked 'clear questionnaire' which allows the user to reset the form to its original state, should they wish to change any of their answers.

HTML has a special set of tags for the construction of forms which involves stating explicitly where the appropriate cgi-script program is to be found. In the FMQA the two script programs associated with the questionnaires have the programs containing the FKB embedded within them. Therefore, in interacting with the FMQA by filling in the questionnaires, the user is interacting with the fuzzy rules and sets. The information contained in the first form is passed by the script program to these embedded programs so that the individual model of the user's WWW experience can be constructed (dotted line on figure 6.1).

The user is then presented with the 'Subject Questionnaire' screen which contains the questions on the subject domain. It is also at this point that the user must choose their original query from the list of AI topics. Table 6.2 summarises the information sought in this form.

Information Sought	Possible Answer
Choice of original query	Pick one from list of 9 possible topics
Experience of AI in general	Indicate a measure on a 20-point sliding scale between the end points *None* and *Very Much*
Experience of topic chosen as original query	Indicate a measure on a 20-point sliding scale between the end points *None* and *Very Much*

Table 6.2 Summary of the Information Sought in the second FMQA Questionnaire.

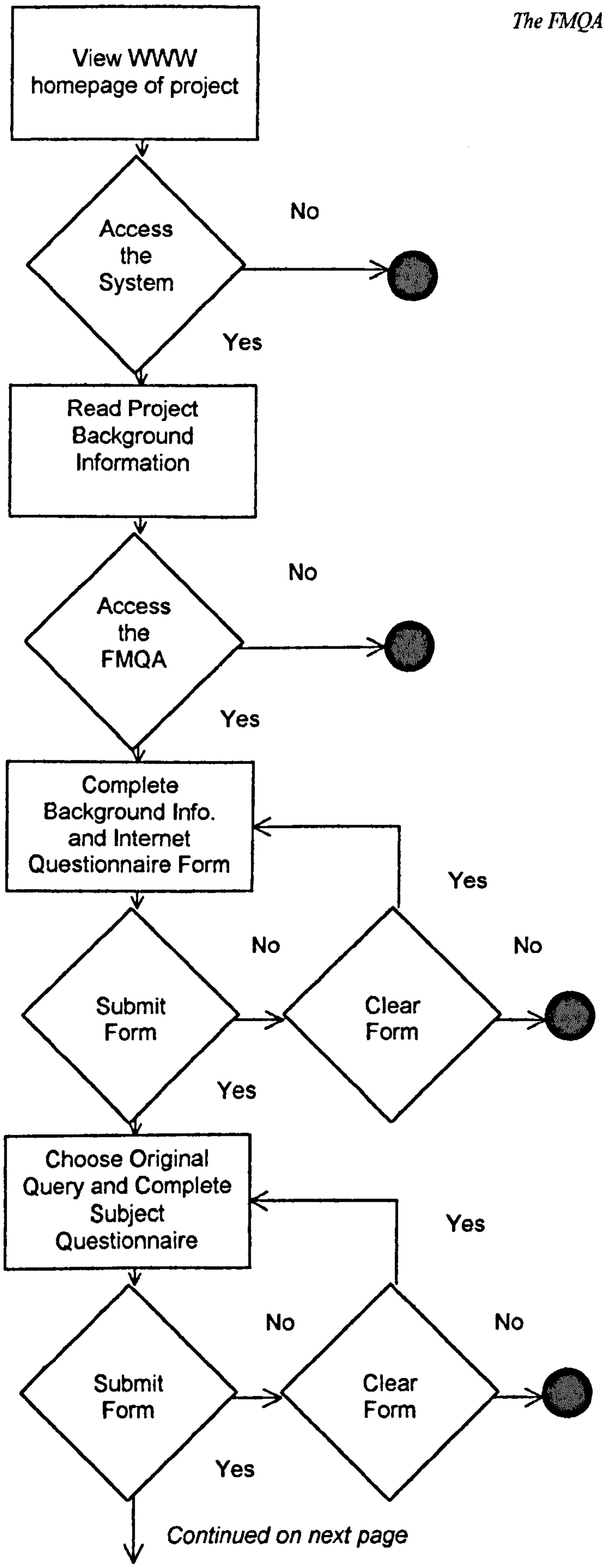
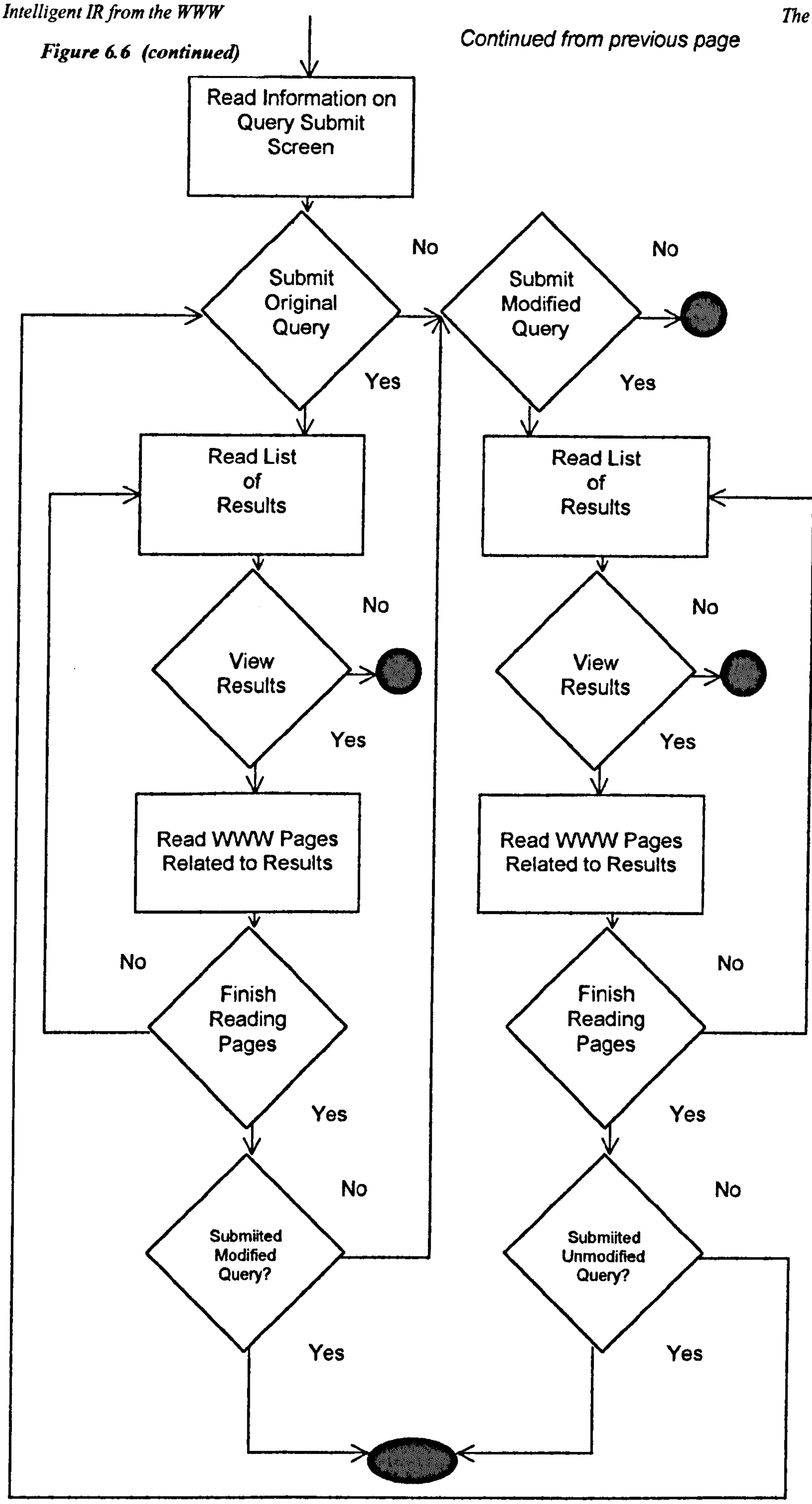


Figure 6.6 Diagram showing Processes and Decisions Involved in a User Interaction with the FMQA.

Continued on next page

Figure 6.6 (continued)

Continued from previous page



This choice and the questionnaire answers are submitted to another cgi-script program in which is embedded not only the FKB programs - which form the rules and sets needed to model the user's subject experience, but also the programs which perform the query refinement.

The information passed with the next and final main screen of the FMQA interface, the 'Query Submit Screen', includes the original and modified queries. This screen also contains an HTML form but the user does not enter any information. Rather, there are two buttons - one marked 'Submit original query', the other marked 'Submit modified query'. The user is then able, by clicking on one of these buttons, to submit either of the queries to Lycos™.

Figure 6.6 presents a diagram showing the processes involved as the user interacts with the FMQA. As can be seen in the diagram, the user has to read a number of WWW pages (or screens on the computer) and has a number of decisions to make during an interaction with the prototype. Any decision which leads to a black circle in the diagram means the end of the interaction or session with the FMQA. After submitting a query, the user moves outside the realm of the FMQA (as indicated by the large dotted-line box in figure 6.1). The user can read the list of results returned by Lycos and then view the WWW pages related to any item. This is an iterative process and only ceases when the user is finished with looking at the pages. In order to view the WWW pages of more than one item in the result list, the user must employ the 'back' button on the WWW browser – this facility is part of the functionality of the browser and nothing to do with the design of the FMQA. However, it was useful in the design of the user study (see chapter 7).

This aspect of the interaction with the FMQA and the query refinement process itself are discussed in more detail in the next section.

6.4 User Interaction and Query Refinement

As discussed in the previous section, the user interacts directly with the prototype FMQA via the WWW forms that are part of the interface. These forms allow the user to enter information and answer sliding-scale questions, and the data collected by these forms are used to build the two individual user models. The models represent the user's knowledge and experience of the WWW and the query subject domain. It was argued in detail in chapter 5 that in order to use the information encapsulated within the models, the final fuzzy sets have to be defuzzified. This defuzzification produces a numerical value that is a measure of the solution to the set. The two numbers are used in the query refinement process to amend the original query.

In the prototype, the first of these numbers - representing a measure of the user's WWW experience and knowledge - is passed to the WWW screen containing the subject domain questionnaire and initial query form ('Subject Questionnaire' box in figure 6.1). When the user fills in the questionnaire and chooses the original query, this first number is also passed to the cgi-script program to which the information from the form is sent.

This second cgi-script program not only builds the user's subject domain model, defuzzifying the final set to produce the numerical measure, but also passes both values and the original query string to a subprogram which performs the query refinement. The query refinement is carried out in the manner discussed in chapter 5. The FMQA can modify an original query in four different ways - *C(Contracted)*, *S(Slight)*, *E(Extra)*, *G(Great)*. Contracted (C) means the original query is replaced with a more general term. In the other cases the query is expanded by the addition of terms, and the greater the expansion, the more terms are added. The user is not exposed to this aspect of the functionality of the FMQA - the process is hidden from view.

When the process is complete, both the original query string and the modified query string are passed back to the cgi-script program which returns to the user the WWW screen represented by the box marked 'Query Submit Screen' in figure 6.1. As noted already, this screen also contains a form which consists of two buttons - one marked 'Submit Original Query', the other marked 'Submit Modified Query'. Behind these buttons, the original and modified query strings are incorporated into URLs, so that depressing one of them means that particular query will be submitted to the existing search engine Lycos™.

After a query is submitted (box marked 'LYCOS' in figure 6.1), procedural control passes outside the realm of the FMQA and over to the search engine. It returns to the user a WWW page consisting of the 'top-ten' 'hits' or results which closely match the query string (see figure 6.7 for an example). These results are a list of URLs which the user can click on in order to access the WWW pages to which the URL points. These pages will contain information which in some way matches the search string. Lycos™ uses an algorithm to index WWW pages and when a query is submitted to the search engine, it is a process of matching the query string to the indexed terms.

Lycos™ works by identifying the 100 terms or words most related to the WWW page being indexed. Combining these terms with the titles, header text (information contained within the <HEAD></HEAD> tags in a WWW page), and an excerpt of the first 20 lines (or 10% of the document), it creates an abstract which can be displayed with the list of results. An inverted file indexing method (Salton, 1989) is used to index the portions the WWW pages thus collected. An inverted file is a list of all the occurrences of words within the database or search engine catalogue. For each word in the database, the search engine keeps a list of the documents containing the word, and often a list of the positions where the word occurs within the documents.

During IR, Lycos™ ranks the documents matching the query according to a relevance score based on these features:

- How many of the query terms are contained within the document
- How frequently the query terms are used within the document
- How close the query terms are to each other in the document (proximity)
- Where the query terms occur within the document (position)
- How closely the query terms match the individual words

For example, in a search for 'Viking Mars Lander', a document containing all three words would rank above one containing only two.

As Mauldin (1997) notes, all major WWW search engines operate in a similar manner to the one described above. They may differ on the amount of the document they index – AltaVista (1998a), for example, indexes the whole WWW document – and in the factors used to rank the retrieved documents. However, these differences are irrelevant for the purposes of this research and the action of the FMQA.

The FMQA is designed to act on the query *before* it is applied to the search engine and this action is not dependent on the workings of the search engine. As will be seen in chapter 7, the performance of the FMQA is tested against the performance of the search engine alone. This testing could have been performed against any search engine.

Lycos™ employs the ranking scheme described above when deciding the order of results displayed to the user (Lycos, 1996). A user is able to scroll down to the bottom of the list and access the next 10, and following groups of 10, results, should they wish.

1) **HARMONY - CP94-202**
HARMONY - CP94-202 Application of **Hardware Based Fuzzy Logic Controller** for Adaptive Pacemakers HARMONY - CP94-202 keywords: **fuzzy logic**, pacemaker ...
http://www.twente.research.ec.org/esp-syn/text/cp94-202.html (3k)
[100%, 4 of 6 terms relevant]

2) **Fuzzy Logic Software Available for A-B Control Coprocessor**
Fuzzy Logic Software Available for A-B **Control** Coprocessor **Fuzzy Logic** Software Now Available for Allen-Bradley **Control** Coprocessor Request FREE Product ...
http://www.ab.com/events/pressrel/pre-95sept/fuzzsoft.html (4k)
[76%, 3 of 6 terms relevant]

3) **FAQ: Fuzzy Logic and Fuzzy Expert Systems 1/1 [Monthly pos**
FAQ: **Fuzzy Logic** and **Fuzzy** Expert Systems 1/1 [Monthly posting] - [6] What is **fuzzy control**? [6] What is **fuzzy control**? Date: 17-MAR-95 The...
http://www.cs.cmu.edu/afs/cs.cmu.edu/Web/Groups/AI/html/faqs/ai/fuzzy/part1/faq-doc-6.html (3k)
[74%, 3 of 6 terms relevant]

4) **Fuzzy Control**
Fuzzy Control Fuzzy Control Researchers: Karl-Erik rz n, and Mikael Johansson The impact of **fuzzy logic** on design of controllers has increased ...
http://control.lth.se/research/fuzzy.html (2k)
[74%, 3 of 6 terms relevant]

Figure 6.7 Partial WWW page of IR results or 'hits' returned by Lycos™.

To return to the FMQA the user must use the 'back' button on the WWW browser. This is a button which forms part of a 'toolbox' of functions and which allows a user to return to the previous WWW page or screen that they were viewing. Returning to the FMQA's 'Query Submit Screen' by using the 'back' button allows the user to submit both the original and the modified query to Lycos™ in turn, and then compare the results. In fact, the users taking part in the user study of the FMQA were asked to do this, in order that the research issue could be addressed. The user study is reported in detail in chapter 7. Firstly, this chapter is concluded with a discussion of the strengths and weaknesses of the prototype FMQA, and how its construction influenced addressing the research issue.

6.5 Conclusions

This chapter has examined the development of the FMQA - a prototype system designed to implement the FUM approach and apply it to IR from the WWW, so that the research issue can be addressed.

Two strengths of the FMQA lie in its simple construction and its WWW interface. Many interfaces to IR tools are quite sophisticated and prospective users are influenced adversely by the need to learn how to navigate and interact with the tools' interfaces. Evidence from the user study of the FMQA (reported in chapter 7) indicates that users find it easy to understand and use. Building the FMQA with a WWW interface ensures that it has compatibility with the interfaces of existing WWW search tools, and thus avoids any bias - in IR results or the user perception of them - being introduced by having different types of interface. The FMQA's interface is also transparent, in that the user is protected from and has no direct contact with the model building and query refinement processes. Another advantage of the current implementation is the use of a list of AI terms as the original query. This avoids the usual problems that may arise when dealing with natural language queries, such as misspellings and parsing.

The user interaction via the questionnaires produces very timely user models which have the strength of representing the user's current experience and knowledge, though this means that the user must fill out the questionnaire each time they begin a session with the FMQA. Additionally, although the models are lost at the end of an interaction, an astute user is able to modify their questionnaire answers (and thus modify their models) by employing the 'back' button on the WWW browser. The 'loss' of the models at the end of the interaction is only a feature of the prototype and any future implementation of the FMQA could easily retain models for use in future user sessions by storing the final fuzzy representations.

A criticism of the current prototype that could be levied against it is that the interface (or link) between the FMQA and Lycos™ is poorly managed. It would be better if the results being returned by the search engine could be collected and manipulated before being displayed to the user - to afford an easy comparison between the different searches, for example. However, to enable this would have meant more resources being spent on development than were strictly necessary to examine the research issue and may have proven difficult to achieve, considering that Lycos™ is a commercial, proprietary system. The user is still able to move between the FMQA and Lycos™ via the functionality of the WWW browser, i.e. using the 'back' and 'forward' buttons, once they have made an initial search using the FMQA.

Though the limitations of and the constraints applied by the current prototype version of the FMQA reduce its applicability as an IR tool, they do not prevent it from being used to address the research issue, which is the primary aim of its development. The research issue, as focused by the examination reported in chapter 5, can be expressed as the following question:

> Whether the application of a prototype system, developed using Fuzzy User Modelling (FUM), can improve the IR results from the WWW, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics?

The development and implementation of the prototype FMQA discussed in this chapter means that the research focus can be narrowed and expressed as the following question:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics?

Addressing the research issue by answering this particular question will not mean the general issue is not examined. Rather, an affirmative answer will satisfy the research issue for a particular set of criteria. The FMQA, though simple and not sophisticated, is designed to allow the research issue to be tested. In order to test the issue an examination of the performance of the FMQA, in comparison with Lycos™, is required. To conduct the examination, a user study of the FMQA was performed and this is reported next, in chapter 7.

CHAPTER 7 - THE USER STUDY:
AN INVESTIGATION OF THE PERFORMANCE OF THE FMQA

7.1 Overview

The previous chapter examined the development and implementation of the prototype FMQA. As stated in chapter 6, the research issue can be expressed as the following statement:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics?

In order to test this issue, a user study of the FMQA's performance was required. This was a study of real WWW users interacting with the prototype in the manner discussed previously. This chapter reports on the design and conduct of such a study. The scope of the user study was designed to allow the research issue to be addressed.

The user study was initially driven by a *User Studies Design Protocol* - a working document, reproduced in Appendix E, which provided a starting point for the development of the parameters of the study. A comparison between that document and the report of the actual study in this chapter will reveal that, whilst the basic structure of the study was adhered to, during development the implementation details did change.

The primary aim of the user study is to provide a means of addressing the research issue. Another objective is to ascertain any differences in results that may exist between different types of user. The conduct of the user study was influenced by previous research which is discussed in section 7.2. Following that, in section 7.3, the aims are described in detail.

The whole focus of this research, embodied in the statement above, is targeted on those WWW users termed 'novice' and 'intermediate'. This factor was paramount in deciding upon suitable users to take part in the study. This process of isolating and choosing suitable user groups took account of the definitions of novice and intermediate discussed earlier in chapter 5 (page 66), and this process is examined in section 7.4.

To make sense of the study and address the research issue requires an evaluation of the user study. Performance measures need to be defined by which the user study may be evaluated. The definition of these measures, and in what manner the corresponding data were collected, is examined in section 7.5. The actual analysis of the collected data, and the results which may be inferred from the study, is left to be covered in the chapter 8.

Once the groups had been chosen and the performance measures defined, the user study could be performed. The timetable and conduct of the study are described in section 7.6. Though the FMQA is a WWW prototype tool, it is not password protected, and thus is accessible to anyone with a WWW browser. However, certain aspects of the study were initially controlled in order to reduce any bias that might be introduced. Criteria such as time accessed, time spent per session and environment are examples of these and are examined in greater detail in section 7.6.

Chapter 7 concludes with an exploration of the advantages and disadvantages of the user study, including what effects, if any, the process might have on any future studies of the FMQA.

7.2 Background

An early study of on-line IR compared four systems capable of interacting with the on-line medical database MEDLINE (Bonham & Nelson, 1988). The study consisted of: a direct comparison of a number of system features, such as cost; comparison of results from a controlled search performed by the authors; and a user study based on observations, questionnaires and interviews with a sample of 11 library users asked to perform their own searches using each system. The main point of the study was to provide support for a decision to select one of the systems for use in a library. This led Bonham and Nelson to note that the:

> "...subjective opinions of library patrons who are potential users and their own observations were valid methods of investigation." (p.26)

Data from the searches, such as query string, retrieved results, time spent searching and cost (MEDLINE was a commercially-based database), were recorded. Questionnaires and interviews were used to gather users' reactions and opinions. Some of the questions are in the form of sliding scale, similar to but less complex than those which are part of the FMQA. Users mentioned retrieval capability as the most important factor but were also influenced by perceived cost whilst on-line. The authors managed to make a choice, despite the limited size of their study, but noted that each system had commendable features. Perhaps the small size of the study prevented conclusive evidence from being collected.

Many studies of the performance of intelligent IR systems have concentrated on measurements of statistical indicators such as 'precision' and 'recall', in which the system under study is compared with another, against a fixed document set (Belkin *et al.*, 1993; Callan & Croft, 1993). These indicators rely on a notion of relevancy (Salton, 1989) and as Tague and Schulz (1988) point out, the relevance of a retrieved document is hard to define and has been defined in different ways, for example with regard to the user, an 'expert' or the original query. Hull (1993) points out that a successful IR process returns documents that satisfy the user's information need and then argues that two assumptions are needed before the retrieved results can be used to evaluate the system under study. Firstly, that all documents can be judged to be relevant or not and secondly, each relevant document is equally important and any value placed on it does not depend upon how many other relevant documents are available. Although reasonable assumptions, these can be questioned, especially when the focus of an evaluation is on the user and their satisfaction. Precision and recall are usually defined as (Hull, 1993):

$$\text{Precision} = r/n \qquad \text{and} \qquad \text{Recall} = r/R$$

where r = No. of relevant documents retrieved
R = Total no. of relevant documents
n = Total no. of documents retrieved

From these definitions, it can be seen that it is difficult to apply these performance criteria in any evaluation of a tool looking to improve IR from the WWW, as it is nigh on impossible to know the total number of relevant documents within a document corpus that is constantly changing. One could test the performance of a WWW-based tool against a fixed subset of the WWW corpus, to which some notion of relevancy had been applied, and then the criteria above could be used. It is doubtful that any results from that kind of evaluation which showed improvement in terms of precision and recall would support a general statement concerning the tool's abilities with regard to the whole WWW.

Instigated in the 1990s, an important series of conferences known as the Message Understanding Conferences (MUCs) (Lehnert & Sundheim, 1991; MUC, 1995) involved performance evaluation of IR systems based on blind test sets of documents. Evaluations used measures of precision and recall based on an information task to extract from the test database of 1700 random documents all those considered relevant to the task, the notion of relevance being predefined. The MUCs led to increased research in the field of intelligent IR, including the development of INQUERY (see chapter 3) and the creation of the TIPSTER collection (Callan & Croft, 1993).

A stated aim of the TIPSTER project is to make intelligent IR research effective through the devising of standards and metrics by which to evaluate systems and test different approaches (Harman, 1992). It is both a collection of documents and a set of sophisticated *intelligent* methods of gathering and representing information. This includes: user interaction to develop synonyms in query expansion; support for diverse interfaces such as forms and command-line; automatic recognition of phrases; and relevance feedback techniques which use 'case-based' reasoning. The architecture of the TIPSTER collection includes representing information needs not as standard Boolean queries but by frame-like data structures called *Topics*. These include fields such as 'domain' which represents the basic subject matter of the search.

The task of creating and coding the topics is a key obstacle in this approach and a considerable overhead as the TIPSTER collection grows. Also, though large, the collection is only 750,000 documents (Callan & Croft, 1993) and it is difficult to see how the approach could be effectively applied to the WWW. There is a question over the measures of performance as well. Precision and recall are used but there is no agreed standard by which to define relevance (Callan & Croft, 1993). For the TIPSTER collection, relevance judgements were obtained by having trained analysts evaluate the 'best' documents retrieved for each topic by a variety of different IR systems. Judgements reported at the first Text REtrieval Conference (TREC) conference are often used (Harman, 1992). INQUERY has also been evaluated against the TIPSTER collection but the tests were limited and the results inconclusive (Belkin *et al.*, 1993).

Turtle and Croft (1991) have evaluated the effectiveness of a IR model based on inference networks (see Chapter 3) by employing it to build networks for two test collections for documents. They describe retrieval performance in terms of precision and recall and focus their research hypothesis in terms of the inference network model performing as well as other probabilistic models. Their results indicated that there is some improvement in precision overall. Another aspect of their research hypothesis, that the use of multiple query formulations and search strategies, e.g. Boolean and probabilistic, will improve upon the IR performance of single query representations, is supported also by their results.

Belkin *et al* (1993) also looked at the effect of multiple query representations on IR system performance, using the TREC database collection and the INQUERY inference network search engine (see Chapter 3). Five, independently generated Boolean query formulations for the TREC topics were produced by the ten different expert on-line searchers. These different formulations were grouped and the groups (and combinations of them) were used as search queries using INQUERY. The results, again based on precision and recall measures, showed that progressive combination of query formulation leads to progressively improving IR performance.

Gauch and Smith (1991) quote research (Blair & Maron, 1985; Borgman, 1986; Fenichel, 1981) which shows that users of IR systems experience difficulties, particularly with developing successful search tactics. They developed a prototype expert system to assist searching by augmenting the capabilities of users through automatic query reformulation and automatic ranking of the retrieved results. Hence, there are some similarities with the FMQA, but the actual AI technique used and the way it is applied are different. They compare the performance of the prototype with that of the user working alone and the user working with the aid of an on-line thesaurus.

Gauch and Smith evaluated the system by using measures of what they term *search effectiveness* and *search efficiency*. Search effectiveness is encapsulated in terms of precision and recall measures involving a database in which the number of relevant documents was known. Search efficiency is defined as the number of Boolean queries the user entered for each of the search questions, and the amount of time spent searching for relevant passages which will answer each question (IR tasks are framed as search questions to be satisfied). The authors also wanted to judge the system in terms of whether it could rank IR results in decreasing order of relevance. This aspect was evaluated by comparing the order of appearance of relevant passages after they had been ranked by the system with a random order of appearance.

Only 12 users took part in the study, so it was somewhat limited. Search effectiveness was not improved overall but they do report a slight improvement in precision. The on-line thesaurus had no bearing on IR improvement. Search efficiency was significantly improved and the third aspect of the evaluation was also supported. Answers to a post-use questionnaire revealed users' opinions of the system's features. Automatic addition of terms was most liked whilst system slowness was, not surprisingly, the most disliked. In conclusion, Gauch and Smith argue that their ES prototype can provide on-line search assistance to improve the efficiency of novice searches, but that much more research is needed to develop a workable, commercial system that employs this technique.

A more user-centred evaluation of an IR system is Tague and Schulz's (1988) examination of users of an OPAC interface. They develop evaluation procedures and measures based upon notions of user friendliness, informativeness and contact time. Interestingly, it is noted that friendliness is usually determined by means of a post-search interview or questionnaire, in which users are asked to judge their satisfaction with aspects of the system, such as ease of use (Tague & Schulz, 1988). Although Stevens (1983) has suggested it is possible to develop more objective criteria for this aspect, such as 'number of people who want to use the system', Tague and Schulz's approach and the approach adopted to evaluation in the user study of the FMQA (see section 7.5) cover user friendliness by means of feedback via a post-interaction questionnaire.

By informativeness, Tague and Schulz (1988) mean the relevance of the retrieved documents according to the subjective judgement of the user. Contact time is the amount of time spent by the user interacting with the system. The authors go on to attempt a formal model of informativeness based on assuming certain relationships between aspects of a user's information need and their informativeness. The model is then employed in such a way that each document in a retrieved set can be assigned an informativeness value by the user. These values are then combined to produce an overall value for the whole set. The authors applied their criteria to two different OPAC user interfaces which were linked to a small database of 487 items. They successfully managed to isolate differences in performance between the two interfaces.

A more recent user-centred approach to IR was evaluated by Morita and Shonoda (1996). They propose a method of information filtering which employs a user-profile acquisition technique based on monitoring users' behaviour and asking for feedback. The method is dependent upon the assumption that a user rejects immediately information they regard uninteresting, as opposed to spending a considerable amount of time digesting information of interest. They tested the assumption with Internet newsgroup readers by monitoring and recording such information as time spent reading an article and whether it was saved or a reply posted. They argue that the dominant factor on the time spent aspect is user interest.

The method of filtering employed is the n-gram or substring method (as described in Chapter 3). Morita and Shonoda argue that this is valid as users of newsgroups are scanning articles and judging them from the impressions of patterns of characters or keywords. The filtering works via each article being scored according to how many substrings it contains which match those in a previously constructed substring database. This database has been constructed by splitting those articles considered interesting by the user in the past into the component substrings. The evaluation of the method consisted of a simulation based on data collected during the experiment to validate the assumption regarding user rejection of information. The criteria used in the evaluation were precision and recall measures. They argue that the results indicate that the filtering method improves upon chance performance, but this claim must be considered in light of the fact there was no real user study. Their method also suffers from the problem of substring saturation in the database - too many highly rated articles by the user implies too many substrings in the database which in turn means any new article will probably attain a high score.

To a certain degree, the FMQA can be considered to be an intelligent user interface within the terms described by Brajnik *et al.* (1996). They describe this as:

> "...a front-end program which interacts with the user and controls an underlying information retrieval system accessing information resources." (p.189)

Brajnik *et al.* report the evaluation of an intelligent user interface, with the emphasis on measuring user satisfaction. The research is of interest since evaluations of user interfaces to IR systems share many of the difficulties of IR system evaluations, such as dealing with relevance (Borgman, 1986; Ennis & Sutcliffe, 1997; Zhang, 1998). Also, users and their information seeking problems are the focus of the evaluation, as is the case with the FMQA. As Brajnik *et al.* point out, one of the consequences of this is that qualitative aspects must be taken into account - aspects which are difficult to define, acquire and quantify, such as users' perception of the retrieved results. Evaluation of user opinion and satisfaction with regard to IR systems has been performed (Bailey & Pearson, 1983; Dalrymple, 1990: Dalrymple & Zweizig, 1992) but often the data acquisition techniques do not seem the most appropriate. Brajnik *et al.* advocate the use of semantic differentials, Likert scales and line-length methods which are long-established methods for acquiring user opinion (Arnold *et al.*, 1997; McCroskey *et al.*, 1997; Mehling, 1959; Osgood *et al.*, 1957). The evaluation techniques used for the FMQA are examined in section 7.4.

The prototype system built by Brajnik *et al.* to test their ideas is called FIRE and has been discussed in chapter 3. To recap, it allows a user to query a database (currently up to 20000 documents), retrieve documents, read and classify them. If not satisfied with the results, a user can begin a semi-automatic reformulation process. FIRE performs this process by selecting, from an appropriate knowledge base, a set of alternative or additional terms with which to modify the query. A main goal of the evaluation was to determine the added value of this reformulation process.

In reporting the evaluation results, the authors note that the reformulation process leads to a slight, but not statistically significant, improvement in performance, in terms of precision and recall. In this case, these measures are computed on the basis of two sets of documents that are already considered relevant. The user behaviour measures are gathered via automatic logging and questionnaires. The results seem to indicate a strong correlation between user satisfaction and the use of the FIRE automatic reformulation aspect. Another conclusion, deduced from the questionnaire answers, is that users prefer a user-controlled interaction, rather than one in which searching is completely controlled by the system. However, perhaps if the user could have confidence in the intelligent interface or system then this preference might change.

Wyle (1996) describes a prototype system called PASADENA developed to apply an information filtering model to the problem of IR from Wide Area Networks (WAN). The filtering model employs user profiles which contain complex user queries that act as filters in the retrieval of information. It is interesting to note that Wyle considers it impossible to measure the effectiveness of his prototype using the classical measures of precision and recall. He applies an effectiveness measure based on usefulness, which uses preference relations contained in each user's ranking of a retrieved set of documents. The notion of relevance in this measure is not an absolute binary judgement. Instead, users' ranking of information relevance relates to other items. Interesting as this notion is, it is difficult to judge as Wyle does not describe any experimental studies in detail.

The traditional approach has little currency in regard to the present research because the document corpus, the information residing on the WWW, is constantly changing. Also, it can be argued that the important aspect here is improving IR results *in the opinion of those most likely to benefit*, i.e. novice and intermediate users of the WWW and subject domain.

Early studies of hypertext systems and users concentrated upon comparing their performance with traditional media. Leventhal *et al.* (1993) report the results of an experiment which examined the performance and navigation strategies of users undertaking a question-answering task. The task is performed using either a hypertext encyclopaedia of facts regarding the fictional character Sherlock Holmes or the traditional paper version. Results showed that, overall, the hypertext users performed marginally better.

Studies of users with regard to the WWW have, until recently, concentrated upon surveys of their requirements, use and preferences. Prime examples are the ongoing Georgia Tech.

Surveys (Pitkow & Kehoe, 1995) and one undertaken by researchers at DMU (Davis & Houghton, 1996). Often, these surveys have highlighted the problems users encounter with the WWW, such as poor response times and 'feeling lost'. Surveys and studies are being performed in order to understand users needs and search methods in specific fields, so that the results can inform the development of a new IR tool. Ferry's (1997) large survey of people involved in the broad areas of art, design, architecture and media is one example. These surveys, and the general WWW surveys, are useful in terms of supplying a snapshot of the current situation with regard to the most important issues concerning WWW users. Their ultimate usefulness in aiding the development of effective IR systems can only be decided by studies of the developed systems themselves.

Recently, with the explosion of WWW resources and the continuing development of tools designed to aid IR from the WWW, user-centred evaluations of IR tools are beginning to emerge. *Anatagonomy* (Kamba *et al.*, 1997) is described by its developers as a personalised newspaper on the WWW. The system acts by monitoring user operations on news articles and then reflecting this behaviour in user profiles. It uses a Java applet to monitor user behaviour and to create HTML pages of information which form the newspaper, according to knowledge about the user's preferences stored in the personal profile. Hence, the term *personalised*. As part of its functionality, the system has to create 'document vectors' of each article it stores. In this manner *Anatagonomy* is sitting between the user and the WWW, downloading and storing articles before they are presented to the user. The 'vector' is a set of words which are included in the article - a parsed representation of the article.

A user profile is represented as a set of pairs of articles and their scores. This is built either by asking the user to read articles and explicitly score them, or by the system assigning a score based on the user's behaviour towards the article. That is, different scores are assigned depending upon whether the user performs such operations as scrolling down, magnifying, displaying in a separate window.

The authors have evaluated the system using 15 users divided into two groups - one providing explicit feedback, the other being monitored to provide implicit feedback. All users were also asked to rank articles via an email message. The authors claim that the results show that the implicit scoring technique and the use of the profile can provide personalised information in a successful manner. However, they concede that there are cases in which explicit scoring by the users is needed, such as if the user reads the whole article and finds it uninteresting. Anatagonomy would give this a high score in that user's profile which would be erroneous and cause problems in future IR sessions. It also seems clear that this approach would have difficulties in dealing with large numbers of articles of different topics, without the user profile becoming large and contradictory.

Recently, Gauch *et al.* (1998) developed a *meta search engine*, ProFusion, for retrieving documents from the WWW. This is basically a tool which submits user queries to a number of existing search engines in parallel and then attempts to intelligently filter and merge the results into one list which is then presented to the user. It is worth mentioning as it has been subject to a user-centered evaluation in which each user was asked to submit the same query to eight different WWW search tools and ProFusion. Each user was asked to provide a relevance judgement for the top 20 retrieved items from each search engine, noting the broken links and the duplicates. The authors noted that ProFusion outperformed the other search engines based on the performance criteria developed.

By its very nature, the present research is user-oriented. Ingwersen (1993) argues that user-oriented IR research aims at improvement of IR effectiveness within the framework of the user and their 'desire for information'(p.83). The user-oriented IR research field has concentrated on considering user's mental models and information seeking behaviour, on an individual level as well as in a wider social context (Borgman, 1986; Cockburn & Jones, 1996; Ennis & Sutcliffe, 1997; Zhang, 1998). There has been substantial research on human-intermediary information interaction, e.g. user and librarian, during IR (van Rijsbergen & Draper, 1997; Ingwersen, 1993) so that knowledge about users' behaviour during IR could inform IR system design, but little on actual evaluation of the performance of IR tools by means of studies of actual users based on criteria such as user satisfaction. As discussed above, within the realm of the WWW, user studies have been performed in the form of surveys of users behaviour, needs and preferences in relation to browsing and searching activities (Catledge & Pitkow, 1995; Cockburn & Jones, 1996; Ferry, 1997) and only recently have tools designed to improve IR from the WWW been subject to evaluation (Gauch *et al.*, 1998; Kamba, 1997).

The next section examines the aims of the FMQA user study, in the light of the above discussion.

7.3 Aims of the User Study

The primary aim of the user study is to provide data about the performance of the FMQA so that the research issue may be effectively addressed. It is also hoped that the data collected may also be used to discover any differences that might exist between distinct categories of user, with regard to their interaction with the FMQA.

The research issue has been formed as a specific question, expressed at the end of chapter 6. This can be slightly reworded and expressed instead as a hypothesis:

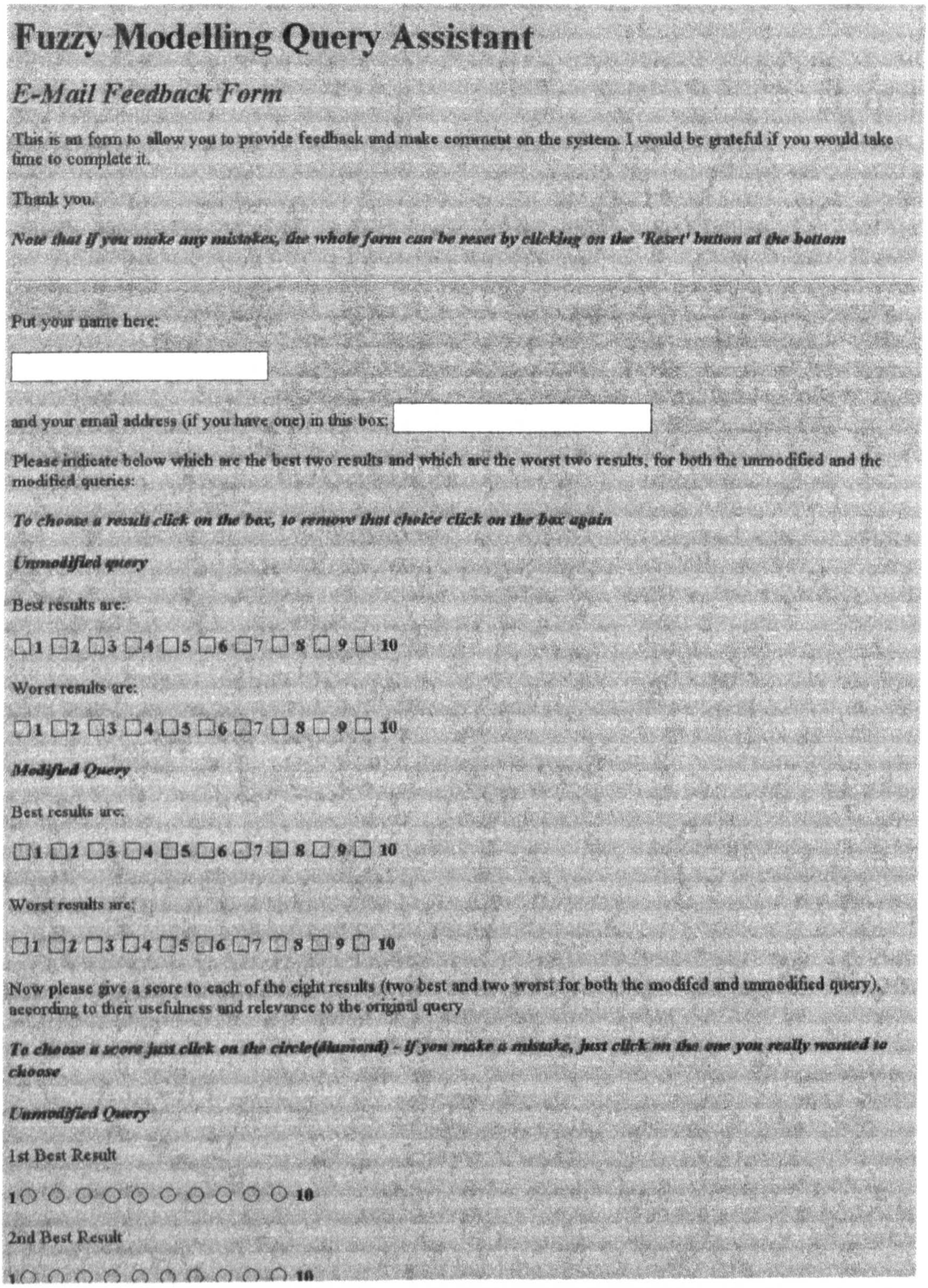
> The application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

The principal aim of the user study is to test this hypothesis. If it is supported then, in a specific case, the main thrust of the research will be upheld. Chapter 9 explores the prospect of generalising any conclusions and evidence provided by the results of the study.

As discussed in chapter 6, the FMQA, when interacting with the user, develops individual user models which represent the users subject domain and WWW experience and knowledge. These models, when defuzzified, produce two numbers, and within the query refinement process these numbers are used to modify a query. In a sense, this process models the users and categorises them as novice, intermediate or expert to a certain degree. The strength of this degree influences the extent of the query modification. Therefore, in terms of the hypothesis, novice and intermediate users within the study group can be identified. The user groups are discussed in more detail in section 7.4.

The other background information collected, such as gender and age range, not only forms an integral part of the user model, but may also allow an analysis to investigate whether any differences in the user groups' reaction to the system is attributable to their AI and WWW experience or to other factors. The information may also allow the analysis to examine if there are any differences for different types of user.

All users were asked to provide explicit feedback on the IR results of both the modified and unmodified queries, and on the design and use of the prototype tool. This step is performed by the user completing an online WWW form as the last act in their interaction with the FMQA. Figure 7.1 displays the form.



Fuzzy Modelling Query Assistant

E-Mail Feedback Form

This is an form to allow you to provide feedback and make comment on the system. I would be grateful if you would take time to complete it.

Thank you.

Note that if you make any mistakes, the whole form can be reset by clicking on the 'Reset' button at the bottom

---

Put your name here:

[ ]

and your email address (if you have one) in this box: [ ]

Please indicate below which are the best two results and which are the worst two results, for both the unmodified and the modified queries:

To choose a result click on the box, to remove that choice click on the box again

Unmodified query

Best results are:

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐ 8 ☐ 9 ☐ 10

Worst results are:

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐ 8 ☐ 9 ☐ 10

Modified Query

Best results are:

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐ 8 ☐ 9 ☐ 10

Worst results are:

☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐ 8 ☐ 9 ☐ 10

Now please give a score to each of the eight results (two best and two worst for both the modifed and unmodified query), according to their usefulness and relevance to the original query

To choose a score just click on the circle(diamond) - If you make a mistake, just click on the one you really wanted to choose

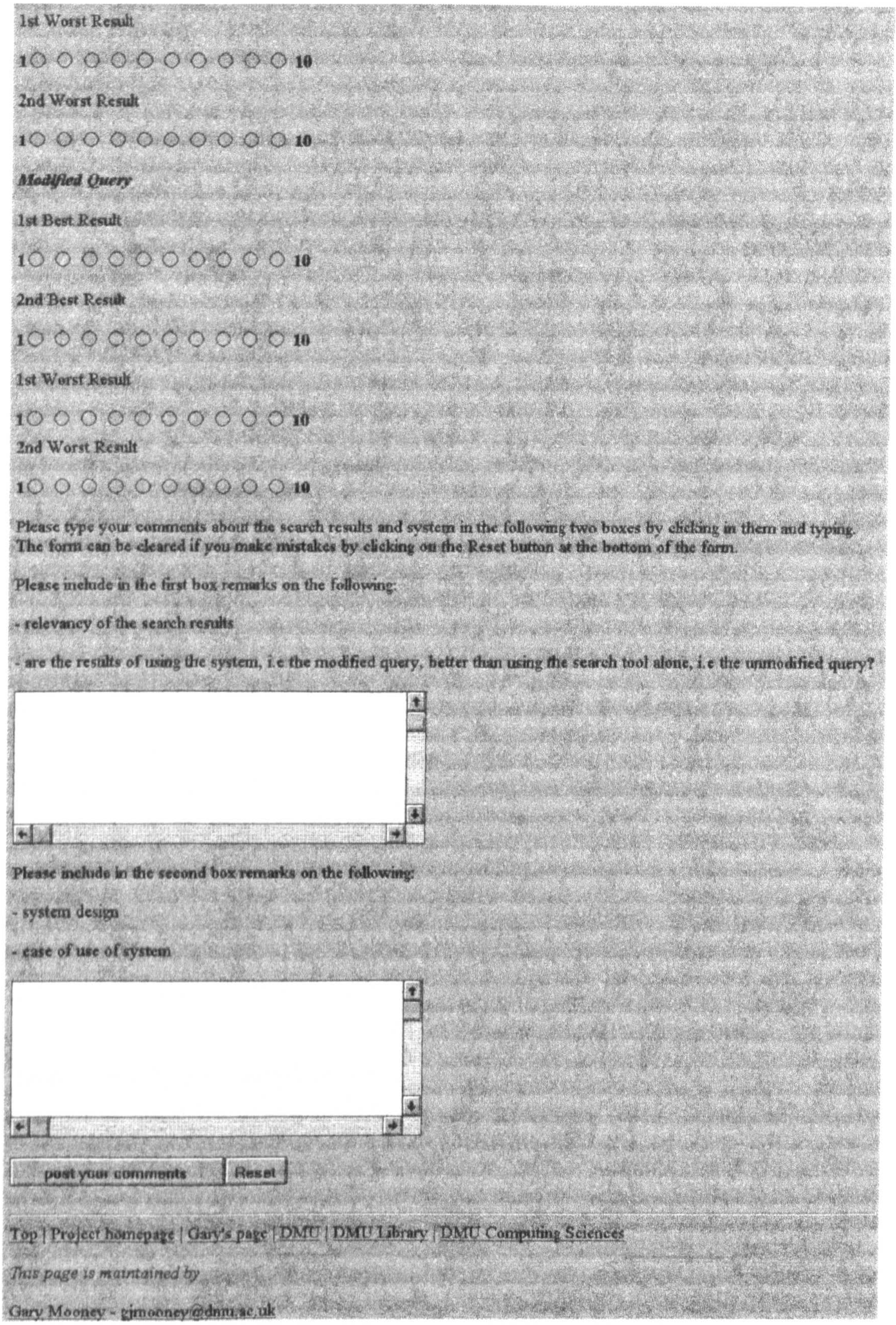
Unmodified Query

1st Best Result

1○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

2nd Best Result

1○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10



1st Worst Result

1 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

2nd Worst Result

1 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

Modified Query

1st Best Result

1 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

2nd Best Result

1 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

1st Worst Result

1 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

2nd Worst Result

1 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10

Please type your comments about the search results and system in the following two boxes by clicking in them and typing. The form can be cleared if you make mistakes by clicking on the Reset button at the bottom of the form.

Please include in the first box remarks on the following:

- relevancy of the search results

- are the results of using the system, i.e the modified query, better than using the search tool alone, i.e the unmodified query?

Please include in the second box remarks on the following:

- system design

- ease of use of system

post your comments | Reset

Top | Project homepage | Gary's page | DMU | DMU Library | DMU Computing Sciences

This page is maintained by

Gary Mooney - gjmooney@dmu.ac.uk

Figure 7.1 Email Feedback WWW Form.

In the light of the previous section's review of IR and user studies, this step is crucial in providing data which will allow the hypothesis to be tested. In focusing the study of the

FMQA on the users, performance measures which can be used to represent and compare user satisfaction are essential. Section 7.5, which deals with the study evaluation, defines and examines the performance measures. How the data related to the measures were collected is discussed in section 7.6 which deals with the conduct of the user study.

In order to perform the study, suitable groups of users were required. The identification and constituency of these is discussed next.

7.4 User Groups

The FMQA is constructed with a WWW user interface and access is not password protected. Thus it is open to use by anyone with access to a WWW browser. In fact, a valid part of the study is the data provided by users who have accessed the system in an ad hoc manner. The monitored results and comment provided by this 'group' are genuine evidence of the performance of the system. Although it is difficult to control factors that may bias the results gathered in this way, they are still worth analysing, if just to make comparison with those collected from the more controlled part of the study (see chapter 8 for discussion of results).

However, for the purpose of collecting data about the system's performance in a more controlled manner, the user study employed a cohort of undergraduate De Montfort University (DMU) students. These user groups were formed from two courses currently running in the School of Computing Sciences at DMU. These were Software Engineering and Computer Science and the two groups chosen were the first year intakes for the academic session 1996/97. First-year undergraduates were targeted as common sense dictated they were more likely to have less experience of AI and the WWW, and thus would probably provide suitable novice and intermediate level users within the definition of these terms given in chapter 5.

Previous research, undertaken through an e-mail questionnaire (Mooney, 1996) – see Appendix J – showed that these groups would provide a suitable cohort for the testing of the system. The questions were designed to gather information about the students' experience of the WWW and of the subject domain of AI. The answers indicated that the students were aware of the subject of AI and of the WWW, but have not been exposed to either to such a great extent. Therefore, they would provide a suitable number of subjects which fit the categories of 'novice' and 'intermediate' users according to the workings of the FMQA.

Additionally, as they will, during their course, acquire more experience and knowledge of AI and of the WWW, they were suitable subjects for any follow-up studies that may take place with a later version of FMQA.

There were some subjects, belonging to both the controlled user groups and the ad hoc group, who were placed in the 'expert' category by the FMQA. The results provided by these subjects were not discarded but inform the analysis as well.

In a sense, then, there are two separate user groups. The 'controlled' group of undergraduates, and the 'ad hoc' group of anyone who accessed the FMQA casually.

The conduct of the user study was influenced by the performance measures designed to evaluate the study. The next section, section 7.5, discusses these performance measures.

7.5 Evaluation

The research reported in section 7.2 shows that traditionally, in examining the performance of IR tools, the measures of precision and recall have been used. It is also noted that these measures depend upon a notion of relevancy and that, often, there is no firm acceptance of an objective definition of relevancy. Instead, there are definitions based on expert's subjective opinion of documents in relatively small, fixed databases (Bonham & Nelson, 1988; Callan & Croft, 1993; Gauch, 1990). User-centred evaluations of IR tools have attempted to find criteria to measure such quantities as 'user satisfaction' and others based on the user's behaviour with the tool (Bonham & Nelson, 1988; Brajnik *et al.*, 1996; Efthimiadis, 1993; Hull, 1993).

In the present research, because the WWW is in an ever-changing flux, it is doubly difficult to define a notion of relevancy and thus use precision and recall as performance measures. Of course, the FMQA could have been tested against a fixed subset of WWW documents which had already been judged for relevancy to particular IR queries by experts in the subject domain. Some WWW IR tools have been evaluated in this way (Cockburn & Jones, 1996; Gauch *et al.*, 1998). However, this would mean that the prototype had been evaluated against a fixed, limited size corpus of WWW documents, and not the WWW itself. Thus, any results from such a narrow evaluation would be limited in their extent and difficult to analyse and interpret in the light of the research hypothesis.

Additionally, as noted at the end of section 7.2, this research is user-oriented. The crux of the research is the notion of improving IR from the WWW for certain types of user. Thus, in any evaluation, the users' opinions and rating of the results and the prototype are paramount, so that the research issue can be effectively addressed.

These two factors led to the rejection of using measures of precision and recall as performance indicators and to the adoption of the following, user-centred, performance criteria:

- user group member noting of the best two and worst two results (out of the first 10) of the FMQA set and of the Lycos™ set (equivalent to the modified and unmodified query sets)

- user group member scored rating of these eight results (scoring each out of 10).

The user was asked to make their decisions on which are the best and worst two results in terms of relevancy, usefulness and precision with regard to the original unmodified query, and then give each a score between 1 and 10.

Other more subjective performance measures also informed the evaluation:

- user group member opinion of the relevancy and usefulness of the results, modified against unmodified query (text-based response)
- user group member opinion of the overall design, ease of use and functionality of the system (text-based response).

Data on the first and last measures were collected to provide objective measures on the performance of the prototype in general terms, and to allow the users responses to be analysed in the light of these factors.

The results were collected automatically and the ratings and opinions solicited from the users as part of a session using the FMQA, when they are asked for feedback (see figure 7.1).

The performance measures were chosen to allow the research issue to be addressed effectively. Data were collected during the conduct of the study in order to satisfy the performance measures. The conduct of the study is discussed next in section 7.6.

7.6 Conduct of the User Study

The conduct of the user study began in October 1996. The first phase, the controlled part, involving the cohort of DMU undergraduates, ran from late October 1996 until early December 1996. The second phase, the ad hoc part, also began at the same time because that is when the FMQA went 'live' on the WWW and continued until July 1997. However, the presence of the FMQA was not advertised, via postings to both internal and external email lists the members of which it was thought might be interested, until after the first phase was completed.

The controlled part took longer than anticipated or than is ideal, owing to a poor response by the user groups involved. Even after numerous attempts to generate interest, the final response in terms of numbers was less than was hoped. The poor response also prevented the controlled part of the study from being completed quickly. Ideally, to prevent biased and incompatible results occurring because of changes to the WWW content and structure (which are ongoing), all subjects should have accessed the system at the same time on the same day in the same environment. This was not practically feasible, so other steps were taken to limit any bias.

Random checks on the IR results returned by the FMQA and Lycos™ suggested that the changes in the WWW content overall did not significantly occur over the period of the first phase. During this phase Lycos™ returned not only the number of hits matching the query but also an estimate of the total number of WWW documents. This number did not substantially change over the period. Therefore, although the goal of implementing the controlled phase quickly was not achieved, it can be argued that the results are not significantly biased by any changes in the WWW.

During the study, a crucial aspect was the time of day it took place. The FMQA is modifying a query before it is submitted to a search engine which resides on a computer in the USA. Thus, speed of access and retrieval of results is subject to the transatlantic link and the amount of network traffic. As it was not feasible for all the members of the controlled user group to access the FMQA at the same time, it was important that the time access did take place for each subject was controlled to some extent, so that their responses to the performance of the system were not influenced unduly by this factor.

Thus, it was decided that it was important that the study took place before 12 p.m. when the traffic is lighter ***and*** that each user access the system at a similar time during the day.

Accordingly, the period 9.30 a.m. to 12.00 p.m. every weekday for the period of the first phase was designated for members of the controlled group to take part.

Where the study was performed was less important and, as already noted, the WWW interface of the FMQA meant that access was possible from any WWW-enabled machine. However, the environment in which a user accesses the FMQA could have some effect on their response and, therefore, to limit any bias that could be introduced in this way, the controlled group's access was restricted. Three machines running the FMQA within the same office were set aside for a subject to access the system at the correct times. However, if a member of the controlled user group accessed the system from a different location, then the results of such an access were still deemed to be suitable for inclusion in the study.

The second phase of the study, in which results of ad hoc access to the FMQA were collected, is, in comparison with the first phase, much less controlled. Access was monitored via the data gathered from the questionnaires and the responses provided via the WWW form. Again, the response, in terms of numbers, was not as great as expected. Often users would interact with the system but choose not to complete the response form. A follow-up email request did produce a response in a few cases. Some ad hoc users chose to respond with comments directly via email. The results of this second phase, both complete and partial, also inform the analysis reported in chapter 8.

From the above discussion, and from the description of the user groups given in the previous section, it can be noted that the first phase of the user study can be considered to be controlled by the following factors:

- the user group constituency (the controlled group members have similar backgrounds, similar ages, and similar experience of the WWW and of AI)
- the timescale (much shorter in the controlled part)
- the time of access to the FMQA
- the environment.

It would have been interesting to compare the results of each phase in the light of these differences, but, as reported in chapter 8, the amount of data collected was insufficient to allow any significant comparison.

After the completion of the study, the results were subject to an analysis, based upon the series of performance measures identified in section 7.5, in order to meet the aims of the study. The results of the study and the subsequent analysis of the data are reported and examined in detail in chapter 8. First, this chapter concludes with a discussion of the advantages and disadvantages of the user study.

7.7 Conclusions

This chapter has examined the design and conduct of a user study of the performance of the prototype tool, the FMQA. This study was undertaken so that the research issue could be effectively addressed.

Originally, there were meant to be two separate studies (see Appendix E), with results from the first informing any changes to both prototype and further studies before the second took place. This would have been ideal, but problems in generating enough user response at the outset, coupled with time restraints on the research, meant that the study had to be restructured. Instead of two separate studies, it became one of two phases, as previously described. However, enough users have taken part in both phases to allow interesting results to emerge and comparisons to be made, as will become evident in chapter 8.

Perhaps more incentive should have been offered to users to take part, but this idea was tempered by recognition that too much inducement might have biased any objective user reaction to the FMQA. The study has been designed so that follow-up studies, with the same or different cohorts of users and with the same or different versions of the FMQA, could be conducted. This might prove interesting, in terms of how users' responses to the system change as they increase their knowledge and experience of the WWW and a subject domain.

A possible weakness of the study lies in the lack of a truly objective performance criterion, such as measurement of precision or recall, with which to focus the results analysis. However, as already noted, these 'objective' measures rely on a notion of relevance and there is great difficulty in defining this notion with regard to the 350 million plus documents residing on the WWW (Altavista, 1999a). Also, the decision was made not to test the FMQA against a small, fixed set of documents - about which relevancy could have been defined - as this, whilst no doubt interesting, would have not provided any evidence to support the research hypothesis.

The user-centred nature of the research is addressed by the exploitation of user-focused performance criteria. Not only has written opinion been sought but also rating of IR results.

This means that there is a numerical measure of user satisfaction in the results and opinion of the system, with which the overall FMQA performance can be judged, and which can be used to quantitatively address the research issue.

Ideally, all subjects in the study would have taken part at the same time and within the same environment. However, this ideal was impossible and a disadvantage of the study is that it is difficult to assess the impact of the environment on the results. The controlled part of the study was conducted to minimise the effects of the environment, and, as stated in section 7.5, that part of the study took place in a short enough period to mean that there were minimal effects caused by changes to the WWW. Also, a strength of the study is the fact it was open to all users of the WWW. This has meant reaction and opinion from different types of user with widely different backgrounds from all over the World.

The insight gained from the experience of conducting the user study would be very useful in any further studies. For example, in hindsight any controlled part of a study should be designed so that subjects took part simultaneously. However, such a condition may have its own effects, for instance an increase in network traffic, which would also be difficult to quantify.

The designed and performed study did produce sufficient and useful results, within the parameters of the chosen performance measures. These results do allow the research issue to be tackled, as will be seen in the next chapter, chapter 8, which examines the results and the subsequent analysis in detail.

CHAPTER 8 - RESULTS

8.1 Overview

The previous chapter examined the preparation and conduct of the user study employed to test the performance of the FMQA. The design and analysis of the user study was focused on providing enough evidence to allow the research issue to be addressed. Although it had some limitations, the user study did provide sufficient results to allow an analysis. These results and the subsequent analysis are reported in detail in this chapter.

The research issue, as expressed during chapter 7, can be formulated as the following question:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics?

In this chapter, in section 8.2, this question is reformulated as a null hypothesis, in such a way that the analysis of the results may provide an answer to the hypothesis and so to the original question. A null hypothesis is one in which an assumption of null effect is stated. This is then statistically tested and either accepted as true or rejected as false (Reid & Boore, 1992). To achieve this goal, an *Analysis Plan* (see Appendix F) was drawn up in which it was ascertained that a number of sub-questions needed to be addressed by the analysis in order for the hypothesis to be answered. These questions are first expressed in this chapter in section 8.2.

It is answering these questions which is the focus of the analysis and of the statistical tests applied to the results data. The choice of tests to apply was also constrained by the limited amount and nature of the data collected. Aspects of the data are discussed in section 8.3 in which the results are presented in summary (see Appendix G for the full raw data). The analysis itself follows in section 8.4. An explanation of each test, the data to which it applies and the sub-question that it addresses, is provided. This is followed in section 8.5 by a discussion of the analysis outcome and the resulting test values, focusing on the implications for the sub-questions and the null hypothesis. The limitations of the analysis are also considered. Finally, in section 8.6, conclusions are drawn from the analysis, with regard to the research issue itself.

8.2 Background

The question stated in the previous section, which is the crux of the whole research, can be restated as a research hypothesis. This could be:

> The results obtained by the application of the FMQA to IR from the WWW improve upon those obtained by using Lycos™ alone, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

In order to investigate this hypothesis using data obtained from the user study, a statistical analysis of the data is required. However, to carry out any analysis of statistical weight, a null hypothesis should be explicitly stated (Reid & Boore, 1992). The research hypothesis given above can be restated as a null hypothesis, as follows:

> There is no significant improvement in the results obtained by the application of the FMQA to IR from the WWW over those obtained by using Lycos™ alone, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

The focus of the analysis, thus, then becomes providing evidence to either reject or accept the null hypothesis. It will become apparent later in this chapter that there is sufficient evidence to reject the null hypothesis. It follows from this that the results of the user study can be shown to support the research hypothesis.

It was ascertained during the development of the plan to analyse the results (see Appendix F for *Analysis Plan* in full) that to address the null hypothesis a number of sub-questions needed to be answered. These are:

(1) Is the FMQA working, is it making a difference to the IR process?
(2) Is there an overall improvement in the results through the use of the FMQA?
(3) Does the FMQA improve results for different categories of user, specifically for those users deemed novice and intermediate?
(4) Does the FMQA improve results over the range of the fixed different topics within the one information domain?

Thus, the focus of the analysis became the answering of these questions by means of appropriate statistical tests. The nature of the questions influenced the portions of the results data employed in the tests. The results are presented in summary next, in section 8.3, before the report of the analysis, which follows in section 8.4.

8.3 Data Summary

This section presents a summary of the data collected during the user study. The information available for analysis consists of the following:

- the data provided via answers to the two questionnaires (see chapter 6, pages 94 and 96, and Appendix G)
- the responses submitted by those users who completed the email feedback form
- the values produced by the fuzzy user modelling process and their interpretation.

From chapter 6 it is apparent that the nature of the study means that there are two distinct user groups - a controlled and an ad hoc group. For the purposes of the analysis, the data collected on the two groups was initially dealt with separately and the summary results are presented independently. As discussed in chapter 6, it is the type of user represented in the control group that the FMQA can assist most. Therefore, in answering the research issue, it is data collected from this group of users which is of paramount importance.

An important point needs to be made here regarding the categorisation of the results for statistical purposes. It will be noted that they have been categorised in a number of ways. Most of these and the values they can have pertain to the questions answered in the questionnaires. For example, the category *Background* can have the values *UK, EU, Overseas* or *Other*. However, some categories and the values are derived during the process of fuzzification and defuzzification which the FMQA performs to modify the original query. Thus, the categories *Internet Model* and *Subject Model* can have the values *Low, Medium* and *High*. These values, as will become evident, represent different ranges into which the defuzzified numbers produced by the final fuzzy user models are placed.

Recall, as discussed in chapter 5, that the action of the FMQA in producing fuzzy user models for two characteristics of the users, and then in defuzzifying these models to produce two numerical values, is not one which allows the comment that a user is 100% in the low, medium or high ranges for either model. Rather, the whole principle of fuzzy logic and modelling (Zadeh, 1994) means that the user can only be considered to be within a region to a certain degree, i.e. they can be considered to be in the low range to a certain degree and *simultaneously* in the medium region to a different degree.

However, in the prototype the defuzzified values are used to modify the original query by first considering which range the value lies in, low, medium or high. In doing so, the user can be considered to be *novice, intermediate* or *expert*, respectively, and thus the research hypothesis can be addressed as the FMQA, through this process, allows the users to be classified in one of the three categories. The range boundaries are such that a defuzzified FUM value is placed in the range to which it belongs to the greatest degree.

The other categories that are derived are the 'By Model' and the 'Query Modification'. The first of these is simply a count of the number of users falling into each possible combination of the 'Internet Model' and 'Subject Model' categories. The second provides a count of the number of users whose query was modified in each possible way. Recall from chapter 6 that in the prototype the FMQA can modify an original query in four different ways - *C(Contracted), S(Slight), E(Extra), G(Great)* - and these are the values allowed in his category. Contracted (C) means the original query is replaced with a more general term. In the other cases the query is expanded by the addition of terms, and the greater the expansion, the more terms are added.

The feedback category is divided into those who provided *Full* feedback - i.e. fully completed the email feedback form - those who provided *Partial* feedback - incomplete forms - and those who provided *None*.

The final category, the *Results*, refers to the list of top ten results returned by Lycos™. It is a measure of the number of distinct sets of results produced by the user group. The measure was obtained by combining the number of distinct original queries with the ways in which the FMQA had modified each query.

The next two subsections present the summary statistics for both the controlled and ad hoc group. Where appropriate, there is further explanation of the data.

8.3.1 User study (Controlled Phase) - Summary Statistics

DATA CATEGORY		VALUE
Number of participants		**39**
Course:	Computer Science	**10**
	Software Engineering	**28**
	Management Science	**01**
Age Range	18-21	**35**
	22-25	**01**
	No response	**03**
Gender	Male	**37**
	Female	**02**
Background	UK	**38**
	No Response	**01**
Position	Undergraduate	**39**

Table 8.1 Background Information by Category - Controlled Phase

The user group is dominated by computing students (Computer Studies/Software Engineering), reflecting the fact these students were targeted by the initial email questionnaire (see chapter 5) as likely to prove to be suitable subjects. The user group members are predominantly in the age range *18-21* and are all UK nationals. They are all undergraduates as expected.

Unfortunately, the user group members are overwhelmingly male, thereby preventing any possible analysis of gender differences because of insufficient sample sizes of each gender. Perhaps this is to be expected, given the targeting of DMU computing undergraduates as a suitable group for the prototype study.

Original Query	Machine Learning	**11**
	Neural Networks	**08**
	Fuzzy Logic	**07**
	KBS	**04**
	AI Programming Languages	**03**
	NLP	**02**
	Genetic Algorithms	**01**
	No Response	**03**

Table 8.2 Number of Users per Original Query Topic - Controlled Phase

Table 8.2 displays the number of users per each possible choice of original query. There is a spread amongst the possible queries which means that the analysis is not biased or disadvantaged by a concentration of users choosing one particular original query.

WWW model (defuzzified value)	Low (= < 6)	**00**
	Medium (> 6 = < 14)	**10**
	High (> 14)	**28**
	No response	**01**
Subject model (defuzzified value)	Low (= < 6)	**25**
	Medium (> 6 = < 14)	**11**
	High (> 14)	**00**
	No Response	**03**

Table 8.3 Number of Users in Each Model Range (WWW and Subject) - Controlled Phase

For purpose of query refinement, the defuzzified value for each fuzzy representation of the user's WWW and Subject experience is classified as Low, Medium or High, according to the value in the ranges outlined. This is a breakdown of the number of users in each range for both WWW and subject experience as defined by the fuzzy user models.

By model category		**Subject model**		
		Low	Medium	High
WWW model	Low	**00**	**00**	**00**
	Medium	**05**	**05**	**00**
	High	**20**	**06**	**00**

No category possible **03** (users did not respond properly in questionnaires)

Table 8.4 Number of Users by Model Category (WWW-Subject) - Controlled Phase

The query refinement process is dependent on the classification for both WWW and Subject. Thus, there are nine separate categories possible in combining the classifications. This is a breakdown of the number of users in each possible combination. However, the fact that there are no users in five of the classifications means that effectively there are only four for the purposes of the analysis and these can be thought as forming a two-way table. The majority of users fall into the *high-low* WWW-Subject combination indicating that they see themselves as having a considerable amount of experience and knowledge of the WWW and little of the subject. This was slightly unexpected but perhaps indicates that students are being exposed to the WWW at an earlier stage in their lives than previously thought.

Query Modification	C (Contract)	**05**
	S (Slight)	**20**
	E (Extra)	**11**
	G (Great)	**00**
	None	**03**

Table 8.5 Frequency of Each Type of Query Modification - Controlled Phase

In refinement the Query is either Contracted (C), i.e. original replaced with a narrower term, or expanded by the addition of terms. This expansion can be Slight (S), Extra (E) or Great (G). The type of refinement is dependent on which category the user placed. This is a simple breakdown of how many users in each type of refinement.

Feedback	Fully	**17**
	Partially	**05**
	None	**17**

Table 8.6 Number of Users Providing Feedback - Controlled Phase

A total of 17 users fully completed the email response form, 5 partially and 17 provided no response. Partial response means that only part of the email form was completed – usually, the user provided textual feedback but did not rate or rank their results as requested (see example of response form on pages 116-117).

Textual Response			
	Results	Modified better	**18**
		Neither better	**01**
		Unmodified better	**01**
		Response unclear	**02**
		No response	**17**
	System	Well designed, easy to use	**21**
		Room for improvement	**01**
		No response	**17**

Table 8.7 Number of Users in Each Classification of Textual Response - Controlled Phase

The users were requested to provide written comment of the FMQA's performance and design on the email response form. The table above represents an analysis of the responses. The analysis depends on judging which category the textual response belongs to by examination of the words. As can be noted, the textual responses from the controlled group were very favourable, in terms of their opinion of the results provided via the FMQA and of the prototype's design and ease of use.

Results			**Modification**			
			Contract	Slight	Extra	Great
	Original Query	Machine Learning	√	√	√	**X**
		Neural Networks	√	√	√	**X**
		Fuzzy Logic	**X**	√	√	**X**
		KBS	**X**	√	√	**X**
		AI Prog. Langs	**X**	√	**X**	**X**
		NLP	**X**	√	√	**X**
		Genetic Algorithms	**X**	√	**X**	**X**

Key: √ = Set of IR results with modification of that original query exists
X = Set of IR results with modification of that original query does not exist

Table 8.8 Types of Sets of IR results stored by Original Query and Modification - Controlled Phase

There are 21 distinct sets of results. A 'set' of results is the list of ten URLs returned by Lycos™ as being the closest match to the query. Sets of results have been stored for each original query chosen and for each type of refined query produced by the FMQA in interacting with the users. A '√' in the table indicates the set exists, a 'X' that it does not.

Results vs. User				**Modification**			
			Original	Contract	Slight	Extra	Great
	Query	Machine Learning	**11**	**04**	**06**	**01**	**00**
		Neural Networks	**08**	**01**	**03**	**04**	**00**
		Fuzzy Logic	**07**	**00**	**05**	**02**	**00**
		KBS	**04**	**00**	**01**	**03**	**00**
		AI Prog. Langs	**03**	**00**	**03**	**00**	**00**
		NLP	**02**	**00**	**01**	**01**	**00**
		Genetic Algorithms	**01**	**00**	**01**	**00**	**00**

Table 8.9 Number of Users per Original Query and per Modified Query (Classified by Original Query and Modification) - Controlled Phase

The number of users per original query and per type of modified query, for each original query chosen. Each user appears once in the 'Original' total for each type of query and once in a category of refinement. It can be noted that the majority of users in this group had their original query modified slightly by the FMQA. This reflects the fact that the FMQA considered the majority of the users to be 'novice' or 'intermediate'.

8.3.2 User study (Ad-Hoc Phase) - Summary Statistics

The group of users that provided data for this phase of the user study is rightly termed ad-hoc because there was no attempt at controlling any aspect of their constituency or the environment in which they took part. Rather, during the period that the FMQA was 'live', then as it has a WWW interface it was accessible from anywhere by anyone with a WWW browser and access to the Internet. The study of the prototype was advertised and announced on various appropriate e-mail lists but, as no data on where and when users heard about the study were collected, then it is difficult to analyse the statistics on the users' general characteristics in any meaningful and proper way.

DATA CATEGORY		VALUE
Number of participants		**75**
Course		**Not Applicable**
Age Range	18-21	**03**
	22-25	**23**
	25+	**49**
Gender	Male	**63**
	Female	**12**
Background	UK	**30**
	EU	**16**
	Overseas	**29**
Position	DMU Staff	**13**
	Postgraduate	**21**
	Undergraduate	**06**
	Other	**35**

Table 8.10 Background Information by Category - Ad-Hoc Phase

The course category is not applicable as the ad-hoc group of users are not constrained to be DMU undergraduates. Of those who responded the users are predominantly over 25, but the number of users with this category value that actually provided proper feedback was insufficient to make it possible to perform any analysis breakdown by this category.

Again, as with the control group, the user group members are overwhelmingly male, thereby again preventing any possible analysis of gender differences because of the imbalance in the sample sizes of each gender. There is a good spread between the three possible values of the category *Background.*

Postgraduates and undergraduates from outside DMU are not included in 'Other' but in the relevant category. 'Other' is the dominant category, perhaps reflecting the open nature of this phase of the user study, and its dominance means there is little to be gained from analysis of this aspect of the users. Although asked to elaborate if choosing this category, most users did not explain their position further.

Original Query	KBS	**15**
	Distributed AI	**14**
	Neural Networks	**13**
	Fuzzy Logic	**08**
	NLP	**07**
	Machine Learning	**05**
	Genetic Algorithms	**04**
	AI Programming Languages	**02**
	Semantic Networks	**02**
	Bayesian Networks	**02**
	No Response	**03**

Table 8.11 Number of Users per Original Query Topic - Ad-Hoc Phase

The number of users per possible choice of original query. There is a spread amongst the possible queries which means that any analysis is not biased or disadvantaged by a concentration of users choosing one particular original query.

WWW model (defuzzified value)	Low (= < 6)	**00**
	Medium (> 6 = < 14)	**13**
	High (> 14)	**59**
	No response	**03**
Subject model (defuzzified value)	Low (= < 6)	**24**
	Medium (> 6 = < 14)	**27**
	High (> 14)	**21**
	No Response	**03**

Table 8.12 Number of Users in Each Model Range (WWW and Subject) - Ad-Hoc Phase

The breakdown of users in the above table is arrived at in exactly the same manner as in the corresponding controlled phase statistics.

By model category		Subject model		
		Low	Medium	High
WWW model	Low	**00**	**00**	**00**
	Medium	**09**	**03**	**01**
	High	**15**	**24**	**20**
No category possible	**03** (Users did not respond properly in questionnaire)			

Table 8.13 Number of Users by Model Category (WWW-Subject) - Ad-Hoc Phase

As before there are nine separate categories possible in combining the classifications and this is a breakdown of the number of users in each possible combination. There is a greater spread of numbers in the categories but again there are three with no data. This is perhaps an indication that none of those taking part in the study consider themselves to be novices with regard to the WWW and the FMQA, in processing the users' questionnaire answers, has reflected that. There is a more reasonable spread in the other combinations but the majority of the users fall into a category containing the high classification of the WWW model value and over half of the 75 are in the medium and high classifications of the subject model value. This indicates that the majority of the users in this group have been considered by the FMQA to be in the 'expert' category or towards the high end of the 'intermediate' category.

Query Modification	C (Contract)	**09**
	S (Slight)	**14**
	E (Extra)	**28**
	G (Great)	**21**
	None	**03**

Table 8.14 Frequency of Each Type of Query Modification - Ad-Hoc Phase

In refinement the Query is either Contracted (C), i.e. original replaced with a narrower term, or expanded by the addition of terms. This expansion can be Slight (S), Extra (E) or Great (G). The type of refinement is dependent on which category the user placed. This is a simple breakdown of how many users in each type of refinement.

Feedback	Fully	**10**
	Partially	**07**
	None	**58**

Table 8.15 Number of Users Providing Feedback - Ad-Hoc Phase

A total of 17 users fully completed the email response form, 7 partially and 58 provided no response. Partial response means that only part of the email form was completed – usually, the user provided textual feedback but did not rate or rank their results as requested (see example of response form on page 116).

Textual Response			
	Results	Modified better	**09**
		Neither better	**04**
		Unmodified better	**02**
		Response unclear	**02**
		No response	**58**
	System	Well designed, easy to use	**08**
		Room for improvement	**05**
		Response unclear	**04**
		No response	**58**

Table 8.16 Number of Users in Each Classification of Textual Response - Ad-Hoc Phase

The users were requested to provide written comment of the FMQA's performance and design on the email response form. The table above represents an analysis of the responses. The analysis depends on judging which category the textual response belongs to by examination of the words. As can be noted, the textual responses from the controlled group were favourable, in terms of their opinion of the results provided via the FMQA and of the prototype's design and ease of use. (The full responses are given in Appendix G).

Results			**Modification**			
			Contract	Slight	Extra	Great
	Original Query	KBS	√	√	√	√
		Distributed AI	X	X	√	√
		Neural Networks	X	√	√	√
		Fuzzy Logic	√	√	√	√
		NLP	√	√	X	X
		Machine Learning	X	X	√	X
		Genetic Algorithms	X	√	√	√
		AI Prog. Langs	X	√	√	X
		Semantic Networks	X	√	X	√
		Bayesian Networks	X	√	X	X

Key: √ = Set of IR results with modification of that original query exists
X = Set of IR results with modification of that original query does not exist

Table 8.17 Types of Sets of IR results stored by Original Query and Modification - Ad-Hoc Phase

In this case there are 34 distinct sets of results (made up of 10 original query sets and 24 sets relating to modified queries - the '√'s in the table). As before, sets of results have been stored for each original query chosen and for each type of refined query produced by the FMQA in interacting with the users. The notation is the same as in the controlled phase.

Results vs. User				Modification			
			Original	Contract	Slight	Extra	Great
	Query	KBS	**15**	**04**	**01**	**05**	**05**
		Distributed AI	**14**	**00**	**00**	**07**	**07**
		Neural Networks	**13**	**00**	**05**	**04**	**04**
		Fuzzy Logic	**08**	**04**	**02**	**01**	**01**
		NLP	**07**	**01**	**01**	**03**	**02**
		Machine Learning	**05**	**00**	**00**	**05**	**00**
		Genetic Algorithms	**04**	**00**	**01**	**02**	**01**
		AI Prog.Langs	**02**	**00**	**01**	**01**	**00**
		Bayesian Networks	**02**	**00**	**02**	**00**	**00**
		Semantic Networks	**02**	**00**	**01**	**00**	**01**

Table 8.18 Number of Users per Original Query and per Modified Query (Classified by Original Query and Modification) - Ad-Hoc Phase

The number of users per original query and per type of refined query, for each original query chosen. As before, each user appears once in the 'Original' total for each type of query and once in a category of refinement. It can be noted that, in the ad-hoc phase, the majority of users had their original query modified 'extra' or greatly by the FMQA. This reflects the fact that the FMQA considered the majority of the users to be 'expert' or at the high end of 'intermediate'.

Before reporting the data analysis in section 8.4, an important set of data required for the analysis, not examined above, needs to be considered. These data are that provided by the users when asked to consider and rate the results returned by the search engine Lycos™. The amount of data is too voluminous to be presented in full here and is collected in Appendix G. In the following sub-section the nature of the data is discussed and example portions are given to highlight its importance.

8.3.3 User Rating of the IR Results

In this sub-section the data provided by the users when asked to consider and rate the IR results is briefly discussed. Recall from chapter 7 that the users were asked to decide which were the best two and the worst two results for each set of ten results returned, one set each representing the response of Lycos to the submission of the modified and unmodified queries. They were then asked to rate each of the eight results chosen, giving each a score out of ten. The data collected from this process are crucial to the whole research, given the user-centred focus stated earlier (see chapter 6). It is the data on which the main analysis is based and thus upon which the fate of the research hypothesis rests.

The total number of users who fully completed the results rating part of the user study was 36: 22 from the controlled phase and 14 from the ad hoc phase. Though this was less than hoped for when compared with the total numbers which took part in the study, the amount reflects the voluntary nature of the study and especially of the feedback part which was conducted via email.

As with the summary statistics given in section 8.3.1, the data collected for each phase of the study were initially dealt with separately. The data are presented in full in spreadsheet form in Appendix G. A portion of the spreadsheet for the controlled phase data is shown below in figure 8.1.

record no.	best score (unmod)	2nd best score (unmod)	best score (mod)	2nd best score (mod)	worst score (unmod)	2nd worst score (unmod)	worst score (mod)	2nd worst score (mod)
10	5	5	9	8	1	3	6	6
21	9	8	9	9	1	2	4	4

Figure 8.1. Portion of the user scoring of the IR results.

As can be seen the user ratings or scores are categorized as follows:

best score (unmod)	best unmodified result score
2nd best score (unmod)	2nd best unmodified result score
best score (mod)	best modified result score
2nd best score (mod)	2nd best modified result score
worst score (unmod)	worst unmodified result score
2nd worst score (unmod)	2nd worst unmodified result score
worst score (mod)	worst modified result score
2nd worst score (mod)	2nd worst modified result score

The scores given by each user who completed fully the scoring process comprise the data used in the majority of the statistical tests.

The total amount of data collected during the user study, reported above in the summary statistics and the user scoring of IR results, is sufficient to enable a detailed statistical analysis to be completed. There is a good spread of numbers of different queries and different degrees of modification. In total, there are 72 pieces of data referring to users' quantitative opinion of the IR results. This means that a fairly detailed analysis, one which considers different strata of the data, can be performed and thus the research hypothesis can be addressed with some confidence. The analysis is reported next, in section 8.4.

8.4 Analysis of the User Study Results

The analysis is driven by the need to address the null hypothesis. As already noted, from the Analysis Plan (Appendix F) it was ascertained that in order to satisfy this need a number of sub-questions had to be answered (see above page 126).

Thus, the statistical tests chosen to provide the basis of the analysis were selected in the knowledge that they needed to offer evidence with which to answer the sub-questions.

Before describing each test in detail, how it was applied and the consequent results, a summary of which tests have been applied to which questions is given.

(i) Is the FMQA working, is it making a difference to the IR process?
- this question is addressed by analysing the summary statistics and examining the means of the scores given by the users.

(ii) Is there an overall improvement in the results through the use of the FMQA?
- this question is also addressed by analysis of the summary statistics and examining the user responses in terms of the scores. Analysis of that data for both user groups (ad-hoc and controlled) is performed using Wilcoxon's test (Reid & Boore, 1992).

Additionally, McNemar's test was performed on the textual comments provided via the email feedback form (Reid & Boore, 1992).

(iii) Does the FMQA improve results for different categories of user, specifically for those users deemed novice and intermediate?
- using the summary statistics to identify the user scores data which pertains to the different categories of user and the different degrees of modification, the mean scores were analysed.

(iv) Does the FMQA improve results over the range of fixed different topics within one information domain?

- using the summary statistics once more to identify which user scores data is associated with different categories of query, the mean scores were analysed.

By using the tests to provide answers to these questions, the null hypothesis and thus the overall research question can be addressed. The tests and their outcomes are now reported.

8.4.1 Does the FMQA make a difference?

There is little point in performing any detailed analysis if the FMQA is making no difference to the overall IR process. Thus, the first tests were done to ascertain whether the FMQA was really making a difference to the process.

unmodified query	textual response
Genetic Algorithms	Response Unclear
Natural Language Processing	Similar Results
	Modified Results Better
AI Programming Languages	Modified Results Better
	Modified Results Better
Neural Networks	Modified Results Better
	Modified Results Better
	Unmodified Results Better
Machine Learning	Modified Results Better
	Modified Results Better
	Modified Results Better
	Response Unclear
Knowledge-based Systems	Modified Results Better
	Modified Results Better
	Modified Results Better
	Modified Results Better
Fuzzy Logic	Modified Results Better
	Modified Results Better
	Modified Results Better
	Modified Results Better
	Modified Results Better
	Modified Results Better

Figure 8.2 Portion of User Study Data Showing User Comment on Results Against Unmodified Query (Controlled Phase)

Examining the 'textual' response category of the controlled user group summary statistics provided the first evidence for a *difference*. Of the 22 users who responded, 18 indicated they thought the modified query produced better IR results and only one expressed the opinion that there was no difference between the modified and unmodified IR results. Neither is this difference confined to a minority of categories of query. A comparison of textual response by original unmodified query for the controlled group shows that, for the majority of the query categories, the users considered the FMQA to be making a difference. This becomes obviously clear by considering figure 8.2 which displays the relevant portion of the data.

(N.B. Recall that the value of the category *textual response* was decided by analysis of the textual feedback provided by the users during the study.) For the ad-hoc phase, there are less data to examine in this way but it is still easy to see that the FMQA is making a difference.

unmodified query	**textual response**
Bayesian Networks	Unmodified Results Better
Natural Language Processing	Similar Results
Genetic Algorithms	Similar Results
Knowledge-based Systems	Modified Results Better
AI Programming Languages	Modified Results Better
Fuzzy Logic	Modified Results Better
	Modified Results Better
Neural Networks	Modified Results Better
	Modified Results Better
	Modified Results Better
Distributed AI	Response Unclear
	Modified Results Better
	Similar Results

Figure 8.3 Portion of User Study Data Showing User Comment on Results Against Unmodified Query (Ad-hoc phase)

To examine the difference more quantitatively, the means of the ratings or scores given by the users to the set of IR results were taken in the following ways.

The mean of the scores given for the two best results for both the modified and unmodified queries, for each user providing feedback in the controlled phase of the user study, were taken by summing the scores and dividing by two. That is:

$$Mean\ Best\ Score\ (mod) = (Best\ Score\ (mod) + 2^{nd}\ Best\ Score\ (mod))/2$$

This produced mean scores shown in Table 8.19 below. Two overall best score means – the means of the means - were produced by summing the means for each user and dividing the answer by the number of users providing feedback. This was done to give an indication of any overall difference between the mean best scores for the modified and unmodified queries. Whereas doing this may not produce any significant result, it does provide a quantitative measurement of the difference in the rating of the IR results for each query. The mean of the mean best scores for the modified, *Mean Best Score (mod)*^{MEAN}, and the unmodified, *Mean Best Score (unmod)*^{MEAN} were:

Mean Best Score (mod)^{MEAN} = *7.69*

Mean Best Score (unmod)^{MEAN} = *6.83.*

Mean Best Score (unmod)	Mean Best Score (mod)
8	9
6	6.5
4	6.5
5.5	8
5	6.5
8.5	9.5
6	6
6.5	9
error	5.5
8	6.5
5.5	7
6.5	7.5
8	8.5
8.5	8.5
5	7.5
8.5	9
7.5	9
8.5	8
4.5	3
7.5	8
8	10
8	8

Table 8.19 Mean Best Scores for both the Modified and Unmodified Queries - Controlled Phase.

The above process was repeated for the scores given by the users to the worst two results, for both the modified and unmodified queries, termed ***Mean Worst Score (mod)*** and ***Mean Worst Score (unmod)***, respectively, again in the controlled phase of the user study. For example

Mean Worst Score (mod) = (Worst Score (mod) + 2^nd^ Worst Score (mod))/2

As before an overall mean score for the worst results, the mean of the means, both for the modified and unmodified queries, is obtained by summing the mean scores and dividing by the number of users providing feedback. Table 8.20 displays the mean scores whilst the means of the means are given below:

Mean Worst Score (mod)^{MEAN} = *3.38*

Mean Worst Score (unmod)^{MEAN} = *2.62*

Mean Worst Score (unmod)	Mean Worst Score (mod)
2.5	2.5
6	6
4.5	3.5
1.5	5.5
6.5	error
2.5	2
2.5	2
2.5	3.5
3.5	3.5
2.5	4
3	1
1.5	2.5
1.5	1
1.5	3.5
2	1
1	4.5
3	4.5
1	1
2	6
1.5	4
1.5	1.5
8	7
2	3

Table 8.20 Mean Worst Scores for both the Modified and Unmodified Queries - Controlled Phase.

The above calculation of the mean best and worst scores and subsequent means of the means was repeated for the data collected during the ad-hoc phase. Tables 8.21 and 8.22 show the mean scores whilst the means of the means are given below.

Mean Best Score (mod)^{MEAN} = 7.25

Mean Best Score (unmod)^{MEAN} = 6.93

Mean Worst Score (mod)^{MEAN} = 3.15

Mean Worst Score (unmod)^{MEAN} = 2.62.

Mean Best Score (unmod)	**Mean Best Score (mod)**
4	4.5
8.5	8
4	6
7.5	9.5
7	7
8	5
5.5	5.5
5.5	5.5
9.5	10
4.5	7
9	9
9	10
9	9
7.5	7.5
7.5	7

Table 8.21 Mean Best Scores for both the Modified and Unmodified Queries - Ad-Hoc Phase.

Mean Worst Score (unmod)	**Mean Worst Score (mod)**
1	1
5	4
1	1
2.5	4
1.5	1
1	1
1.5	2
1	1.5
6	7
1.5	2
none	none
6	7
5	7
1	2.5

Table 8.22 Mean Worst Scores for both the Modified and Unmodified Queries - Ad-Hoc Phase.

There is a difference between the mean score for the modified and unmodified query IR results and this is sufficient evidence to confirm that, from the results provided by the user-centred study, the FMQA is making a difference to the overall IR process.

8.4.2 Does the FMQA provide overall improvement?

This second question is, in affect, an extension of the first. It has been shown that the FMQA makes a difference to the IR process, but does involving it mean overall improvement. As discussed in chapter 7, *improvement* in the terms of the user-centred study is measured by the users' responses to the sets of IR results returned, both textual and numeric in terms of the scores. To answer this question, the main tests employed were McNemar's test (Reid & Boore, 1992), which was performed on the data within the 'textual response' category, and Wilcoxon's test (Reid & Boore, 1992), which was applied to the user scoring of the IR results. These tests were applied to data from both the controlled and ad-hoc phases of the user study.

However, the results from section 8.4.1 can also be used here to provide evidence to answer sub-question 2. A cursory examination of the mean scores for the modified and unmodified query IR results show that the modified means are, on the whole, higher than the corresponding unmodified means. The overall mean scores are unanimously higher and both these factors indicate that, on the whole, the users think there is improvement in using the FMQA. And, as already noted, 18 out of the 22 controlled user group respondents indicated textually that the modified query results were better. The outcome of performing McNemar's test on these data is reported below.

Test 1: McNemar's test on paired data from the controlled phase of the user study

Description: This test is particularly applicable to cases which consider 'before and after' situations in which the subject is used as its own control. In this case it can be applied to the textual response category of the users' feedback data in which they are asked to decide whether the umodified or modified query IR results are better. Recalling that the null hypothesis is that the FMQA is making no difference to the IR process, then in any distribution of users providing feedback the expected or control response would be 50% for each category of *modified better* and *unmodified better.* By performing McNemar's test on the actual distribution of responses, a test of how significant the data is from that expected is possible and evidence to address the null hypothesis is provided.

Result There were 19 users who expressed a preference for either the unmodified or modified IR results. Thus, in tabular form the actual and expected response distributions are:

	Modified better	Unmodified Better
Actual	18	1
Expected	9.5	9.5

Table 8.23 McNemar's Test on Paired Data (Textual Response) - Controlled Phase.

The test statistic, $\chi^2 = 15.2$.

To interpret this figure, statistical tables provide values which indicate levels of significance. The degree of freedom is 1 and χ^2 (0.1%) = 10.828. This implies that the result is significant to 0.1% level or that $P < 0.001$.

Test 2 McNemar's test repeated for the ad-hoc phase data

Description The same test was repeated but using the textual response data from the ad-hoc phase of the user study.

Result Only 11 users from this phase expressed a preference - the distributions are:

	Modified better	Unmodified Better
Actual	9	2
Expected	5.5	5.5

Table 8.24 McNemar's Test on Paired Data (Textual Response) - Ad-Hoc Phase.

The test statistic is $\chi^2 = 4.4$.

From tables, χ^2 (5%) = 3.84, therefore the ad-hoc response distribution, whilst not as significant as the controlled distribution, is still significant to the 5% level, i.e $P < 0.05$.

Test 3 McNemar's test repeated for the combined ad-hoc and controlled distributions

Description The data from the whole user study are tested using McNemar's test.

Result Combining the two phases produced a total of 30 users expressing a preference, giving the following distributions:

	Modified better	Unmodified Better
Actual	27	3
Expected	15	15

Table 8.25 McNemar's Test on Paired Data (Textual Response) - Both Phases Data Combined.

$\chi^2 = 19.2$.

As before, χ^2 (0.1%) = 10.828, and hence the overall response distribution is highly significant at the 0.1% level, P < 0.001.

The tests applied to the numeric and textual responses of the users obviously indicate that the FMQA is making some difference and it seems some improvement to the IR process. For example, the mean score of the best result in the controlled phase is higher for the modified query in 16 out of 22 cases, and for the worst result is equal or higher for the modified query in 15 out of 22 cases. The figures for the ad-hoc phase are 12 out of 14 and 11 out of 14, respectively.

In the controlled phase, the overall average score for the best results from the modified query is 7.69, as compared with 6.83 for the unmodified query. The figures for the worst results are 3.38 and 2.62, respectively. For the ad-hoc phase, the best result means are 7.25 and 6.93, the worst result means 3.15 and 2.62. Whilst the figures are closer, there is still evidence here from the users that the FMQA is making overall improvement to the IR process.

The results from applying McNemar's test to the textual response add weight to this opinion. In the controlled phase 18 out of 19 who expressed a preference, preferred the modified results, giving a χ^2 value which implied significance at the 0.1% level. Even the less numerous ad-hoc data implied significance at 5% and overall the significance in favour of the modified results was again 0.1%.

Though the means tests are good for indicating differences and improvement within different categories, another test can be performed on the scores as they are, in effect, paired data. That is, the values given by the users to each of the four results from the modified and original query IR results can be paired as follows:

best score (unmod) and ***best score (mod)***
2^nd^ best score (unmod) and ***2^nd^ best score (mod)***
worst score (unmod) and ***worst score (mod)***
2^nd^ worst score (unmod) and ***2^nd^ worst score (mod)***

Wilcoxon's test (Reid & Boore, 1992) can be performed to see if there is any significant difference between the paired sets of data. The test is a non-parametric test which is particularly applicable here as it is designed for paired readings – the test is useful because it takes account not only of the difference in the 'before and after' situation, but also of the extent of that difference. Therefore, the outcome will provide further evidence that there is significant difference between the two sets of data, arising from the modified and unmodified query results, and reinforce the supposition that the FMQA improves the performance of the search engine.

Wilcoxon's test is a signed ranks test which involves ranking the differences between the paired data. Differences between the two scores are ranked. Positive and negative differences are separated, signs are ignored and the smaller total is the statistic. The value is compared with values in a statistical table to assess significance. Note that, interestingly, in this case if the calculated statistic is greater than the table value the null hypothesis would be accepted.

The test was performed upon the data for the paired datasets outlined above and the actual data used and the calculation involved in each case can be found in Appendices G and H. In the tables below, Tables 8.26 and 8.27, the test results are given for the ***Mean Best Scores*** and ***Mean Worst Scores*** for the controlled phase, respectively. Also included in the tables are the P-values, the mean of the means and the *Median* of the means. Tables 8.28 and 8.29 show the results for the ad-hoc phase data. The results for the other paired datasets, for both phases, can be found in Appendix H.

Controlled Phase

Statistic	**Value**
Mean of ***Mean Best Scores (mod)***	7.69
Mean of ***Mean Best Scores (unmod)***	6.83
Median of ***Mean Best Scores (mod)***	8
Median of ***Mean Best Scores (unmod)***	7.5
T-value (Wilcoxon test)	25
Degrees of freedom	18
T(5%) (from tables)	40
P-value	< 0.05

Table 8.26 Wilcoxon's Test Statistic, T, P-values, Mean of and Median of Mean Best Scores – Controlled Phase.

Statistic	Value
Mean of ***Mean Worst Scores (mod)***	3.38
Mean of ***Mean Worst Scores (unmod)***	2.62
Median of ***Mean Worst Scores (mod)***	3.5
Median of ***Mean Worst Scores (unmod)***	2.5
T-value (Wilcoxon test)	24.5
Degrees of freedom	15
T(5%) (from tables)	25
P-value	< 0.05

Table 8.27 Wilcoxon Test Statistic, T, P-values, Mean of and Median of Mean Worst Scores – Controlled Phase.

From the two tables it can be seen that the paired datasets show significant difference at the 5 % level for both the ***Mean Best Scores*** and ***Mean Worst Scores***. This significance is repeated in the majority of the results for the other paired datasets in the controlled phase (see Appendix G).

To interpret the test results reported in tables 8.26 and 8.27, the null hypothesis formulated at the beginning of this chapter needs to be recalled (see page 126). With that in mind, the significance at the 5% level displayed by the data implies that the null hypothesis can be rejected on this evidence. In other words, the results of applying Wilcoxon's tests to these data add support to the assertion that the FMQA is improving the IR process.

Note from the two tables that the ***Median*** of the mean scores, for both the best and worst results, is higher for the modified query dataset. This adds further support to the assertion that the FMQA is improving the process.

Ad-hoc Phase

The Wilcoxon tests were repeated for the user scored data collected during the ad-hoc phase. The outcome was somewhat different and is shown in tables 8.28 and 8.29.

Statistic	Value
Mean of ***Mean Best Scores (mod)***	7.25
Mean of ***Mean Best Scores (unmod)***	6.93
Median of ***Mean Best Scores (mod)***	7
Median of ***Mean Best Scores (unmod)***	7.5
T-value (Wilcoxon test)	12
Degrees of freedom	9
T(5%) (from tables)	5
P-value	No Sig.

Table 8.28 Wilcoxon Test Statistic, T, P-values, Mean of and Median of Mean Best Scores – Ad-hoc Phase.

Statistic	Value
Mean of ***Mean Worst Scores (mod)***	3.15
Mean of ***Mean Worst Scores (unmod)***	2.62
Median of ***Mean Worst Scores (mod)***	2
Median of ***Mean Worst Scores (unmod)***	1.5
T-value (Wilcoxon test)	11.5
Degrees of freedom	11
T(5%) (from tables)	10
P-value	No Sig.

Table 8.29 Wilcoxon Test Statistic, T, P-values, Mean of and Median of Mean Worst Scores – Ad-hoc Phase.

Therefore, even though the mean of the mean scores, both best and worst, show a difference in favour of the modified scores, Wilcoxon's test result implies that this difference is *not* significant. This result is generally repeated in the other Wilcoxon tests on the ad-hoc paired datasets (see Appendix G).

Hence, contrary to the situation implied by the analysis of the controlled phase data, the evidence provided by the ad-hoc data seems to imply that the null hypothesis is correct and that the FMQA is making no significant difference to the IR process for the ad-hoc users. The ***Median*** of the mean scores for this phase provides conflicting evidence – for the best scores, the unmodified query value is greater, but the situation is reversed for the worst scores.

Two reasons could perhaps account for the difference between the controlled and ad-hoc results here. One is the lack of data to apply to the tests from the ad-hoc phase. Only 14 out of 75 users who took part in the ad-hoc phase completed the scoring part of the study. The nature of the Wilcoxon test (Reid & Boore, 1992) means that the degrees of freedom (dof) depends on the number of ranked differences between the paired data sets. The less number of pairs at the outset means a smaller dof which reduces the sensitivity of the test. The amount of data available for the tests may be insufficient to display the difference that, according to the controlled phase results, the FMQA is making to the IR process.

The second reason may be that the users who provided the data in the ad-hoc phase are at the boundary or outside the type of user the FMQA is thought to be able most to aid.

Therefore, although the above test results provided evidence to support the premise that the FMQA is improving the IR process overall and especially for the users in the controlled phase, further analysis was needed to discover whether this improvement can be considered to be generally affecting all aspects of the user study, e.g. across all categories of fuzzy user model. This analysis might also shed more light on the effect the FMQA had, if any, during the ad-hoc phase.

8.4.3 Does the FMQA improve the IR Process for different categories of user?

Further analysis on the mean scores was performed in order to provide evidence to answer this question. The mean score data was stratified according to two different categories of the summary statistics, model category and type of modification. This was done for both the controlled and ad-hoc phases of the user study. In each case, the results for the controlled phase are discussed first.

Mean Scores versus Model Category – Controlled Phase

Mean Best Score (unmod)	Mean Best Score (mod)	Mean Worst Score (unmod)	Mean Worst Score (mod)	Model Category
8	9	2.5	3.5	Med-Low
6	6.5	1.5	2.5	
4	6.5	1.5	1	
5.5	8	1	4.5	Med-Med
5	6.5	3	4.5	
8.5	9.5	2.5	2.5	High-Low
6	6	6	6	
6.5	9	1.5	5.5	
8	6.5	3.5	3.5	
5.5	7	2.5	4	
6.5	7.5	3	1	
8	8.5	1.5	3.5	
8.5	8.5	1	1	
5	7.5	2	6	
8.5	9	1.5	4	
7.5	9	1.5	1.5	
8.5	8	2	3	
4.5	3	4.5	3.5	High-Med
7.5	8	2.5	2	
8	10	2	1	
8	8	8	7	

Table 8.30 Mean Score Data Stratified by Model Category - Controlled Phase.

Here, the mean score data is stratified according to factor of ***Model Category***. Recall, as discussed in chapter 6 (page 90), for purpose of query refinement, the defuzzified value for each fuzzy representation of the user's WWW and subject experience is classified as *Low*, *Medium* or *High*, according to which range of values it belongs to, to the greatest degree. From this classification, there are then nine possible values of Model Category, *Low-Low* through to *High-High*. However, interestingly for the controlled phase, all users were classified in one of four of the possible nine values (see section 8.3.1, page 130). Therefore, the mean score data stratified by Model category produces table 8.30.

Taking the first value in the Model Category column, *Medium-Low*, this means that the modelling process has interpreted any user with this value to be at an *Intermediate* level with regard to WWW experience and at a *Novice* level with regard to subject experience. As can be seen, there are three rows of data associated with this model value. Examination of the mean scores shows that in every case ***Mean Best Score (mod)*** and ***Mean Worst Score (mod)*** are greater than ***Mean Best Score (unmod)*** and ***Mean Worst Score (unmod)***, respectively. Taking the mean the data in these rows produces the following values:

Mean Best Score (mod)^{MEAN} ***(Med-Low)*** = 7.33

Mean Best Score (unmod)^{MEAN} ***(Med-Low)*** = 6.00

Mean Worst Score (mod)^{MEAN} ***(Med-Low)*** = 2.33

Mean Worst Score (unmod)^{MEAN} ***(Med-Low)*** = 1.83

As expected, these values show that, for users considered to be novice with regard to AI and at an intermediate level with regard to the WWW, the FMQA provides an improvement to the IR results.

For the model category *Med-Med*, the data provides inconclusive evidence as there are only two rows to consider. However, what data there is supports the notion that for this category of user as well, the FMQA offers improvement. Again, the modified query scores are higher than the unmodified values in each case.

The next model category *High-Low* offers the most data and the best opportunity to support the claims for the FMQA. There are 12 rows of data to consider in this category which is grouping together those users considered by the modelling process to be *Expert* in terms of WWW experience and to be novice in terms of their subject experience. Examining the data for the mean best scores, it can be seen that ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in eight instances, equal in two and is only less than the unmodified score (by half a point) in one case. For the mean worst scores, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score***

(unmod) in six, equal to in five, and again only less than in one case. Taking the mean of the scores produces the following values:

Mean Best Score (mod)^{MEAN} ***(High-Low)*** = 8.00

Mean Best Score (unmod)^{MEAN} ***(High-Low)*** = 7.08

Mean Worst Score (mod)^{MEAN} ***(High-Low)*** = 3.45

Mean Worst Score (unmod)^{MEAN} ***(High-Low)*** = 2.37

These values seem to indicate that for this category of user the FMQA is improving the IR process.

The final model category which was identified in the controlled phase was *High-Med.* There are four rows of data and here the data does not completely follow the same pattern as in the other model categories. For the best mean scores, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in two cases, equal to in one, and less than in one instance. This produces mean values of:

Mean Best Score (mod)^{MEAN} ***(High-Med)*** = 7.25

Mean Best Score (unmod)^{MEAN} ***(High-Med)*** = 7.00

This indicates a slight advantage in favour of the FMQA modified query but the difference is small. The suspicion that the data in this model category is bucking the trend is supported by examining closely that portion of table 8.30 which covers the worst mean scores:

Mean Worst Score (unmod)	Mean Worst Score (mod)	Model Category
4.5	3.5	High-Med
2.5	2	
2	1	
8	7	

Table 8.31 Portion of Table 8.30 for Mean Worst Score Data Stratified by Model Category High-Med - Controlled Phase.

As is obvious, the worst unmodified mean score, ***Mean Worst Score (unmod)*** is greater than the modified, ***Mean Worst Score (mod)***, in all four cases, giving overall mean values of:

Mean Worst Score (mod)^{MEAN} ***(High-Med)*** = 3.37

Mean Worst Score (unmod)^{MEAN} ***(High-Med)*** = 4.25

One explanation of this maybe that the users considered to be in this model category, whilst still considered to be at an *Intermediate* level in the subject area, are considered to be *Expert* in terms of WWW experience and therefore are perhaps at the boundary of the type of user that the FMQA is likely to aid the most. Recall that, as discussed in chapter 5 (see page 63), it was thought that the FMQA was most likely to help novice and intermediate users, both in terms of WWW and subject experience. One of the surprises of the whole user study was that none of the users was modelled as being a novice in terms of the WWW, perhaps reflecting the rapid spread of the use of the global network. The relatively small failure of the FMQA to improve the process for the *High-Med* model category users, in terms of the IR results they considered the worst, is greatly offset by the rest of the results of the analysis in this section. An analysis which shows that for three types of modelled user the FMQA does make a positive difference.

The question now arises as to whether the FMQA repeated this promising performance during the ad-hoc part of the user study. Whilst the subjects of the controlled part were identified and chosen as likely to be of a type that the FMQA could aid (as has proven to be the case), there was no control over the constituency of the ad-hoc group. Therefore, it was worth performing the same stratification of the mean score data for this phase of the user study to see how the FMQA performed in more 'live' conditions.

Mean Scores versus Model Category – Ad-Hoc Phase

As in the controlled phase analysis, the data when stratified by model category produced interesting results. Examining the summary statistics shows that all users that took part in this phase of the study were modelled to be in six of the model categories (see table 8.13, page 135 in this chapter), and that none of these were considered to be in the *Low* category for WWW experience. Unfortunately, the stratification of the user scores shows that all those who responded to the email feedback request positively fell into three model categories, as is evidenced in table 8.32.

However, the data within these three categories is sufficient for the analysis to allow some conclusions to be drawn. The first aspect to note is that all the users who provided feedback were considered by the modelling process to be in the *High* category of WWW experience, in other words to be *Expert* in this category. This means that, in this part of the analysis, it is possible to compare directly the effect of the FMQA for the three different categories of subject model that the FUM process can attribute to a user, *Low*, *Medium*, and *High*, and the WWW experience factor can be ignored.

Mean Best Score (unmod)	Mean Best Score (mod)	Mean Worst Score (unmod)	Mean Worst Score (mod)	Model Category
8.5	8	5	4	High-Low
7.5	9.5	2.5	4	
7	7	1.5	1	
4	4.5	1	1	High-Med
4	6	1	1	
5.5	5.5	1.5	2	
9.5	10	6	7	
4.5	7	1.5	2	
9	9	6	7	
7.5	7	1	2.5	
8	5	1	1	High-High
5.5	5.5	1	1.5	
9	10	none	none	
7.5	7.5	5	7	

Table 8.32 Mean Score Data Stratified by Model Category - Ad-Hoc Phase.

For the first model category, *High-Low*, there are three rows of data. The best mean modified score, ***Mean Best Score (mod)*** is greater than the unmodified score, ***Mean Best Score (unmod)*** in two cases and less than it in one. In the worst mean scores, the situation is reversed, but when the means of these data are taken the following results are achieved:

Mean Best Score (mod)^{MEAN} (High-Low) = 8.16

Mean Best Score (unmod)^{MEAN} (High-Low) = 7.66

Mean Worst Score (mod)^{MEAN} (High-Low) = 3.00

Mean Worst Score (unmod)^{MEAN} (High-Low) = 3.00

Therefore, the FMQA can be said to be making a small but positive difference for those users considered to be *Novice* in the subject model category.

The second model category, *High-Med*, provides more data and more support for the FMQA improving the IR process. There are seven rows of data and examining this data shows that ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in four cases, equal in two, and only less in one. In the worst mean score data, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in five cases and equal in the other two. This produces mean results as follows:

Mean Best Score (mod)^{MEAN} (High-Med) = 7.00

Mean Best Score (unmod)^{MEAN} (High-Med) = 6.28

Mean Worst Score (mod)^{MEAN} (High-Med) = 3.21

Mean Worst Score (unmod)^{MEAN} (High-Med) = 2.57

These results show that, for the ad-hoc users at the *Intermediate* level in terms of subject experience, the FMQA also has a positive effect in aiding the IR process. Also, unlike the same category in the controlled phase the improvement is seen across all the data.

The final model category that appears in the table, *High-High*, should be populated with data that indicates the FMQA has no positive effect for these users. These users are considered by the modelling process to be *Expert* in terms of both WWW and subject experience, and are thus not thought to be the type of user the FMQA can aid. However, in terms of the best mean scores ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in one case, equal in two, and less than in one. In the worst mean scores, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in two case and equal in the other. Taking the means of these data produced the following values:

Mean Best Score (mod)^{MEAN} ***(High-High)*** = 7.00
Mean Best Score (unmod)^{MEAN} ***(High-High)*** = 7.50
Mean Worst Score (mod)^{MEAN} ***(High-High)*** = 3.16
Mean Worst Score (unmod)^{MEAN} ***(High-High)*** = 2.33

Hence, the FMQA has, for this category of user the subject and WWW expert, a slightly worse performance in terms of the best results but a better performance in terms of the worst results.

Overall, for the ad-hoc phase, the results of the above analysis support the assertion that the FMQA makes an improvement for different types of user, in terms of the model category in which they are placed by the modelling process.

A second way to consider the user is not by the model category they are placed in but by the type of modification the FMQA applies to the query. Recall from the summary statistics (page 128) that there are four possible types of modification – *C (Contracted)*, *S (Slight)*, *E (Extra)* and *G (Greatly)*. The user scored data can be stratified by this category to see if the FMQA improves the process across all types of modification. This analysis is reported next.

Mean Scores versus Type of Modification – Controlled Phase

The stratification of the mean scores by type of modification produces Table 8.33.

Mean Best Score (unmod)	Mean Best Score (mod)	Mean Worst Score (unmod)	Mean Worst Score (mod)	Query Modification
8	9	2.5	3.5	C
6	6.5	1.5	2.5	
4	6.5	1.5	1	
8.5	9.5	2.5	2.5	S
6	6	6	6	
6.5	9	1.5	5.5	
8	6.5	3.5	3.5	
5.5	7	2.5	4	
6.5	7.5	3	1	
8	8.5	1.5	3.5	
8.5	8.5	1	1	
5	7.5	2	6	
8.5	9	1.5	4	
7.5	9	1.5	1.5	
8.5	8	2	3	
4.5	3	4.5	3.5	E
7.5	8	2.5	2	
8	10	2	1	
8	8	8	7	
5.5	8	1	4.5	
5	6.5	3	4.5	

Table 8.33 Mean Score Data Stratified by Type of Query Modification - Controlled Phase.

As can be noted, there are three types of modification represented in the controlled phase mean score data. None of the users who provided complete feedback in this phase of the study was modelled as requiring their original query to be modified *Greatly (G)*. This is in keeping with the user cohort being targeted as probably consisting of novice and intermediate level users, and with the design of the FMQA (see chapter 5) which applies more extensive modification to the original query the more experienced the user is considered to be by the system.

There are three rows of data pertaining to the modification factor *C* and in all three cases, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)***. In two out of three, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)***. Taking the mean of these scores produces the following values:

Mean Best Score (mod)^{MEAN} ***(C)*** = 7.33

Mean Best Score (unmod)^{MEAN} ***(C)*** = 6.00

Mean Worst Score (mod)^{MEAN} ***(C)*** = 2.33

Mean Worst Score (unmod)^{MEAN} ***(C)*** = 1.83

These results show that for this modification factor, the modified query results are considered by the controlled phase users to be better overall.

For the modification factor *S* there are 12 rows of data. Here, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in eight cases, eqial to in two and less than in two cases. For the mean worst results, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in six cases, equal to in five and less than in only one instance. These promising results produce the following values when the means are taken.

Mean Best Score (mod)^{MEAN} ***(S)*** = 8.00

Mean Best Score (unmod)^{MEAN} ***(S)*** = 7.25

Mean Worst Score (mod)^{MEAN} ***(S)*** = 3.46

Mean Worst Score (unmod)^{MEAN} ***(S)*** = 2.38

These results show that for this modification factor as well the FMQA is having an overall positive effect on the IR process.

The final modification factor in table 8.33, *E*, relates to six rows of mean score data. For the best mean scores, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in four cases, equal in one and less than in one case. For the worst mean scores the situation is different – ***Mean Worst Score (mod)*** is only greater than ***Mean Worst Score (unmod)*** in two cases and is less than in the other four. However, taking the mean of the scores produces the following values:

Mean Best Score (mod)^{MEAN} ***(E)*** = 7.25

Mean Best Score (unmod)^{MEAN} ***(E)*** = 6.41

Mean Worst Score (mod)^{MEAN} ***(E)*** = 3.75

Mean Worst Score (unmod)^{MEAN} ***(E)*** = 3.50

Therefore, overall there is improvement in the modified scores in comparison with the unmodified scores and this improvement is not restricted to only one type of modification. In other words, the aid to the IR process the FMQA appears to provide for the controlled phase cohort of users is independent of the degree of modification made to the original query. The next subsection examines whether this general effect was repeated during the ad-hoc phase.

Mean Scores versus Type of Modification – Ad-Hoc Phase

Stratification of the mean score data by type of modification produces Table 8.34.

Mean Best Score (unmod)	Mean Best Score (mod)	Mean Worst Score (unmod)	Mean Worst Score (mod)	Query Modification
8.5	8	5	4	S
7.5	9.5	2.5	4	
7	7	1.5	1	
4	4.5	1	1	E
4	6	1	1	
5.5	5.5	1.5	2	
9.5	10	6	7	
4.5	7	1.5	2	
9	9	6	7	
7.5	7	1	2.5	
8	5	1	1	G
5.5	5.5	1	1.5	
9	10	none	none	
7.5	7.5	5	7	

Table 8.34 Mean Score Data Stratified by Type of Query Modification - Ad-Hoc Phase.

As can be noted, three of the possible four types of modification appear in the table, but in this table it is *S*, *E* and *G*. None of the ad-hoc users who provided scores of the IR results were modelled by the system as requiring their original query to be *C (Contracted)*. This may reflect that the ad-hoc users, accessing the system in a 'live' situation from many different remote locations, considered themselves to be fairly experienced in terms of the WWW and the query subject.

From table 8.34, it can be seen that three rows of data relate to the modification type *S*. Considering the best mean scores, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in one case, equal to in one case and less than in the other instance. For the worst mean scores, ***Mean Worst Score (mod)*** is less than ***Mean Worst Score (unmod)*** in two cases and is greater than in the other case. Taking the mean of these scores produces the following values:

Mean Best Score (mod)^{MEAN} (S) = 8.16

Mean Best Score (unmod)^{MEAN} (S) = 7.66

Mean Worst Score (mod)^{MEAN} (S) = 3.00

Mean Worst Score (unmod)^{MEAN} (S) = 3.00

Therefore, overall there is a slight improvement afforded by the use of the FMQA.

For the modification type *E* there are seven rows of data and ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in four cases, equal in two and less than in one case. In the worst mean score data, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in five cases and equal in two. This produces the following overall mean values:

Mean Best Score (mod)^{MEAN} (E) = 7.00

Mean Best Score (unmod)^{MEAN} (E) = 6.28

Mean Worst Score (mod)^{MEAN} (E) = 3.21

Mean Worst Score (unmod)^{MEAN} (E) = 2.57

These results show that the modified scores show improvement over the unmodified scores generally for this modification type.

The final modification type, *G* is related to four rows of data for the best mean scores and three for the worst mean scores (one user provided only partial feedback). Here, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in one case, less in one case and equal in two. Considering the worst mean scores, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in two cases and equal in the other instance. Taking the mean of these scores produces the following values:

Mean Best Score (mod)^{MEAN} (G) = 7.00

Mean Best Score (unmod)^{MEAN} (G) = 7.50

Mean Worst Score (mod)^{MEAN} (G) = 3.16

Mean Worst Score (unmod)^{MEAN} (G) = 2.33

Hence, these results show that for the IR results achieved via the FMQA modifying the query greatly, the ad-hoc users consider the unmodified query IR results to be better. This outcome is at odds with the other results. The difference is slight and it may be that the lack of data to base the analysis on is a factor preventing any difference from emerging. Another reason may be that the FMQA, designed to help users who are novice or at an intermediate level, was operating at the boundary of its effectiveness as it was interacting with users who are considered to be expert. However, the mean worst scores do show improvement for the modified scores and reinforce the other analysis results that show the positive effect of the FMQA is not confined to one modification type.

The considered conclusion to be drawn from the analysis of the user-scored data, stratified by model category and by type of modification, is that the FMQA is making a difference, and that difference is, on the whole, positive. This improvement occurs within different categories of user, and is not confined to one particular type of user, in terms of their model category and the extent to which the FMQA modifies their original query.

The next question to address is whether the FMQA provides improvement generally across different original or unmodified queries or whether the improvement is confined to one or two topics. This was answered by analysing the user-scored data after stratification by the category of original query. This analysis is reported next.

8.4.3 Does the FMQA improve the IR Process for Different Topics of Original Query?

Controlled Phase

Stratifying the mean score data by the original or unmodified query produces table 8.35.

Mean Best Score (unmod)	Mean Best Score (mod)	Mean Worst Score (unmod)	Mean Worst Score (mod)	Unmodified Query
8.5	9.5	2.5	2.5	Genetic Algorithms
6.5	9	1.5	5.5	AI Programming Languages
6	6	6	6	Natural Language Processing
4.5	3	4.5	3.5	
7.5	8	2.5	2	Neural Networks
8	9	2.5	3.5	
8	6.5	3.5	3.5	
5.5	7	2.5	4	Machine Learning
6.5	7.5	3	1	
6	6.5	1.5	2.5	
4	6.5	1.5	1	
8	8.5	1.5	3.5	Knowledge-based Systems
8	10	2	1	
5.5	8	1	4.5	
5	6.5	3	4.5	
8.5	8.5	1	1	Fuzzy Logic
5	7.5	2	6	
8.5	9	1.5	4	
7.5	9	1.5	1.5	
8	8	8	7	
8.5	8	2	3	

Table 8.35 Mean Score Data Stratified by Original or Unmodified Query - Controlled Phase.

Examination of the table reveals that there is insufficient data to consider the first three queries in any detail. The one row of mean score data for *Genetic Algorithms* and *AI Programming Languages* is not enough to draw any conclusions even though the mean modified query scores are equal or better than the unmodified query scores. Similarly, for the unmodified query *Natural Language Processing* the situation is inconclusive because of the paucity of data.

However, the other four queries featured in the table deserve further consideration as the data is numerous enough to provide more evidence on the effect of the FMQA. Looking at the three rows of data for the query *Neural Networks* shows that in two cases the mean best modified score, ***Mean Best Score (mod)***, is greater than its counterpart, ***Mean Best Score (unmod)***. For the mean worst scores, the situation is the modified score is better in one instance, equal in another and worse in the third. Taking the mean of the scores in these three rows produces the following values:

Mean Best Score (mod)*^{MEAN}* *(Neural Networks)* = 7.83
Mean Best Score (unmod)*^{MEAN}* *(Neural Networks)* = 7.83
Mean Worst Score (mod)*^{MEAN}* *(Neural Networks)* = 3.00
Mean Worst Score (unmod)*^{MEAN}* *(Neural Networks)* = 2.83

As can be seen there is very little difference if any between the best and worst scores when the mean is calculated. This perhaps reflects a combination of the natural smoothing out effect that taking the mean has with the fact there is only three rows of data to consider.

Considering the data which pertains to the original query *Machine Learning* sheds more light on the effect of the FMQA. There are four rows of data to consider and in every case, for the mean best scores, ***Mean Best Score (mod)*** is higher. Again, for the mean worst scores the situation is even, with ***Mean Worst Score (mod)*** being higher than ***Mean Best Score (unmod)*** in two cases and vica versa. Producing the means of the data in these rows gave the following results:

Mean Best Score (mod)*^{MEAN}* ***(Machine Learning)*** = 6.88
Mean Best Score (unmod)*^{MEAN}* ***(Machine Learning)*** = 5.50
Mean Worst Score (mod)*^{MEAN}* ***(Machine Learning)*** = 2.83
Mean Worst Score (unmod)*^{MEAN}* ***(Machine Learning)*** = 2.83

These show a marked improvement in the best mean score for the modified query results and at least no loss in performance in the worst mean scores.

For the original query *Knowledge-based Systems* there are four rows of data as well but here the effect of the FMQA is clearer. Again, ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in all four cases but the users have attributed greater differences to the scores overall. For the worst mean scores, ***Mean Worst Score (mod)*** is clearly greater than ***Mean Worst Score (unmod)*** in three cases. Taking the mean of the data produces the following results:

Mean Best Score (mod)^{MEAN} (KBS) = 8.25

Mean Best Score (unmod)^{MEAN} (KBS) = 6.63

Mean Worst Score (mod)^{MEAN} (KBS) = 3.38

Mean Worst Score (unmod)^{MEAN} (KBS) = 1.88

These results show that, for the original query *Knowledge-based Systems*, the FMQA is making a marked improvement in the IR results.

For the final query in table 8.35, *Fuzzy Logic*, there is the most amount of data to consider – six rows. Again, the data shows a bias in favour of the modified query scores: ***Mean Best Score (mod)*** is greater than ***Mean Best Score (unmod)*** in five cases and equal in the sixth; ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in three cases, equal in two and only less (by 1 mark) in one instance. Taking the mean of this data produces the following results:

Mean Best Score (mod)^{MEAN} (Fuzzy Logic) = 8.33

Mean Best Score (unmod)^{MEAN} (Fuzzy Logic) = 7.66

Mean Worst Score (mod)^{MEAN} (Fuzzy Logic) = 3.75

Mean Worst Score (unmod)^{MEAN} (Fuzzy Logic) = 2.66

The values above reinforce the notion that the FMQA is not only making a difference but is making an improvement to the IR results. The analysis reported in this section indicates that, on the whole, this improvement cuts across a number of the original queries and is not confined to one or two, and therefore the FMQA is able to improve the IR process for different users with different needs. The ad-hoc phase data are considered next.

Ad-Hoc Phase

As in the controlled case, the mean score data was stratified by original or unmodified query. This produced Table 8.36.

Mean Best Score (unmod)	Mean Best Score (mod)	Mean Worst Score (unmod)	Mean Worst Score (mod)	Unmodified Query
9.5	10	6	7	AI Programming Languages
8	5	1	1	Natural Language Processing
9	9	6	7	Genetic Algorithms
4.5	7	1.5	2	Machine Learning
7.5	7	1	2.5	
4	6	1	1	Neural Networks
9	10	none	none	
7.5	9.5	2.5	4	Fuzzy Logic
4	4.5	1	1	
8.5	8	5	4	Bayesian Networks
7	7	1.5	1	
5.5	5.5	1.5	2	Distributed AI
5.5	5.5	1	1.5	
7.5	7.5	5	7	

Table 8.36 Mean Score Data Stratified by Original or Unmodified Query - Ad-Hoc Phase.

As can be noted, any incisive analysis was hampered by the paucity of data. Only 14 users in this phase completed the scoring part of the feedback section of the study (only 13 fully), reflecting the entirely voluntary nature of the study. For the majority of original query values, there is only one row of data, and therefore, for these, nothing conclusive can be drawn except to say the FMQA is making some difference. In those cases where there are two rows of mean score data, *Fuzzy Logic*, *Machine Learning*, and *Bayesian Networks*, the FMQA is improving the IR results for the first two and only marginally failing in the third query. Again, two rows of data are not enough to give an authoritative opinion.

In the final original query category, *Distributed AI*, there are three rows of data. Here the mean best scores are identical in all three cases whereas in the mean worst scores, ***Mean Worst Score (mod)*** is greater than ***Mean Worst Score (unmod)*** in all three instances. The means of these values are:

Mean Best Score (mod)^{MEAN} (Distributed AI) = 6.83

Mean Best Score (unmod)^{MEAN} (Distributed AI) = 6.83

Mean Worst Score (mod)^{MEAN} (Distributed AI) = 3.50

Mean Worst Score (unmod)^{MEAN} (Distributed AI) = 2.50

These values imply that the FMQA is making a small positive difference in the instance of Distributed AI as the original query. Examining the table it can be argued that this is the case for the majority of different original queries, and thus the performance of the FMQA in the ad-hoc phase mirrors its performance in the controlled phase of the user study. However, owing to the lack of data, this assertion can only be weakly made and hence it is difficult to conclude that, for the ad-hoc phase, the FMQA provides positive aid to users with different query needs.

The final section of this chapter, section 8.5, draws the results of the analysis together and discusses them in general, offering conclusions that can be made in the light of the analysis.

8.5 Discussion and Conclusions

The preceding section reported on the analysis of the user study results. The crux of the analysis was a series of statistical tests applied in order to provide evidence with which to address a number of questions. The questions were formulated in such a way that answering them would allow the null hypothesis to be addressed. These questions were:

(1) Is the FMQA working, is it making a difference to the IR process?

(2) Is there an overall improvement in the results through the use of the FMQA?

(3) Does the FMQA improve results for different categories of user, specifically for those users deemed novice and intermediate?

(4) Does the FMQA improve results over the range of the fixed different topics within the one information domain?

The overall outcome of the analysis implies that evidence exists to answer each of these questions with a *Yes*. The overall mean scores produced by examining the user rated IR results clearly show that the FMQA is making a difference to the IR process and this difference can be interpreted as an overall improvement in the IR results in the opinion of the users. In section 8.4.2 McNemar's test on the textual feedback responses implied that this improvement was significant for both the controlled and ad-hoc groups, and significant to the 0.1% level for the controlled group.

The fact that the user-scored data represented paired data - unmodified and modified query scores - allowed Wilcoxon's test to be performed to test the significance of the difference in the scores. These tests showed that the difference is significant for the controlled group. Therefore, for these users, mostly categorized by the FMQA as novice or intermediate, the improvement provided is significant.

The situation seems different for the ad-hoc group. Whilst the mean overall scores show a small positive difference between the modified and unmodified IR query results, when Wilcoxon's test were applied this difference was shown to be not significant. As already noted, the lack of data in this group and the fact that the majority of the group's constituents are thought to be expert may be contributing factors to this outcome. The FMQA is making a difference but the amount of data collected is insufficient to allow a judgement on the significance of this difference. Addtionally, this difference may be limited, in the ad-hoc group case, by the fact the FMQA, as designed (see chapter 6), is thought to provide most aid to novice and intermediate users.

Further evidence that the FMQA improves the IR process for different types of user was provided in section 8.4.3 in which sub-question 3 was addressed. The mean scores were stratified by model category and by type of modification. The result of this process showed that, for almost all types of modification and different model categories represented in the data, the overall mean modified scores were higher than the unmodified scores. The only contradictory results were for the model category *High-High* and the modification type *G (Greatly)*, for the ad-hoc group of data. Again, this reversal can be accounted for by noting that the users providing the data in these categories were considered to be expert. Overall, the outcome of stratifying the data in this manner supports the assertion that the FMQA improves the IR results for more than one type of user, and that it provides most aid to novice and intermediate users

Section 8.4.4 reported in the results of stratifying the mean score data by original or unmodified query in order to address sub-question 4. Though the data was sparse for some topics of query, especially in the ad-hoc group, the analysis showed that the improvement afforded by the FMQA is exhibited across a number of topics, for example knowledge-based systems, fuzzy logic and machine learning.

The stratification of the data by various categories highlights the limitations of the analysis. One limitation is that the amount of data in each stratification was insufficient to allow further analysis, such as Wilcoxon's test, to test the significance of the difference within each category. Another limitation was the inability to perform any analysis by categories such as gender, background and age. This was because there was insufficient spread of data within these categories to warrant any stratification.

Despite these limitations, the overall conclusions to be drawn from the analysis is that the FMQA is making a difference, that this difference represents an overall improvement, and that the improvement is not limited to one type of user or to one particular query topic.

In other words, the analysis of the user study provides sufficient evidence to reject the null hypothesis and thus accept the research hypothesis stated on page 125 at the beginning of this chapter. This means that the research issue, the crux of this thesis, can be addressed in the light of the outcome of the analysis. The research issue, recall, was expressed as the following question:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics?

The results of the analysis imply that the answer to this question is *Yes*, the application of the FMQA to IR from the WWW does improve upon the results returned by Lycos™, for the particular conditions outlined in the research issue.

The next and final chapter of this thesis, chapter 9, presents an argument for the generalising of the results of the research. The whole research is reviewed and its strengths and weaknesses discussed. Limitations of the research and the effect of hindsight on its conduct are considered. Finally, conclusions are drawn and possible future directions for the research are offered.

CHAPTER 9 - CONCLUSIONS

9.1 Overview

The preceding chapter reviewed the data collected during the user study and then described the analysis of this data, reporting on the results of the analysis. The interpretation of the results was that the FMQA does provide overall improvement to the IR process in the particular conditions outlined by the research issue.

Before presenting an argument for the case that this research outcome can be generalised to other conditions, this final chapter of the thesis reviews the whole research, looking again at how the research focus was established and considering the strengths and weaknesses of the research during its development and conduct. Section 9.2 restates the focus of the research before presenting a review of the findings of the investigative research covered in chapters 2,3 and 4. In section 9.3 the Fuzzy User Modelling (FUM) approach and its implementation in the prototype FMQA are reviewed. This is followed by section 9.4 which discusses the development and conduct of the user study, and the subsequent reporting and analysis of the study results, originally reported in chapters 7 and 8, respectively. Section 9.5 offers the argument for generalising the research outcomes and discusses possible future directions for the research. The thesis is concluded with some final thoughts in section 9.6.

9.2 Literature Review

This section reviews the findings of chapters 2,3 and 4 which reported on the investigative research performed to underpin this research. Recall that the focus of the research was encapsulated by the following statement:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

However, before this final focus was achieved (in chapter 6, page 104), research into broader areas was performed. Initially, in chapter 1 of the thesis, *Introduction*, the background to the research was presented and this led to the first discussion of how the research could be focused. Recall that it is the on-going telecommunications and information revolution (Masuda, 1982; Tarjanne, 1997) which is fuelling an exponential explosion in on-line information sources that was the main factor behind the research. Perhaps the biggest development in recent years has been the part of the Internet (Lane-Lawley & Summerhill, 1994) known as the WWW (Berners-Lee *et al.*, 1992). This became the information upon which the research focused, primarily because its ever-growing, ever-changing nature means it is becoming one of the most popular information sources (however, it will be argued below that the research outcomes are not confined to the WWW).

It was then noted in chapter 1 that the increasing number and type of information sources had raised users' expectations and this had placed libraries and other information providers under increased pressure to provide access to these sources. This pressure, and the financial constraints faced by providers, has led to the belief that IR systems needed to incorporate the expert knowledge of the information professional with some understanding of the user's preferences and experience within the IR process. In other words, these systems needed to become 'intelligent' in some way, working with the user to satisfy their IR goal, improving the IR process, and easing the burden on the information professional.

Another factor behind the need for intelligent IR tools and a major consideration in this research was also raised first in chapter 1. This was the limitation of existing IR tools, particularly the problems of data irrelevancy and redundancy from which they suffer.

Therefore, the initial broad concern of this research was how the IR process, from information sources on the WWW, could be improved by being imbued with some sense of intelligence and with some representation of the user's experience and knowledge. Hence, it seemed natural to investigate the application of AI techniques to provide the intelligent component and to examine user modelling techniques to capture and represent the user's experience and knowledge.

In section 1.2, then, the initial research issue took shape – whether, and to what extent, the performance of IR from the WWW could be improved through the application of AI and user modelling techniques.

In order to address this issue, a number of areas were subjected to investigative research. An investigation of AI and IR was undertaken in order to decide which AI approach to employ. This was reported in chapter 3, but was predicated on background research looking at IR and electronic sources of information in general which was reported in chapter 2 and was undertaken primarily to underpin the background to the subject of the thesis. The main conclusions from chapter 2 were that there is sustained and growing interest in the research and development of electronic sources of information (Bansler *et. al*, 1984; Bourne, 1980; Follet *et. al*, 1993; Hartley *et. al*, 1987; Parfell, 1987; Saffedy, 1993; Saltzer, 1992), that this development is increasingly focused on networked sources (Collier, 1993; Elib, 1998; Eyre, 1997), and, perhaps most importantly, that integration of information sources is being driven by the WWW (Bishop *et al.* 1997; Fox *et al.*, 1997; Salampasis *et al.*, 1996).

Chapter 3 reported the investigation of *Intelligent Information Retrieval*, and the outcome of this investigation led to further focusing of the research issue. The examination showed that many aspects of AI have been used in IR, producing systems with some limited success. The main conclusion was that many of the AI approaches previously employed were inapplicable owing to the nature of the WWW. Many of the previous AI approaches relied heavily upon knowledge of the information corpus or database. The ever-changing nature of the WWW precluded their consideration in this research. It was concluded that to effectively apply intelligence to IR from the WWW, the research should focus upon query refinement. Two areas of soft computing, fuzzy logic and neural networks, showed promise and were further evaluated. This evaluation suggested that Fuzzy Logic (FL) was the AI approach to use and that the technique could be used to model the WWW user's experience and knowledge. Here, there arose a limitation of the thesis' research. Ideally, a number of AI techniques should have been chosen as candidates and then tested fully by the development of prototypes and subsequent comparison of their performance. However, limitations on time and resources prevented this course of action and a decision had to be made based upon the nature of the AI technique and the existing research.

The choice of FL is, however, a strength of the research. Its ability to encapsulate the linguistic and semantic 'vagueness' of human reasoning and conceptualising – activities that take place during the IR process – make FL a suitable proposition for solving the problems of IR (Chang & Chen, 1998; Hosono *et al.*, 1985; Mansfield & Fleischman, 1993; Nakamura & Iwai, 1982; Radecki, 1976; Terano *et al.*, 1992; Yager & Larsen, 1993; Zadeh, 1993).

Thus, the final conclusion was that FL should be used to represent user information and this information should then be employed to refine a user's query. Therefore, the research focus was narrowed to whether, and to what extent IR from the WWW could be improved by the application of FL and user modelling techniques to refine the user's query.

Following the decision explained in chapter 3 that FL should be used to model users' experience and knowledge, an investigation into user modelling and IR in general was performed to underpin the approach. This investigation was discussed in chapter 4, *The Role of User Modelling in IR.* It was revealed there that user modelling is a huge on-going area of research and that proper overviews of the work are difficult to achieve. Overviews are hampered by the existence of various definitions of the term 'user model'. It was noted that, within the realm of this research, it was those definitions that apply to a representation of the user's abilities, limitations beliefs and goals, and to a model of a typical user that might be referenced by a system designer that were considered relevant. Underlying issues involved in user modelling, such as the dimensions proposed by Kass and Finin (1988), were discussed before the chapter concentrated upon user modelling and intelligent information retrieval.

There has been much research in this field (Chang *et al.*, 1993; Chen and Norcio, 1992; Kok, 1991b; Jennings *et. al.*, 1992; Logan *et al.*, 1994; McTear, 1993) and the conclusion was that many of the early attempts could be criticised for not being tested on large information corpuses in real situations. The recent work on employing user models to improve IR from the WWW in particular was also discussed. There have been a number of systems developed which employ some sort of representation of the WWW user to filter information delivered via the WWW (Ackerman *et al.*, 1997; Cullen, 1997; Eichmann, 1996; Eichmann & Wu, 1996; Kamba *et al.*, 1997) with limited success. A crucial point made is that there has been a shift away from the information corpus itself towards the user, through the development of user models in the form of profiles or stereotypes (Bell, 1996; Etzioni *et al.*, 1997; Rich, 1979; Kerr, 1997; Yeates, 1998). A number of the systems suffer from having insufficiently flexible models and are unable to adapt to changing user characteristics.

The main conclusion was that user modelling could provide a way to capture and represent user characteristics which can then be used to assist IR from the WWW. This outcome combined with previous conclusions, led to the decision to use FL to develop user models that are adaptive and flexible, and that also intelligently represent information about the user. Noting that this approach was novel, chapter 5, *Intelligent Information Retrieval - A New Approach Employing Fuzzy User Modelling*, discussed it in detail. The findings of that chapter and chapter 6, which covered the implementation of the approach, are reviewed in section 9.3.

9.3 Fuzzy User Modelling and the Prototype

In chapter 5 a major problem with IR was noted – that the prospective user of IR systems often has only a vague concept of their information need or is unable to express that need with sufficient clarity and precision. However, the point is made that by applying FUM techniques to gather and represent users' knowledge and experience, an information need – formulated as a query – could be refined in the light of that knowledge and thus any IR results retrieved from a WWW search might be improved. Thus, the research issue was restated as:

> Whether the application of a prototype system, developed using FUM, can improve the IR results from the WWW, for novice and intermediate users, employing queries in one information domain.

The research issue was restricted to address one information domain to make the scope of the research manageable. It is argued below that it can be extended to cover other domains. During the conduct of the research, it was assumed that any system developed to assist IR from the WWW would most likely benefit those users whose experiences and knowledge of the WWW and of the information domain was not great, and thus the research issue was limited to addressing these types of user. (However, the results of the user study suggested that many users had more WWW experience than initially expected but their information domain knowledge was limited.)

The rest of chapter 5 discussed the FUM approach in detail. The concept behind the modelling of the WWW user was discussed. It was noted that, though the general concepts of novice, intermediate and expert had some relation to the established notion of user stereotypes (Rich, 1979), intuitively and in reality a user would move through these categories in a continuous way as they grew in experience and knowledge. This point was emphasised by the employment of sliding-scale questions to acquire the information regarding the user's experience and knowledge. The use of questionnaires in knowledge acquisition is well documented (Kobsa & Wahlster, 1989; Nessen, 1987) and answers to questions could be used to indicate membership of a category to a certain degree and build an individual user model which is flexible. Modelling of the user in this way is related to the concept of user profiles (Bell, 1996; Etzioni *et al.*, 1997) as an individual model can be considered to be a profile.

Chapter 5 then examined how FL could be employed to implement the user modelling technique. It was noted that in terms of the research, two different associated semantics are relevant. One is the use of fuzzy logic to represent incomplete or vague states of information (Zadeh, 1978) – a certain aspect of vagueness is inherent in the use of open-ended questions. The second, more applicable, semantics is that in which preferences are expressed between more or less acceptable solutions with respect to the constraint (Bellman & Zadeh, 1970). It was noted that the gradeness introduced by the use of fuzzy sets makes it the ideal technique to employ to create an individual user model from a series of overlapping default models. The answers to the questions set out in the questionnaires could be used to create the user model via the use of fuzzy logic because of this property of gradeness.

The actual shape of the fuzzy sets employed to represent the models and the genesis and combination of the fuzzy rules in the FKB were then discussed. The shape of the sets chosen was influenced by work in fuzzy logic and control processes (Kosko, 1992; Lee, 1990) but was decided intuitively. This is a weakness of the research but, as noted in chapter 5, to uncover the shapes from empirical evidence would have involved a long study of the users which was outside the terms of reference and scope of the research. However, it was noted that the shapes are not chosen arbitrarily but relied on the intuition that the greater the knowledge and experience of the user, the less novice (and consequently the more expert) they become.

The formulation of the FKB through the combination of fuzzy rules in FAM was then examined and presented. The rules were needed so that the answers to the questions could be used to modify the fuzzy user models. A strength of the research was the use of the FKB as fuzzy rules are much more flexible than traditional expert system rules. The knowledge gathered by the question-answer process could be retained and propagated throughout the modelling process.

In order to apply the knowledge contained within the user models, the final individual fuzzy sets have to be defuzzified, so that the knowledge can be interpreted and employed in query refinement. Issues of defuzzification were discussed and it was argued that the method employed (Cox 1994; Lee 1990), was as good as any other in this context. The current understanding of defuzzification relies more on heuristics than rigorous mathematics (Cox, 1994).

The chapter then discussed the interpretation of the numbers or scalars produced by defuzzification. It was suggested that one interpretation could be the higher the number, the greater the degree of knowledge and experience about the WWW or subject domain the user has. With this in place, the use of the numbers to develop a query refinement approach could then be offered.

As already noted, the research issue was limited to consider one information domain and thus two models are developed, one for the WWW and one for the information domain of AI. Though this could be thought of as a limitation of the research, it is not a limitation of the FUM approach as models for other information domains could easily be developed in the same way. This restriction allowed the research to retain its focus of being the application of a new approach to IR from the WWW.

The scope of the user's query was discussed next and chapter 5 notes that, again because of resource constraints and the research focus, the query was limited to a choice of keyword phrases representing major AI topics. Query refinement in other IR systems and research was then investigated and the prospect of applying existing refinement techniques examined. Some of these techniques rely on a detailed indexing of the IR database documents (Harman, 1992) and thus were deemed inapplicable to the present research. However, the approach of refining queries with regard to lexical-semantic relations did prove promising (Vorhees, 1994). As discussed in detail in chapter 5, the query refinement approach adopted in the present research became one in which the original query is refined by modification through recourse to a small lexicon of semantically related terms.

The query refinement process then became a matter of interpreting the defuzzified user model numbers in terms of how they were to be used to modify the original query. The basic concept of the method was that the lower the numbers the more the original query was expanded by addition of terms from the lexicon.

Chapter 5 concluded that the FUM approach had some weaknesses and some strengths. One weakness is that the individual models are fixed after the questionnaires have been completed, there being no provision for the use of user feedback or other technique to adapt the models. However, the models are very current as new ones are created each time the user accesses the system. A strength of the research is the direct knowledge acquisition via the questionnaires.

Another weakness was the intuitive way in which the fuzzy set representations and the FKB have been developed. However, this heuristic approach to the development is not unjustified, as there is support for it in previous fuzzy logic research (Baaklini & Mamdani, 1975; Brae & Rutherford, 1979; King & Mamdani, 1975; Van den Berg & Van Dijk, 1997). The overlapping fuzzy sets are an effective way to represent linguistic concepts such as novice and expert.

To test the research issue it was decided to employ FUM to refine queries before they were submitted to an existing WWW search engine. This would set up the possibility of comparing IR results for both the original (unmodified) and refined (modified) queries and thus offer a way for the research issue to be addressed. To do this necessitated the development and implementation of a prototype system and then its employment in a user study. Chapter 6 reported on the design and building of the system, termed the *Fuzzy Modelling Query Assistant* (FMQA).

A brief investigation of a number of ways in which fuzzy sets could be represented computationally led to the employment of a method developed by Cox (1994) which uses hybrid C/C++ code to model fuzzy sets as vectors. However, chapter 6 reported that, for the purposes of this present research, the code had to be extended and developed, as well as being re-engineered, compiled and implemented upon a platform which hitherto had not been done. For example, the largest fuzzy system described by Cox (1994) has five fuzzy rules, whereas eventually the FKB of the FMQA contained over 100. This accomplishment was a marked success of the research but its achievement used a lot of the resources available to the research in terms of time. This factor had an impact on the user study.

The aspects of the FMQA user interface – the on-line questionnaires, the query submit screen, the link between the FMQA and the search engine – were then discussed. A WWW interface was chosen and it consisted of a series of interlinked WWW pages or 'screens'. What may seem an obvious choice now was not so apparent in late 1995/early 1996, and the reasons for using it were not the popularity of the WWW but that it ensured compatibility with the interfaces of existing WWW search tools and thus avoided any bias being introduced by having different types of interface.

Lycos™ (Lycos, 1996) was chosen as the search to which the query – modified and unmodified – would be submitted. This was because, at that time, Lycos™ was thought to have the most complete indexed catalogue of WWW documents. It could be argued that presently it has been eclipsed by the catalogue at Alta Vista (Altavista, 1999a).

The implementation of the FUM approach was fully described, particularly how the fuzzy set representation in C/C++ was embedded in the WWW interface through the use of forms and CGI-scripts. The WWW forms allowed the user to provide information via the questionnaires and also to choose the original or unmodified query. The system automatically builds the models, defuzzifies the results, and uses the numbers in conjunction with the query to produce a modified query. The user did not see this functionality but was presented with a 'Query Submit Screen' which contained a WWW form that allowed either the unmodified or modified query to be submitted to Lycos™.

After query submission procedural control passed outside the realm of the FMQA and over to the search engine. The search engine returned a WWW page consisting of hypertext links to the 'top-ten' WWW documents which, according to the Lycos™ indexing algorithm, closely matches the query. To submit another query, say the modified after the unmodified, the user had to employ the 'back' button on their WWW browser.

This loss of control over procedure and the inability of the FMQA to collect and manipulate IR results before presentation to the user could be used to levy the criticism that the interface between the FMQA and Lycos™ was poorly designed. With hindsight, and the development in WWW applications such as Java (*Sun Microsystems*, 1998) since the prototype was built, the interface could be improved to allow control over the IR results to be retained. However, at the time of prototype development, it was not crucial to the examination of the research issue and it may have been difficult to achieve, given the commercial nature of Lycos™.

The strengths of the FMQA lay in its simple construction and WWW interface. Avoiding the sophistication of some IR tools obviated the need for the participants of the user study to learn how to navigate a difficult interface, something that may have adversely influenced the outcome of the study. Whereas the information collected via the questionnaires is lost every time a user finishes a session with the FMQA, this was seen as a strength of the system. The rebuilding of the fuzzy models each time meant that the FMQA was always using information that represented the user's current experience and knowledge. The 'loss' of the models was only a feature of the prototype and any future implementation could retain them.

It was concluded that the FMQA was sufficiently developed to be used to address the research issue – though its development meant the focus of the research was refined again to be expressed in the following statement:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

This is the research issue that was restated at the beginning of this chapter and was addressed by the conduct and analysis of the user study. Section 9.4 reviews the user study and the subsequent analysis of its results.

9.4 User Study and Data Analysis

The design and conduct of the user study was reported in chapter 7, *The User Study – An Investigation of the Performance of the FMQA*. This section reviews the findings of that chapter and those of chapter 8 which reported the analysis of the study results.

This was a study of real WWW users interacting with the FMQA. The study was driven by a *User Studies Design Protocol* (Appendix E) although, during implementation, some aspects of the study changed. Originally, there were meant to be two studies, with results from the first informing any changes to both prototype and further studies before the second took place. As already mentioned, the FMQA development had eroded the research resources and there were time constraints on the study. Problems with generating enough initial user interest exacerbated this situation and meant that the study had to be restructured.

The study became one of two phases – *Controlled* and *Ad-Hoc*. Certain types of user were targeted and asked to interact with the system in a more controlled environment, the results of which became part of the *Controlled Phase.*. However, as the FMQA had a WWW interface it was possible to place it in a 'live' situation and allow access to anyone with a suitable WWW browser. Results collected from user interaction in this way were considered to be part of the *Ad-Hoc Phase*. Chapter 7 describes the identification and choosing of suitable user groups for the controlled phase, which took place via an email questionnaire. This process was driven by the need to find users which could be deemed novice or intermediate in terms of the definitions given in chapter 5, so that the results of the user study could be used to address the research issue. This was the primary aim of the user study and this and other aims were detailed in chapter 7.

In order to satisfy the aims, some performance criteria had to be devised, justified and the study conducted so that the performance of the FMQA could be measured in terms of these criteria. An investigation, reported in chapter 7, into study of intelligent IR systems in terms of their performance, revealed that many studies had concentrated on measurement of statistical indicators such as 'precision' and 'recall', in which the system under study is compared with another, against a fixed document set (Belkin *et al.*, 1993; Callan & Bruce Croft, 1993; Harman, 1992). It was argued in chapter 7 that, as these criteria depend on a notion of relevancy, it is difficult to apply them to an evaluation of a tool trying to aid IR from the WWW, as it is almost impossible to count the number of relevant documents in an ever-changing document corpus.

Rather, an argument is made for a user-centred evaluation of the FMQA, citing Ingwersen's (1992) assertion that user-oriented IR research aims at improvement of IR effectiveness *within* the framework of the user and their desire for information. Consequently, chapter 7 describes the development of performance criteria based on the user ranking and rating the best and worst results in the 'top ten' list returned to them by Lycos™, both for the modified and unmodified queries. The users were also asked to provide explicit textual feedback on the performance of the FMQA and on the design and ease of use of the system – this feedback also proved invaluable when the results were analysed. Chapter 7 described how the user feedback data – both the scores and the textual portion – was collected and evaluated during the study. One weakness of the study is the lack of a truly objective performance criterion but the criteria chosen were designed to allow the research issue to be addressed from the user's point of view. The users were asked to provide feedback by completing a WWW form and submitting the information which was emailed to the researcher. One drawback of this procedure is that it led to less data that was suitable for analysis than was expected. The feedback part of the user interaction with the FMQA was entirely voluntary and this was reflected in the proportion of users who completed the feedback in comparison with the number who took part in the study overall (39 out of 114).

Though it would have been ideal, it was not feasible to have all users in the controlled phase take part at the same time. However, certain elements were controlled which effectively set these users apart from those who made up the ad-hoc phase. These elements were: the user group constituency, the timescale of the phase, the time of access to the FMQA, and the environment.

One weakness of the study is that it is difficult to access the impact of the environment on the results, especially for those users in the ad-hoc phase, and therefore these factors were controlled in the first phase in order to minimise their effects. In hindsight, any controlled part of the user study should be designed so that subjects took part simultaneously, but such a condition might have its own effects, such as increased network traffic, which would be difficult to quantify.

Despite the problems already considered, the user study produced sufficient and useful data to allow a proper analysis. The user study results and the subsequent analysis were reported in chapter 8. The main thrust of the analysis of the results was to provide evidence to address the research issue. In chapter 8 this issue was reformulated as the following null hypothesis

> There is no significant improvement in the results obtained by the application of the FMQA to IR from the WWW over those obtained by using Lycos™ alone, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

The analysis was then focused on providing evidence to accept or reject this hypothesis. As noted in chapter 8, to achieve this goal it was ascertained in an *Analysis Plan* (see Appendix F) that a number of sub-questions needed to be addressed. The chapter then concentrated on a presentation of the user study results in summary before presenting the analysis in full. The controlled and ad-hoc phase data were analysed separately. Mean scores from the user ratings of the results were calculated, presented, and interpreted in terms of the four sub-questions. McNemar's test (Reid & Moore, 1992) was performed on the textual feedback and Wilcoxon's test (Reid & Moore, 1992) on the paired user-scored data (ratings for the modified and unmodified results) to test the significance of any difference. Finally, the mean score data was stratified by a number of categories of the user study data, e.g. unmodified query, to examine if any difference between the paired data was repeated across numerous possible values of these categories.

The main conclusions of the analysis, as reported in chapter 8, were that the four sub-questions could be answered affirmatively, especially for the data from the controlled phase. In other words, there was sufficient evidence to reject the null hypothesis and thus the crux of the research was addressed with the answer that the application of the FMQA to IR from the WWW *could* improve upon the results returned by Lycos™, in the particular conditions stated in the research issue.

The next section, section 9.5, offers a case for generalising the research outcomes to other conditions. It also discusses possible future directions for this research.

9.5 Generalising and Extending the Research

An examination of the construction of the FMQA and the technique of FUM employed will show that this research outcome can be generalised in certain ways. The particular information domain chosen for the prototype, *AI*, has no special intrinsic significance with regard to the nature of this research. True, some of those users which took part in the user study had some knowledge of the AI domain and the extent of this knowledge stretched from those who considered themselves to have very little (mainly confined to the controlled group) to those who expressed considerable domain knowledge (mainly within the ad-hoc group). Recall that the focus of the prototype FMQA was on improving IR results for those with limited or intermediate domain knowledge.

The results reported in chapter 8 appear to support this focus though there was also some favourable responses from ad-hoc users whom were labelled by the prototype as being 'expert'. This evidence points towards a conclusion that the FMQA could be extended to cover other information domains, by the development of other query lists and associated semantically related lexicons, and if then focused on aiding users seeking information within those domains whom have little or average knowledge of those domains, it is likely that there would be a similar research outcome. A possible future development of the research would be to expand the scope of the FMQA to cover more information domains.

As stated in chapter 5, the shape of the fuzzy sets used in the models – the membership functions – was decided heuristically. The research outcome, that the FUM approach to IR from the WWW can provide improvement in the IR results, supposes a certain choice of membership function. One interesting future direction for the research would be to examine the effects of different membership functions on the performance of the FMQA. This development was considered in the present research but time and resource constraints prevented it from being performed.

One generality of the research, not particularly highlighted before, is that the approach is not confined to the use of one particular search engine. Lycos™ was only chosen because of its position of superiority at the time, but it would be very easy to develop implementations of the FMQA which would work with other search engine catalogues and, in fact, with other information corpuses. This is because the coupling between the FMQA and the search engine is what is termed *weak*, and there is nothing in the FUM approach that is dependent on the choice of information corpus. With hindsight, it would be better if the FMQA retained control over the presentation of the IR results to the user, and doing this in any future implementation would allow another possible area of research to be examined. This is the effects of filtering and manipulating the IR results before presentation and according to the information contained in the user models.

Other possible future directions for the research could be the development of more sophisticated, persistent user models that can be directly modified by the users. Additionally, the development of more comprehensive query refinement by allowing user-generated queries and by the research and construction of more semantically rich lexicons. The second development may perhaps be difficult to achieve, and would no doubt raise difficult and interesting questions about knowledge acquisition and representation. Questions such as the meaning of words and the semantic relations between them, and the representations of these relations. The thorny issue of which sources to use for knowledge acquisition would arise as well, raising questions such as which sources can be considered as the definitive expert sources for a particular information domain. However, this kind of research would allow the development of the FMQA to enable its use in more general situations. Therefore, it can be argued that the research outcome can be generalised to a certain extent in a number of ways. It is likely that the outcome would be the same if different information domains were used and if the queries had been applied to different search engines.

The development of more sophisticated query refinement, through the creation of semantically rich lexicons coupled with the use of user feedback on IR results to modify user models, might allow the FMQA to increase its ability to improve IR overall and for different types of user, including those considered to be expert. This is because it is hoped that its improving abilities would be honed by such developments.

The final section of this chapter and the thesis, section 9.6, presents some concluding remarks.

9.6 Final Thoughts

The work reported in this thesis was begun in August 1994 when the WWW was only 2-3 years old (Berners-Lee *et al.*, 1992) and could still be considered to be the preserve of the academic, the computing professional and the technologist. In the intervening five years, it has grown in size and importance at a prodigious rate, finding an ever-increasing use in schools, in the home, for leisure and work-related activities, and, most recently, in the area of commerce (Bowyer, 1997; Earthweb, 1998; Galla, 1996; Imdb, 1998; *Mecklermedia*, 1998; Teachers.net, 1998; *Tesco*, 1998). What was a fairly large corpus of information in August 1994, 500,000 documents, had, in May 1999, mushroomed spectacularly to a huge amount of something in the region of 350 Million (Altavista, 1999a). One of the interesting outcomes of the ad-hoc phase of the user study was that many of the respondents came from outside academia. Now more than ever, what we, the users of the WWW, need are tools which can intelligently assist our arduous task of finding the information we require in this mountain of documents.

This research has shown that it is possible to take established tools and techniques from the fields of AI and user modelling, apply them in a simple but straightforward manner to the problem of IR and WWW, and achieve an outcome which implies improvement in terms of the results in the opinion of those who matter, the users of the WWW. Whereas the specific application of the techniques may not, perhaps, be one which could be used to build a general all-purpose intelligent IR tool that can cover all information domains, it has been argued that there is scope for using and extending the FUM technique to cover other information domains and for the development of the technique itself through, for example, more sophisticated user models.

Research and development of possible tools to aid IR continues at a pace (Agentweb, 1998; Bishop *et al.*, 1997; CIIR, 1997; Fox *et al.*, 1996; Maes, 1994; Salampasis *et al.*, 1996). At a recent conference, John (1998) heard the majority of the panel of a plenary session (Fuzz-IEEE, 1998) state that an important future area of research for fuzzy logic is in the field of network IR. As already noted, the combination of fuzzy logic and IR research is not new (Radecki 1976; Yager & Larsen, 1993) but the application of such research to on-line IR is relatively recent (Chang & Chen, 1998). The research reported in this thesis shows one way forward for the use of fuzzy logic and user modelling in IR, and that, though there are problems and pitfalls to be overcome, there are gains to be achieved now and are bright possibilities to be sought in the future.

REFERENCES

Ackerman M *et al.* (1997) Learning Probabilistic User Profiles *AI Magazine* Summer, pp. 47-56.

Adams R J (1990) *Communication and Delivery systems for Librarians* Gower, Aldershot.

Agentlist (1997) 'The Software Agents Mailing List' at http://www.cs.umbc.edu/agentslist/

Agentweb (1998) 'UMBC Agentweb' available at http://www.cs.umbc.edu/agents.

Allen R B (1990) User Models: Theory, Method and Practice *International Journal of Man-machine Studies* **32**, pp. 511-543.

Alta Vista (1999a) 'Main page', available at http://altavista.digital.com

Alta Vista (1999b) 'What is Refine?, available at http://altavista.digital.com/av/content/help-refine.htm.

Anick P G, Vaithyanathan S (1997) 'Exploiting clustering and phrases for context-based information retrieval' **in** *Proceedings of ACM SIGIR'97 International Conference on Research and Development in Information Retrieval* pp. 314-323.

Ariadne (1998) 'Ariadne Project Description', available at http://ariadne.uni.ch/project.htm.

Archie (1995) The archie server at McGill University can be accessed by telneting to: archie@archie.mcgill.ca:23.

Arens, Y, Knoblock, C A and Shen, W-M (1996) Query reformulation for dynamic information integration *Journal of Intelligent Information Systems* **6** (2/3), June, pp. 99-130.

Armstrong *et al.* (1995) 'Web Watcher: A Learning Apprentice for the World Wide Web' **in** *Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogenous Distributed Resources* Stanford, CA, March.

Arnold W E, McCroskey J C, Prichard S V O (1997) 'The Intensity Component of Semantic Differential Scores for Measuring Attitude', available at http://www.as.wvu.edu/~jmccrosk/48.htm.

Atkinson R D, Yokley J R (1993) Multiplatform CD-ROM networking *CD-ROM Professional* **6** (3), pp. 73-81.

Baaklini M, Mamdani (1975) Perspective Methods for Deriving Control Policy in a Fuzzy-Logic Controller *Electronic Letters* **11**, pp. 625-626.

Bailey J E, Pearson S W (1983) Devlopment of a Tool for Measuring and Analysing Computer User Satisfaction *Management Science* **29** (5), pp. 530-545.

Balabanovic M, Shoham Y (1997) Fab: Content-based, collaborative recommendation *Communications of the ACM* **40** (3), pp. 66-72..

Bansler J *et al.* (1985) 'Filing and Retrieval of Unstructured Information: Some System Considerations' **in** *ESPRIT '84: Status Report of On-going Work* J Roukens & J F Renuart (eds) Elsevier, Amsterdam, pp. 295-319.

Barcaroli B *et al* (1996) 'Knowledge-based Access to the Network', available at http://www.info.unicaen.fr/~serge/3wia/workshop/papers/paper6.html.

Basch R (1995) The marriage of fax and on-line *ONLINE* **19** (1), January/February, pp. 12-21.

Bassano J C, Braunwarth M, Mekaouche W (1992) Dialect 2: A Multi-experts system for information retrieval **5** (1), pp. 43-60.

Bayardo R *et al.* (1997) '*InfoSleuth*: Semantic Integration of Information in Open and Dynamic Environments', available at http://www.mcc.com/projects/infosleuth/ papers/sigmod97.fm.html.

Beaumont I (1996) 'User Modelling and Hypertext Adaptation in the Tutoring System Anatom-Tutor' at http://www.education.uts.edu.au/projects/at/beaumont.html.

Beilstein (1999) 'Beilstein Online' available at: http://www.beilstein.com/beilst_2.html.

Belew R K (1989) 'Adaptive Information Retrieval' **in** *Proceedings of the 12th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval* Cambridge, MA, June 25-28, pp. 11-20.

Bell, D.A. (1992) Framework for Query Optimization in Distributed Statistical Databases *Information and Software Technology* **34** (6), pp. 363-378.

Bell G (1996) 'HyBis - An Adaptive Navigation System for Hypermedia' at http://www.psyc.nott.ac.uk/~grb/papers/HCI96-doctoral-position.html.

Belkin N J, Seeger T, Wersig G (1984) Simulation of a distributed expert-based information provision mechanism **3** pp. 122-141.

Belkin, N J and Croft W B (1992) Information Filtering and Information Retrieval: Two Sides of the Same Coin? *Communications of the ACM* **35** (12), pp. 29-38.

Belkin, N J, Marchetti P G and Cool, C.(1993) BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval *Information Processing & Management* **29** (3), pp. 325-344.

Belkin N J *et al.* (1993) 'The Effect of Multiple Query Representations on Information Retrieval System Performance' **in** *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 339-346.

Bellman R E, Zadeh L A (1970) Decision Making in a Fuzzy Environment *Mathematical Analysis and Applications* **17**, pp. 141-164.

Berners-Lee T (1995a) 'An Executive Summary of the World-Wide Web Initiative' at http://www.w3.org/hypertext/WWW/Summary.html.

Berners-Lee T *et al.* (1992) World-Wide Web: The Information Universe *Electronic Networking: Research, Applications and Policy* **1** (2), Spring, pp. 52-58.

Bishop A P *et al.* (1997) 'Building a University Digital Library: The Need for a User-Centered Approach', available at http://anshar.grainger.uiuc.edu/dlisoc/ monterey.final_copy.html

Bjorner S (1995) Viewing, downloading and printing images - part 2: using Dialog link *ONLINE* **19** (1), January/February, pp. 41-45.

Blair D C, Maron M E (1985) 'An Evaluation of Retrieval Effectiveness for a Full-text Document Retrieval System' *Communications of the ACM* **28** (3), pp. 289-299.

BLCMP (1998) 'Talis Web Product Overview', available at http://www.blcmp.org.uk/products/talisweb.htm.

Bonham M D, Nelson L L (1988) An Evaluation of Four End-User Systems for Searching MEDLINE *Bulletin of the Medical Library Association* **76** (1), January, pp 22-31.

Borgman C L (1986) The User's Mental Model of an IR System: An Experiment on a Prototype On-line Catalog *International Journal of Man-machine Studies* **24**, pp. 47-64.

Botman A M, Kok A J, Siklossy L (1987) 'Using User Models and Automatic Expertise Acquisition in Man-Machine Interaction' **in** *Proceedings 7th International Workshop on Expert Systems and their Applications*, Avignon, pp. 389-398.

Bourne C P (1980) On-line systems: history, technology and economics *Journal of the American Society for Information Science* **31** (3), May, pp. 155-160.

Boyle C, Encarnacion A (1991) 'A User Model Based Hypertext Documentation System' **in** *Proceedings of the IJCAI Workshop* W4.

Boyle C, Encarnacion A (1993) 'Meta-Doc: An Adaptive Hypertext Reading System' **in** *User Modeling and User-adapted Interaction.*

Bowyer R (1997) 'WebTV Internet Access-A Quick Review and Other TV/Internet Devices', available at http://www.biug.org/notes/webtvby.html.

Brachman R J *et al.* (1991) 'Living with CLASSIC: When and How to Use a KL-ONE-Like Language' **in** *Principles of Semantic Networks: Explorations in the Representations of Knowledge* John Sowa (ed.), Morgan-Kaufmann, San Mateo, CA, pp. 401-456.

Brae M, Rutherford D A (1979) Theoretical and Linguistic Aspects of the Fuzzy Logic Controller *Automat.* **15** (5), pp. 553-577.

Brajnik G, Guida G, Tasso C (1987) User modelling in intelligent information retrieval *Information Processing and Management* **23**, pp. 305-320.

Brajnik G, Guida G, Tasso C (1988) 'IR-NLI II: Applying Man-machine Interaction and Artificial Intelligence Concepts to Information Retrieval' **in** *Proceedings 11th International Conference on Research and Development in Information Retrieval* Grenoble, pp. 387-400.

Brajnik G, Mizzaro S, Tasso C (1996) Evaluating User Interfaces to Information Retrieval Systems: A Case Study on User Support *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 128-136.

Braun E (1994) *The Internet Directory* Fawcett Columbine/Ballantine, NY.

Brimicombe A (1997) 'A Fuzzy Set Universal Translator for Linguistic Hedges Used in Geographical Enquiry', 10[th] European Colloquium on Theoretical and Quantitative Geography, available at http://esmt1.tuwien.ac.at/EUCOL/c-3-6.htm.

Broglio J, Callan J P, W B Croft (1996) 'An Overview of the INQUERY System as Used for the TIPSTER Project', available at http://ciir.cs.umass.edu/info/inquery.html.

Brook Long K, Fowler J, Barber S (1992) 'VNS Retriever: Querying MEDLINE over the Internet' in *Proceedings of Summer 1992 USENIX Conference*, June 8-12, San Antonio, TX, pp. 81-91.

Brown M (1996) 'Fuzzy Sets and Systems' at http://www.abo.fi/~rfuller/fuzs.html .

Brown J (1995) 'Preliminary WWW OPAC', Pilkington Library, University of Loughborough at: http://liba.lut.ac.uk/main.html.

Buchanan B G, Shortliffe E H (eds) (1984) *Rule-Based Expert Systems* Addison-Wesley, Reading, MA.

Bush V (1945) 'As we may think' *The Atlantic Monthly* July.

Callan J P, Croft W B (1993) 'An Evaluation of the Query Processing Strategies Using the TIPSTER Collection' in *Proceedings of the 16[th] Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 347-355.

Catledge L D, Pitkow J E (1995) Characterising Browsing Strategies in the WWW *Computer Networks and ISDN Systems* **27** pp.1065-1073.

Cavanagh J M A (1989) Library applications of knowledge-based systems *Reference Librarian* **23**, pp. 1-19.

CCITT/ISO (1988) *The Directory, Part 1: Overview of Concepts, Models and Services* CCITT Draft Recommendations X.500/ISO DIS 9594-1, CCITT/ISO, Gloucester, UK, December.

Celentano A, Fugini M A, Pozzi S (1995) Knowledge-based document retrieval in office environments: The Kabiria system *ACM Transactions on Information Systems* **13** (3), pp. 237-268.

Chang C-S, Chen A L P (1998) Supporting conceptual and neighborhood queries on the World Wide Web *IEEE Transactions on Systems, Man and Cyberbetics-Part C: Applications and Reviews* **28**, May, pp. 300-308.

Chang S-K, Costabile M F, Levialdi S (1993) Modeling users in an adaptive visual interface for database systems *Journal of Visual Languages and Computing* **4**, pp. 143-159.

Chen J E, Otto K N (1995) Constructing membership functions using interpolation and measurement theory *Fuzzy Set and Systems* **73**, pp. 313-327.

Chen H, Dhar V (1990) User misconceptions of information retrieval systems *International Journal of Man-machine Studies* **32**, pp 673-692.

Cheney D (1995) 'WAISgateII (WWW to WAIS gateway)' at:
http://www.wais.com/waisgate2.html/.

CIIR (1997) Center for Intelligent Information Retrieval, available at http://ciir.cs.umass.edu/info/ciirbiblio.html.

Clark S H, Mischo W H (1992) On-line public access catalog retrieval structures and techniques: with references to recent developments in the United States and Great Britain *IUATL Proceedings [New Series]* **1** pp. 113-120.

CMU (1996) 'CMU Artificial Intelligence Repository - Fuzzy Logic Packages' at http://www-cgi.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/fuzzy/systems/ .

Cockburn A, Jones S (1996) Which way now? Analysing and Easing Inadequacies in WWW Navigation *International Journal of Human-Computer Studies* **45** (1), pp.105-130.

Cohen P R, Kjeldsen R (1987) Information retrieval by constrained spreading activation in semantic networks *Information Processing and Management* **23** (4), pp. 255-268.

Collier M W (1993) 'The Future of Academic Library Collections in the Electronic Environment' **in** *Proceedings of 'Taming the Electronic Jungle' Seminar, May* National Acquisitions Group/UK Serial Group, pp. 106-112.

Collier M W (1995) 'Defining the Electronic Library' **in** *Proceedings of 1st ELVIRA (Electronic Library and Visual Information Research) Conference* M Collier & K Arnold (eds) Milton Keynes, May 1994, Aslib, London, pp. 1-5.

Collier M W, Ramsden A, Wu Z (1993) 'The Electronic Library: Virtually a Reality' **in** *Opportunity 2000: Understanding and Serving Users in an Electronic Library, Proceedings 15th International Essen Symposium*, Universitatsbibliothek, Essen, 12-15 October 1992, pp. 135-146.

Coutaz J *et al.* (1992) Critical issues: User Modelling *Engineering for Human-Computer Interaction (IFIP Transactions A - Computer Science & Technology)* **A-18**, pp. 419-423.

Cox E (1994) *The Fuzzy Systems Handbook - A Practitioner's Guide to Building, Using and Maintaining Fuzzy Systems* Academic Press, New York.

Crane P (1992) 'Multifunction Optical Offers Versatility - New Standards Will Combine Rewritable and WORM for Data Storage' **in** *The Best of CD-ROM Professional: Timeless Tips & Practical Advice* (ed. N Herther), Online Inc., Wilton, CT, pp. 144-146.

Cullen C (1997) Filtering information from the Internet: the Borges information filtering service project in Dublin City University *Vine* **104**, pp. 45-50.

Cusack, E (1991) Refinement, Comformance and Inheritance *Formal Aspects of Computing* **3** (2), pp. 129-141.

Dalrymple P W (1990) Retrieval by Reformulation in Two Library Catalogs: Towards a Cognitive Model of Searching Behaviour *Journal of the American Society for Information Science* **41** (4), pp. 272-281.

Dalrymple P W, Zweizig D L (1992) Users' Experience of Information Retrieval Systems: An Exploration of the Relationship between Search Experience and Affective Measure *Library and Information Science Research* **14**, pp. 167-181.

Danilowicz, C., 1994, Modelling of User Preferences and Needs in Boolean Retrieval Systems *Information Processing & Management* **30** (3), pp. 363-378.

Dern D (1995) Come shop with me *.net* issue 4, March, p. 30.

DIALOG (1998) 'The Dialog Corporation – Keeping the World Better Informed', available at http://www.dialog.com.

Digital Library Group – DMU (1999) 'The De Montfort University Digital Library System', available at: http://www.dlib.dmu.ac.uk/About/.

Doszkocs T E (1991) 'Library Applications of Neural Networks' **in** *Proceedings 6th Annual Conference on Computers in Libraries* Mecklermedia, Westport, CT, pp. 44-48.

Doszkocs T E, Reggia J, Lin X (1990) Connectionist models and information retrieval *ARIST (Annual Review of Information Science and Technology)* **25**, pp. 209-260.

DMU (1997) 'Research Projects within the Division of Learning Development', available at http://www.dmu.ac.uk/dept/admin/dld/research.html.

DMU (1998a) 'The DMU Library OPAC', available at http://www.library.dmu.ac.uk (via telnet).

DMU (1998b) *The DMU CD-ROM Network* De Montfort University Kimberlin Library, The Gateway, Leicester, LE1 9BH, UK.

Dubois D, Prade H (1980) *Fuzzy Sets and Systems: Theory and Applications* Academic Press, NY.

Dubois D *et al.* (1997) 'Introduction - A Manifesto: Fuzzy Information Engineering' **in** *Fuzzy Information Engineering* eds D Dubois & H Prade, pp. 1-8.

Duval E (1998) 'HOME – Hypermedia Object Management Environment', available at http://www.cs.kuleuven.ac.be/~erikd/MANUAL/home.html.

Earthweb (1998) 'The Intranet Journal', available at http://www.intranetjournal.com.

Efthimiadis E N (1993) 'A User-Centred Evaluation of Ranking Algorithms for Interactive Query Expansion' **in** *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 146-159.

Eichmann D (1996) 'A User Agent for the Web', available at http://ricis.cl.uh.edu/agents/user_agent_overview.html.

Eichmann D, Wu J (1996) 'Sulla – a user agent for the Web', available at http://mingo.info-science.uiowa.edu/eichmann/www-s96/Overview.html.

elib (1998) 'elib: Electronic Library Programme', available at http://www.ukoln.ac.uk/services/elib.

Emtage A, Deutsch P (1992) 'Archie: An Electronic Directory Service for the Internet' in *Proceedings of the 1992 Winter USENIX Conference*, USENIX, Sunset Beach, CA, pp. 93-110.

Ennis M, Sutcliffe A (1997) 'Cognitive Models in IR: Predicting Task Support for Users' Needs', available at http://www.psyc.bbk.ac.uk/research/projects/cogent/wkshp1/me.html.

ERCOMS (1997) 'ERCOMS project' available at http://www.dmu.ac.uk/dept/admin/dld/research.html.

Etzioni O, Weld D (1994) A Softbot-Based Interface to the Internet *Communications of the ACM* **37**(7), July, pp. 72-76.

Etzioni O *et al.* (1997) 'A scalable comparison-shopping agent for the world-wide web', available at http://www.cs.washington.edu/homes/etzioni/.

Euzenat J (1996) 'Knowledge Bases as Web Page Backbones', available at http://www.info.unicaen.fr/~serge/3wia/workshop/papers/paper10.html.

ExamNet (1999) 'ExamNet', available at http://www.dlib.dmu.ac.uk/Services/examnet.shtml.

Excite (1997) 'Excite – home page', available at http://www.excite.com.

Eyre J (1997) 'ELISE – The Introduction' Electronic Library Image Service for Europe, available at http://severn.dmu.ac.uk/elise/.

Feigenbaum E A (1977) 'The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering' in *Proceedings IJCAI-77*, pp. 1014-1029.

Feigenbaum E A, McCorduck P *The Fifth Generation* Michael Joseph, London, p. 88.

Fenichel C H (1981) Online Searching: Measures that Discriminate Among Users with Different Types of Experience *Journal of the American Society of Information Systems* **32** pp. 23-32.

Ferry A (1997) 'ADAM (Art, Design, Architecture and Media) – 1996 Survey of User Information Needs and Search Methods Results', available at http://adam.ac.uk/adam/reports/survey/intro/html.

Fielding R T (1994) 'Maintaining Distributed Hypertext Infostructures - Welcome to MOMspiders Web' in *Proceedings 1st International W3 Conference (WWW94), Geneva, May 25-27* (to be published by Elsevier).

FIGIT (1994) '"FIGIT Framework" – Follet Implementation Group on Information Technology, Framework for Progressing the Initiative', available at http://www.jisc.ac.uk/pub/c4-94.html.

Finin T *et al.* (1995) 'KQML 95' available at http://www.cs.umbc.edu/kqml/kqml95/kqml95.html.

Follett *et al.* (1993) 'The Follet Report – Joint Funding Council's Libraries Review Group Report', available at http://www.ukoln.ac.uk/services/papers/follett/report/.

Fourcher A (1997) 'The Videoconference Resource', available at http://www.videoconferencing.com/.

Fox E A (1987) Development of the CODER system: a testbed for artificial intelligence methods in information retrieval *Information Processing and Management* **23** (4), pp. 341-366.

Fox E A (1997) 'Networked Digital Library of Theses and Dissertations', available at http://www.dlib.org/september97/theses/09fox.html.

Frei H P, Stieger D (1995) The Use of Semantic Links in Hypertext Information Retrieval *Information Processing and Management* **31**(1), pp. 1-14.

Frolich P, Nejdl W, Wolpers M (1999) 'KBS-HYPERBOOK – An Open Hyperbook System for Education' 8th WWW Conference, Toronto, Canada, May 11-14, available at: http://www8.org.

Fu L-M (1994) *Neural Networks in Computer Intelligence* McGraw-Hill, New York, 460pp.

Fung R, Del Favero B (1995) Applying Bayesian networks to information retrieval *Communications of the ACM* **38** (5), pp. 42-48.

Fuzz-IEEE (1998) 'Plenary Session Remarks' **in** *Proceedings of the 7th International Conference on Fuzzy Systems FUZZ-IEEE'98.*

GALE (1998) 'The Gale Group – Reference Information for Libraries, Schools and Businesses', available at http://www.gale.com.

GASS (1999) 'GASS - Good Alternative Simple Solution', available at http://www.dlib.dmu.ac.uk/~gass/.

Gallant S I *et al.*(1992) HNC's *Matchplus* system *SIGIR Forum* **26** (2), pp. 34-38.

Galliers J R (1992) 'Autonomous Belief Revision and Communication' **in** Gardenfors P (ed.) *Belief Revision* Cambridge University Press, Cambridge, pp. 220-246.

Gamiel K (1995) 'CNDIR Research and Development', WWW/Z39.50 Integration, at: http://vinca.cndir.org.

GANNET (1993) 'GANNET for IR: An Integrated Framework', available at http://ai.bpa.arizona.edu/papers/gannet93/.

Gauch S (1990) 'Evaluation of an Expert System for Searching in Full Text' **in** *Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 255-277.

Gauch S, Smith J B (1991) Search Improvement via Automatic Query Reformulation *ACM Transactions on Information Systems* **9** pp. 249-280.

Gauch S, Wang G, Gomez M (1998) 'ProFusion: Intelligent Fusion from Multiple, Distributed Search Engines', available at http://medoc.springer.de:8000/jucs_2_9/profusion_intelligent_fusion_from/html/paper.html.

Genesereth M R, Fikes R E (1994) 'Knowledge Interchange Format – V 3.0, Reference Manual' available at http://logic.standford.edu/kif/Hypertext/kif-manual.html.

Genesereth M R, Ketchpel S P (1994) Software Agents *Communications of the ACM* **37**(7), July, pp. 48-54.

Gilbert D, Janca P (1996) 'IBM Intelligent Agents', IBM Corporation, Research Triangle Park, NC, available at http://www.raleigh.ibm.com/ iag.iaghome.html.

Gilder G (1994) 'The Bandwidth Tidalwave' *Forbes ASAP*, December, available at http://pdomain.uwindsor.ca/faculty/socsci/comstudies/Costclarsa/tidal.html.

Goldberg D *et al.* (1992) Using collaborative filtering to weave an information tapestry *Communications of the ACM* **35** (12), pp. 61-70..

Good I J (1961) A casual calculus (I) *British Journal of Philosophy and Science* **11**, pp. 305-318.

Gopher (1995) More infomation about Gopher services and an extensive guide to Gopher can be found through ftp to: boombox.micro.umin.edu in the ' pub/gopher ' and ' pub/gopher/docs ' directories

Gordon M (1988) Probabilistic and genetic algorithms *Communications of the ACM* **31** (10), pp. 1208-1218.

Govert N, Pfeifer U (1997) 'SFgate – The WWW Gateway for freeWAIS-sf', available at http://amaunet.cs.uni-dortmund.de/projects/ir/SFgate/SFgate.html.

Gralla P (1996) 'How Intranets Work', 2D Press, available at http://www.developer.com/reference/library/IS62764411/ewtoc.html.

Greef de P, Breuker J, Jong de T (1988) *Modality: An Analysis of Functions, User Control and Communication in Knowledge-based Systems* ESPRIT Project P1098, Deliverable D6.

Guan T, Wong K F (1999) 'KPS – A Web Information Mining Algorithm' 8th WWW Conference, Toronto, Canada, May 11-14, available at: http://www8.org.

Guilfoyle C (1997) 'Intelligent Agents and Information Retrieval' in *Information Interchange Report*, available at http://www.techapps.co.uk/iiarttagt.html.

Haines D, Croft W B (1993) 'Relevance Feedback and Inference Networks' in *Proceedings of the 1993 ACM-SIGIR Conference* pp. 2-11.

Hall J H (1986) *Online Bibliographic Databases* 4th edn., Aslib, London.

Han, J (1994) Constraint-Based Query Evaluation in Deductive Databases *IEEE Transactions on Knowledge and Data Engineering* **6** (1), pp. 96-107.

Harman D (1993) 'Overview of the First TREC Conference' in *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 36-47.

Harman D (1994) 'Overview of the Third TREC Conference' in *Proceedings of the 3rd Text Retrieval Conference* Association for Computing Machinery, Gaithersburg, MD, November.

Hartley J, Noonan A, Metcalfe S (1987) *New Electronic Information Services: An Overview of the UK Database Industry in an International Context* Gower, Aldershot.

Harrenstein K, Stahl M, Feinler E (1985) *NICName/WHOIS* RFC 954, SRI International, October.

Haykin S S (1994) *Neural Networks: A Comprehensive Foundation* Macmillan, NY, 696pp.

HCI (1998) 'Human Compuer Archive (User Modelling)', at http://lpac.ac.uk/SE2-HPC/Articles/GenerateHtml/hci.model.html.

Hebb D O (1949) *The Organisation of Behaviour* Wiley, New York.

Heery R (1996) 'Review of metadata formats *Program* **30** (4), October.

Heckerman D, Mamdani A, Wellman M P (1995) Real-world applications of Bayesian networks *Communications of the ACM* **28** (3), pp. 24-26.

Hellendoorn H, Thomas C (1993) Deffuzification in Fuzzy Controllers *Intelligent and Fuzzy Systems* **1** pp. 109-123.

Heller S (ed.) (1990) *The Beilstein Online Database* ACS Symposium Series 436, American Chemical Society, Washington, DC.

Herther N (1985) CD-ROM technology: a new era for information storage and retrieval *Online* **9** (6), November, pp. 17-28.

Hosono, K T. *et al.* (1985) Weighted document retrieval systems based on fuzzy set theory approaches, *Library and Information Science* pp.137-147.

Hopfield J J (1982) 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities' **in** *Proceedings of the National Academy of Science* **79**, pp. 2554-2558.

Hoppe T (1996) 'The MIHMA Project: a Web Information Service Based on Description Logics', available at http://www.info.unicaen.fr/~serge/3wia/workshop/papers/paper25.html.

Howard R A, Matheson J E (1981)' Influence Diagrams' **in** *Readings on the Principles and Applications of Decision Analysis* R A Howard, J E Matheson (eds), Strategic Decisions Group, Menlo Park, CA, Vol. 2, pp. 721-762.

Huibers T, van Linder B (1996) 'Formalising Intelligent Information Retrieval Agents' at http://www.dis.strath.ac.uk/people/ruairi/theo.html.

Hull D (1993) 'Using Statistical Testing in the Evaluation of Retrieval Experiments' **in** *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 329-338.

Humpherys B L, Lindberg D A (1989) 'Building the Unified Medical Language System' **in** *Proceedings of the 13th Annual Symposium on Computer Applications in Medical Care* November 5-8, IEEE Computer Society Press, Washington, DC.

Hsu C-N, Knockblock C A (1995) 'Using Inductive Learning to Generate Rules for Semantic Query Optimization' **in** *Advances in Knowledge Discovery and Data Mining* G Piatetsky, U Fayyad (eds), chapter 17, MIT Press, Boston, MA.

Imdb (1998) 'The Internet Movie Database: A database of more than 160,000 titles', available at http://www.uk.imdb.com.

Ingwersen P (1993) Polyrepresentation of Information Needs and Semantic Entities: Elements of a Cognitive Theory for Information Retrieval Interaction **in** *Information Retrieval Interaction,* Taylor Graham, London, chapter 5, pp. 83-122.

INSPEC (1998) *INSPEC Database On CD-ROM* Silverplatter Databases, 10 Barley Mow Passage, Chiswick, London W4 9PU, UK.
Information at http://www.silverplatter.com/catalog/insp.htm.

Iwadera T, Kimoto H (1992) 'The Effect of a Dynamic Word Network on Information Retrieval' **in** *Proceedings of the SPIE Conference (SPIE Volume 1710 Science of Artificial Neural Networks)* **2** (1), pp. 622-630.

Jennings A, Higuchi H, Liu H (1993) A user model neural network for a personal news service *Australian Telecommunication Research* **27** (1), pp. 1-12.

John R I (1998) 'Private communication with the author'.

Jones C H (1998) 'Extranets', available at http://www.uis.edu/~chjones/extranet.htm.

Jones, J (1991) User modelling and advice giving in intelligent help systems for Unix *Information and Software Technology* **33** (2), pp. 121-133.

Jones E K, Roydhouse A (1995) Intelligent Retrieval of Archived Meteorological Data *IEEE Expert* December, pp. 50-57.

Kahle B *et al.* (1993) Interfaces for distributed systems of information servers *Journal of the American Society for Information Science* **44** (8), pp. 453-467.

Kahn R E, Cerf V G (1988) 'An Open Architecture for a Digital Library System and a Plan for its Development' Corporation for National Research Initiatives Technical Report, March.

Kamba T, Sakagami H, Koseki Y (1997) ANATAGNOMY: A Personalized Newspaper on the World Wide Web *International Journal of Human-Computer Studies* **46** pp. 789-803.

Kantrowitz M *et al.* (1997) 'Fuzzy Logic FAQ', available at http://www.cs.cmu.edu/Groups/AI/html/faqs/ai/fuzzy/part1/faq-doc-0.html.

Kaplan C, Fenwick J, Chen J (1993) Adaptive hypertext navigation based on user goals and context *User Modelling and User-Adapted Interaction* **3** pp. 193-220.

Kass R (1989) 'Student Modelling in Intelligent Tutoring Systems - Implications for User Modelling' **in** *User Models in Dialog Systems* Kobsa A, Wahlster W (eds) Springer, Berlin, pp. 386-410.

Kass R, Finin T (1988) Modelling the user in natural language systems *Computational Linguistics* **14** (3), pp. 5-22.

Keates S, Robinson P (1997) 'Human Factors and User Modelling for Motion-Impaired Users' at http://rehab-www.eng.cam.ac.uk/projects/lsk12/usermodel.hmtl.

Kerr M (1997) NewsAgent for Libraries *VINE* **104** pp. 51-53.

King P J, Mamdani E H (1977) The Application of Fuzzy Control Systems to Industrial Processes *Automat.* **13** (3), pp.235-242.

Kirk T (1996) 'Knowledge Based Access to Information on the World Wide Web', available at http://www.info.unicaen.fr/~serge/3wia/workshop/papers/paper20.html.

Klir G J, Folger T (1988) *Fuzzy Sets, Uncertainty and Information* Prentice-Hall, Englewood Cliffs, NJ, 355pp.

Kobsa A (1990) Modelling the User's Conceptual Knowledge in BGP-MS – A User Modelling Shell System *Computational Intelligence* **6** (4), pp. 193-208.

Kobsa A, Wahlster W (eds) (1989) *User Models in Dialog Systems* Springer, Berlin, 472 pp.

Kok A J (1991a) A formal approach to user models in data retrieval *International Journal of Man-machine Studies* **35**, pp. 675-693.

Kok A J (1991b) A review and synthesis of user modelling in intelligent systems *The Knowledge Engineering Review* **6** (1), pp. 21-47.

Kolodner J L (1993) *Case-based Reasoning* Morgan Kaufmann, San Francisco, CA.

Konstan J A *et al.* (1997) GroupLens: Applying collaborative filtering to Usenet news *Communications of the ACM* **40** (3), pp. 77-87.

Koo T-K J (1996) 'Construction of Fuzzy Linguistic Model', **in** Proceedings of IEEE Conference on Control and Decision, Kobe, Japan, available at http://robotics.eecs.berkeley.edu/~koo/fz_fnapp_cdc96.html.

Kosko B (1992) *Neural networks and fuzzy systems : a dynamical systems approach to machine* Prentice Hall, Englewood Ciffs, NJ.

Koster M (1995) 'World Wide Web Robots, Wanderers and Spiders' at http://web.nexor.co.uk/mak/doc/robots/robots.html/.

Kroeker K L (1997) Emedia News *E-Media Professional,* April, Cambridge, MA.

Krulwich B, Burkey C (1996) 'Learning User Information Interests through Extraction of Semantically Significant Phrases' **in** *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access* Stanford, CA.

Kwok K L (1991) 'Query Learning Using an ANN with Adaptive Architecture' **in** *Proceedings 8th International Workshop on Machine Learning* Morgan-Kaufmann, CA, pp. 260-264.

Kwok, K.L.(1995) A Network Approach to Probabilistic Information Retrieval *ACM Transactions on Information Systems* **13** (3), pp. 324-353.

Laffan S (1996) 'Rapid Appraisal of Groundwater Discharge Using Fuzzy Logic and Topography', at http://ncgia.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/laffan_shawn/s_fetxt4.html

Lahiri S (1994) *Winsock FTP for Windows, Version 3.1* Released into the public domain January 11 1994, information from ' slahiri@magnus.acs.ohio-state.edu '.

Lai J W (1993) 'User Modelling in AI and HCI' at http://www.io.com/~jwtlai/usermodel.html.

Laine-Cruzel, S., Lafouge, T. and Abdallah, N. B., 1996, Improving information retrieval by combining user profile and document segmentation *International Journal for Artificial Intelligence in Engineering* **32** (3), pp. 305-316.

Lane-Lawley E, Summerhill C (1994) *Internet Primer for Information Professionals: A Basic Guide to Internet Networking Technology* Mecklermedia, Westport, CT.

Lang K (1995) 'Newsweeder: Learning to Filter Netnews' **in** *Proceedings of the 12th International Conference on Machine Learning* Tahoe City, CA.

Langton C G (ed.) (1995) *Artificial Life: An Overview* MIT Press, Cambrdige, MA.

Lee C C (1990) Fuzzy Logic in Control Systems: Fuzzy Logic Controller – Part 1 *IEEE Transactions on Systems, Man and Cybernetics* **20** (2), March/April, pp. 415-418.

Leventhal L M *et al.* (1993) Sleuthing in HyperHolmes: an evaluation of using hypertext vs a book to answer questions *Behaviour and Information Technology* **12** (3), pp. 149-164.

Lieberman H (1996) 'An Automated Channel Surfing Interface Agent for the Web', avaiable at http://www.info.unicaen.fr/~serge/3wia/workshop/papers/paper29.html.

Liehnert W, Sundheim B (1991) A Performance Evaluation of Text Analysis Technologies *AI Magazine* **1**, pp. 81-94.

Lindberg D A, Humpherys B L (1990) 'The UMLS Knowledge Sources: Tools for Building Better User Interface' **in** *Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care* November 4-7, IEEE, Los Alamos, CA, DC.

Litteck, H. J and Wallis, P.J.L., 1992, Refinement Methods and Refinement Calculii *Software Engineering Journal* **7** (3), pp. 219-229.

Lu Z, Callan J P, Croft W B (1996) Applying inference networks to multiple collection searching *Umass Technical Report/TR96-42* University of Massachusetts, Amherst, MA.

Lucarella D, Zanzi Z (1993) Information retrieval from hypertext: an approach using plausible inference *Information Processing and Management* **29** (3), pp. 299-312.

Lund University (1997) 'Lund University Electronic Library Homepage' , available at http://www.ubl.lu.se/index_eng.html.

Lycos (1998) 'Lycos – Your personal Internet guide', available at http://www.lycos.com/.

Machovec G S (1993) Veronica: A Gopher Navigational tool on the Internet *Online Libraries and Microcomputers* **11** (10), October, pp. 1-4.

Maes P (1994a) Agents that reduce work and information overload *Communications of the ACM* **37** (7), pp. 31-40.

Maes P (1994b) 'Social Interface Agents: Acquiring Competence by Learning from Users and Other Agents' **in** *Software Agents - Papers from the 1994 Spring Symposium (Technical Report SS 94-03)* O Etzioni (ed.), AAAI press, New York, pp. 71-78.

Maes P (1995) 'Artificial Life Meets Entertainment: Lifelike Autonomous Agents', available at http//pattie.www.media.mit.edu.

Mansfield Jr, W.H. and Fleischman, R.M., 1993, A High-Performance, Ad-Hoc, Fuzzy Query Processing System *Journal of Intelligent Information Systems* **2** (4), November, pp. 397-420.

Manoff M *et al.* (1993) The MIT Libraries Electronic Journals Project: Reports on Patron Access and Technical Processing *MIT Serials Review*, Fall, pp. 15-40.

Martin B K, Rada R (1987) Building a relational database for a physician document index *Medical Informatics* **12** (3), pp. 187-201, July-September.

Masuda Y (1982) Three great social revolutions: agricultural, industrial and informational *Prometheus* **3**, pp. 269-276.

MATLAB (1998) 'The Maths Works Inc – Developers of MATLAB and Simulink', available at http://www.mathworks.com/index.shtml.

Mauldin M L (1991) 'Retrieval Performance in FERRET: A Conceptual Information Retrieval System' **in** *Proceedings of the 14th International Conference on Research and Development in Information Retrieval*, Chicago, October, ACM SIGIR, available at http://fuzine.mt.cs.cmu.edu/mlm/sigir91.

Mauldin M L (1996) 'Lycos: The Catalog of the Internet' at http://www.lycos.com/

Mauldin M L (1997) 'Lycos: Design choices in an Internet search service', *IEEE Expert*, available at http://www.fuzine.com/mim/ieee97.html.

Maybury M T (1995) 'Intelligent Multimedia Interfaces – A Tutorial', at http://www.mitre.org/resources/centers/advanced_info/maybury/Tutorial.Part1.intro/TOC.html

McCahill M (1992) The Internet Gopher protocol: A distributed server information system *Connexions - The Interoperability Report* **6** (7), July, pp. 10-14.

McCray A T, Hole W T (1990) 'The Scope and Structure of the First Version of the UMLS Semantic Network' *Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care* November 4-7, IEEE Los Alamos, CA.

McCroskey J C, Prichard S V O, Arnold W E (1997) 'Attitude Intensity and the Neutral Point on Semantic Differential Scales', available at http://www.as.wvu.edu/~jmccrosk/31.htm.

McCulloch W S, Pitts W (1943) A logical calculus of ideas imminent in nervous activity *Bulletin of Mathematical Biophysics* **5**, pp. 115-133.

McNeill F M, O'Hagan M (1997) 'Fuzzy Systems Engineering', available at http://www.fuzzysys.com.

McTear M F (1993) User modelling for adaptive computer systems: a survey of recent developments *Artificial Intelligence Review* **7**, pp. 157-184.

Mecklermedia (1998) 'The Electronic Commerce Guide', available at http://e-comm.internet.com/.

MEDLINE (1998) 'MEDLINE resource centre', available at http://omni.ac.uk/general-info/internet_medline.html.

Mehling R (1959) A simple test for measuring intensity of attitudes *Public Opinion Quarterly* **23**, pp. 576-578.

Mercer (1997) 'Predicting the Future: Based Upon Aggregated Expectations', available at http://oubs.open.ac.uk/future/OUBS/tools/lrptools.html.

Microsoft (1996) *Internet Explorer* WWW browser, The Microsoft Corporation, available at http://www.microsoft.com.

Microsoft (1997) *Microsoft Encarta* Multimedia Encyclopaedia on CD-ROM, The Microsoft Corporation at http://www.microsoft.com.

Miller G (1990) WordNet: An on-line lexical database *International Journal of Lexicography* **3** (4), special issue.

Miller M A (1991) *Internetworking: A Guide to Network Communication, LAN to LAN; LAN to WAN* Prentice-Hall, Hemel Hempstead.

Minsky M, Papert S (1969) *Perceptrons* MIT Press, Cambridge, MA.

Miyamoto S (1989) Two approaches for information retrieval through fuzzy associations *IEEE Transactions on Systems, Man and Cybernetics* **19** (1), January/February, pp 123-130.

Miyamoto S, Nakayama K (1986) Fuzzy information retrieval based on a fuzzy pseudothesaurus *IEEE Transactions on Systems, Man and Cybernetics* **16** (2), March/April, pp. 178-282.

Mizzaro S *et al.* (1996) 'The FIRE Project', The Flexible Information Retrieval Environment, available at http://www.dimi.uniud.it/labs/fire.html.

Monarch I, Carbonell J G (1987) CoalSORT: A knowledge-based interface *IEEE Expert* Spring, pp. 39-53.

Montulli L (1995) 'Lynx - Home Page' at http://info.cern.ch/hypertext/WWW/Lynx/Status.html/.

Mooney G J (1995a) *Information Retrieval from Electronic Sources* CMS Working Paper No. 62, School of Computing and Mathematical Sciences, DMU, May.

Mooney, G J (1995b) *User Modelling and An Intelligent Agent Applied to the Internet* CMS M.Phil./Ph.D. Transfer Document, School of Computing and Mathematical Sciences, DMU, September.

Mooney G J (1996) 'E-mail questionnaire to DMU Undergraduate Students, 1995/96', contact the author, gary.mooney@iop.org, for details.

Morita M, Shinoda Y (1996) Information filtering based on user behavior analysis and best match text retrieval

Morris A (1990) 'The Use of Expert Systems in the Library and Information Service Environment' in *Computers in Libraries International 90* J Eyre (ed.), Meckler, London, pp 104-107.

MUC (1995) *Proceedings of the 6th Message Understanding Conference* ARPA/Software and Intelligent Systems Technology Office, Morgan Kaufmann, San Francisco, CA.

Mukhopadhyay U, Stephens L, Huhns M (1986) An intelligent system for document retrieval in distributed office environments *Journal of the American Society for Information Science* **37**, pp. 123-135.

Myaeng, S H (1990) Integration of User Profiles: Models and Experiments in Information Retrieval *Information Processing & Management* **26** (6), pp. 719-738.

Nadeau D (1997) 'Introduction to VRML 97', available at http;//www.sdsc.edu/~nadeau/.

Nakamura K, Iwai S (1982) Topological fuzzy sets as a quantitative description of analogical inference and its application to question-answering systems for information retrieval *IEEE Transactions on Systems, Man and Cybernetics* **12** (2), March/April, pp 193-204.

Nassar N (1996) 'Isite Information System', CNDIR (Center for Networked Information Discovery and Retrieval), available at http://www.cndir/org/ir/isite.html.

NCC (1987) *On-line Databases and Electronic Mail Systems: Interconnecting Applications Handbook*, The National Computing Centre, Manchester, UK.

NCSA (1997) 'Multimedia Design for the WWW', The National Center for Supercomputing Applications, University of Illnois at Urbana-Champaign, available at http://www.ncsa.uiuc.edu/General/Training/AdvHTML/multimedia.design.html

Nessen E (1989) SC-UM: user modelling in the SINIX consultant *Applied Artificial Intelligence* **3**, pp. 33-44.

Netfind (1995) A Netfind service is available by telneting to: netfind@monolith.cc.ic.ac.uk:23

Netscape (1994) 'Mosaic Communications change name to " Netscape Communications Corp. "', press release at http://home.netscape.com/info/pressrelease5.html/

Netscape (1995) 'Netscape Communications - Home Page' at http://home.netscape.com/.

Netscape (1998) *Netscape Communicator* available at http://home.netscape.com/.

Neumann B C (1992) The Prospero file system: A global file system based on the virtual system model *Computing Systems* **5** (4), Fall, pp. 407-432.

Newell A, Shaw J C, Simon H A. (1956) 'Empirical explorations of the logic theory machine: a case study in heuristics' in *Proceedings of the Western Joint Computer Conference*, pp 218-239. (Also in *Computers and Thought* (1963) Feigenbaum and Feldman, (Eds), McGraw-Hill, pp 134-152.)

Nielsen J (1990) *Hypertext and Hypermedia* Academic Press, NY.

Nierstrasz O (1995) 'CUIW3 Catalog -Home Page' at http://cuiwww.unige.ch/w3catalog/.

Obraczka K, Danzig P B,Shih-Hao L (1993) Internet resource discovery services *Computer* September, pp. 8-22.

Oliver W D (1997) 'The Singing Tree: A Novel Interactive Musical Interface', available at http://theremim.media.mit.edu/~woliver/singhtml/title.html.

Online and CD Notes (1994a) Announcement of new CD-ROMs, December, pp. 13-14.

Online and CD Notes (1994b) News about new online services, December, pp 11-12.

Osgood C E, Suci G J, Tannenbaum P H (1957) *The Measurement of Meaning* Urbana Press, Urbana, IL.

Paice, C D (1991) A Thesaural Model of Information Retrieval *Information Processing & Management* **27** (5), pp. 433-448.

Paijmas H (1997) Indexing Texts with SMART *Linux Gazette* **13** available at http://www.ios.chalmers.se/~kivilinux/linux/lg/issue13/smart.html.

Paradise (1995) The Paradise service is available by telneting to: dua@paradise.ulcc.ac.uk:23

Parfell M (ed.) (1987) *On-line Databases and E-mail Systems* Information services department, the National Computing Centre, Manchester.

Pazzani M, Muramatsu J, Billsus D (1995) 'Syskill & Webert: Identifying Interesting Web Sites', University of California at Irvine, available at http://www.ics.uci.edu/~pazzani.

Perl (1998) 'Perl Home Page', available at http://www.perl.com.

Phoenix (1997) 'The Phoenix project', available at http://www.dmu.ac.uk/dept/admin/dld/research.html.

Pollit S (1987) CANSEARCH: An expert systems approach to document retrieval *Information Processing and Management* **23**, pp. 119-138.

Rada R *et al.* (1989) Development and application of a metric on semantic nets *IEEE Transactions on Systems, Man and Cybernetics* **19** (1), January/February, pp. 17-30.

Radecki T (1976) Mathematical model of information retrieval system based on the concept of a fuzzy thesaurus *Information Processing Management* **12**, pp. 313-318.

RealPlayer (1998) 'Real Networks – The Home of Real Audio, Real Video and Real Flash', available at http://www.realplayer.com.

Reid N G, Boore J R P (1992) *Research Methods and Statistics in Health Care* Edward Arnold, London, chapter 5-6, pp. 71-107.

Resnick P, Varian H R (1997) Introduction – Recommender systems *Communications of the ACM* **40** (3), pp. 56-58.

Rice S (1997) 'Context and Concept Filtering', available at http://gate.cruzio.com/~autospec/.

Rich E (1979) User modelling via stereotypes *Cognitive Science* **3**, pp. 329-354.

ROSAT (1998) 'UK ROSAT Guest Observer Centre', available at http://ledas-www.star.leac.uk/rosat-goc/.

Rose M T, Cass D E (1987) *ISO Transport Service on Top of the TCP* RFC 1006, May 1987.

Ross T J (1995) *Fuzzy logic with applications* McGraw-Hill, NY.

RS Components (1998) *RS Components Catalogue on CD-ROM* RS Components, information at http://rswww.com.

Rucker J, Polanco M J (1997) Siteseer: Personalised navigation for the web *Communications of the ACM* **40** (3), pp. 73-76.

Rumelhart D E, McCelland J L, PDP research group (1986) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* Vols 1 & 2, MIT press, Cambridge, MA.

Russel R (1994) Z39.50 and SR: An Overview *VINE* No. 97, December, pp. 3-8.

Saffedy W (1993) *Electronic Document Imaging Systems: Design, Evaluation and Implementation* Meckler, Westport, CT.

Salampasis M *et al.* (1996) 'Co-operative Information Retrieval in Digital Libraries', *Proceedings of the 18th Annual Colloquium of the BCS IR* Manchester, UK, March, available at http://osiris.sund.ac.uk/~cs0msa/bcsir96.htm.

Salampasis M (1999) 'An Agent-based Hypermedia Digital Library ', available at: http://osiris.sund.ac.uk/~cs0msa/welcome.html.

Salton G (1989) *Automatic Text Processing* Addison-Welsey, Cambridge, MA.

Salton G, Buckley C (1990) Improving retrieval performance by relevance feedback *Journal of the American Society for Information Science* **41** (4), pp. 288-297.

Salton G, Lesk M E (1971) 'Computer evaluation of indexing and text processing' **in** *The SMART Retrieval System: Experiments in Automatic Document Processing* ed. G Salton, Prentice-Hall, Englewood Cliffs, NJ, pp. 143-180.

Saltzer J H (1992) 'Technology , Networks and the Library of the Year 2000' **in** *Future Tendencies in Computer Science, Control and Applied Mathematics (Proceedings of the 25th Anniversary Conference of INRIA, Paris, France, December) Lecture Notes in Computer Science* **653**, Springer, Berlin.

Satellites International Ltd. (1995) *BRAQUE - A Browse & Query Environment* The Paddock, Hambridge Rd., Newbury, Berks RG14 5TQ.

Savoy J (1993) Searching information in hypertext systems using multiple sources of evidence *International Journal of Man-Machine Studies* **38**, pp 1017-1030.

Scholtes J C (1992) 'Neural Nets in Information Retrieval: A Case Study of the 1987 Pravda' **in** *Proceedings of the SPIE Conference (SPIE Volume 1710 Science of Artificial Neural Networks)* **2** (1), pp. 631-641.

Schwartz M F, Tsirigotis P G (1991) Experience with a semantically cognizant Internet white pages directory tool *Journal of Internetworking: Research and Experience* **2** (1), March, pp. 23-50.

Schwartz M F *et al.* (1992) A comparison of Internet resource discovery approaches *Computing Systems* **5** (4), Fall, pp. 461-493.

Scott P, McDonald D (1998) 'Webcats: Library Catalogues on the WWW', available at http://www.lights.com/webcats.

Secret A (1995) 'Agora - Home Page' at http://info.cern.ch/hypertext/WWW/Agora/ Overview.html/.

Shaffer R (1997) 'Practical Guide to Genetic Algorithms' available at http://cham1.nrl.navy.mil/~shaffer/practga.html.

Shakes J *et al* (1996) 'Dynamic Reference Sifting: A Case Study in the Homepage Domain' **in** Proceedings of the Sixth International WWW Conference, pp. 189-200, available at http://www.cs.washington.edu/research/ahoy/doc/paper.html.

Shoval P *et al.* (1990) ERSE: An expert retrieval system for electronic databases *Expert Systems for Information Management* **3** (2), pp. 85-114.

Silverplatter (1998) *The INSPEC Database on CD-ROM* Silverplatter, 10 Barley Mow Passage Chiswick, London W4 4PH, UK.

Sleeman D H (1985) UMFE: a user modelling front end sub-system *International Journal of Man-Machine Studies* **23**, pp. 71-88.

Smarr L (1995) 'National Center for Super Computing Applications - Home Page' at http://www.ncsa.uiuc.edu:80/General/NCSAIntro.html/.

Smith K F (1986) Robot at the reference desk *College and Research Libraries* **47**, pp. 486-490.

Spring M B (1993) The virtual library *Virtual Reality World* **1** (3/4), November/December, pp. 53-64.

Srihari S N, Lam S W, Hull J J, Srihari R K, Govindaraju V (1994) 'Intelligent Data Retrieval from Raster Images of Documents' *Proceedings of Digital Libraries '94*, The First Annual Conference on the Theory and Practice of Digital Libraries, College Station, Texas, June 1994, pp. 34-40.

Stephenson K (1996) 'History of the Internet' at http://skydive.ncsa.uiuc.edu/ people/kims/talks/pk0396/02.html.

Stevens G C (1983) User-friendly systems: a critical examination of the concept *Behavior and Information Technology* **2** (1), pp. 3-16.

Story G A *et al.* (1992) The RightPages image-based electronic library for alerting and browsing *Computer* **25** (9) September, pp 17-26.

Stratton B (1994) The transiency of CD-ROM? A reappraisal for the 1990s *Journal of Librarianship and Information Science* **26** (3), September, pp. 157- 164.

Sun, W. and Yu, C., 1994, Semantic Query Optimization *IEEE Transactions on Knowledge and Data Engineering* **6** (1), pp. 136-151.

Sun Microsystems (1998) 'JAVA Computing', available at http://www.user.sun.com/java.

Sun Microsystems (1999) 'Jini Technology Software Simplifies Network Computing', available at: http://java.sun.com/products/jini/index.html.

Tague J, Schultz R (1988) 'Some Measures and Procedures for Evaluation of the User Interface in an Information Retrieval System' **in** *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* Association for Computing Machinery, pp. 371-385.

Tan T C, Smith P, Pegman M (1990) RADA – a Research and Development Advisor incorporating artificial intelligence techniques and expert system approaches *Expert Systems with Applications* **1** pp. 171-178.

Teachers.net (1998) 'Teachers.Net - The Teacher Website!', available at: http://www.teachers.net.

Terano K, Asai K, Sujeno M (1992) *Fuzzy Systems Theory and Its Applications* , Academic Press, Boston, MA, chapter 15, pp. 231-246.

Terjanne P (1997) 'The 1998 Telecommunications Revolution: One Month to Go', presentation at International Telecommunications Union (ITU), Study Group 3 Meeting, December 1997, available at http://www.itu.int/itudoc/osg/ptspeech/chron/1997/42249.html.

Tesco (1998) 'The Tesco Homepage', available at http://www.tesco.co.uk.

Thiel T J (1992) 'Integrated CD-ROM and WORM Optical Disk Systems on the Navy's Paperless Ship' **in** *The Best of CD-ROM Professional: Timeless Tips & Practical Advice* N Herther (ed.), Online Inc., Wilton, CT, pp. 51-58.

Thompson R H, Croft W B (1989) Support for browsing in an intelligent text retrieval system *International Journal of Man-Machine Studies* **30** (6), pp. 639-668.

Tkach D S (1998) 'Information Mining with the IBM Intelligent Miner Family, IBM Corporation, available at http://www.software.ibm.com/data/iminer /fordata/download/whitefam3.html.

TPFL (1994) *The CD-ROM Directory* TPFL (Task Force Pro Libra) publishing, London.
Van Brakel P A (1993) Implications of networking CD-ROM databases in a research environment *Southern African Journal of Library and Information Science* **61** (1), pp. 28-34.

TREC, 1996, *Proceedings of the 5th Text Retrieval Conference (TREC-5)* National Institute of Standards and Technology/Defense Advanced Research Projects Agency.

Troll D A (1992) The Mercury project: meeting the expectations of electronic library patrons *Advances in Online Public Access Catalogs* **1**, pp. 114-135.

Turing A (1950) Computing Machinery and Intelligence *Mind* **LIX**, no 2236, October, pp. 433-460.

Turtle H R, Croft W B (1991) Evaluation of an inference network-based retrieval model *ACM Transactions on Information Systems* **9** (3), pp. 187-222.

UKOLN (1998)'The eLib Programme', UK Office for Library and Information Networking, at http;//www.ukoln.ac.uk/services/elib/papers/.

US Library of Congress (1995) 'Z39.50 Resources' at: http://lcweb.loc.gov/Z3950/.

Van Brakel P A (1993) Implications of networking CD-ROM databases in a research environment *Southern African Journal of Library and Information Science* **61** (1), pp. 28-34.

Van den Berg J, van Dijk J (1997) 'How to Create the Simplest FAM-based Fuzzy System', available at http://kaa.cs.few.eur.nl/few/people/jvandenberg/Publ/List/ICONIP-97-abstract.

Van Rijsbergen K, Draper S (1997) 'Proceedings of the Second Mira Workshop – 1: Combining HCI and IR Evaluation', available at http://www.dcs.gla.ac.uk/mira/workshops/padua_procs/hciir.html.

Veljkov M, Hartnell G (1994) *The Internet E-mail System: Pocket Guides to the Internet, Vol. 4* Mecklermedia, Westport, CT.

Veronica (1995) The two main UK servers are available through gopher or WWW at: gopher.ic.ac.uk.2347/7 - Imperial college; info.mcc.ac.uk.2347/7 - University of Manchester.

Vickery A, Brooks H (1987) Expert systems and their applications in LIS *Online Review* **11**, pp. 149-165.

Vielle, L., 1989, Recursive query processing: the power of logic *Theoretical Computer Science* **69** (1), pp. 1-54.

Voorhees E M (1994) 'Software Agents for Information Retrieval' **in** *Software Agents - Papers from the 1994 Spring Symposium (Technical Report SS 94-03)* O Etzioni (ed.), AAAI press, New York, pp. 126-129.

VUBIS (1992) System Overview *VINE* **88** September, pp. 30-36.

Wahlster W, Kobsa A (1986) Dialogue-based user models *Proceedings IEEE* **74** (7), pp. 948-960.

WAIS (1995) 'WAIS Inc. - " home " page' at: http://www.wais.com/newhomepages/company.html/.

Wang Y-C *et al.* (1985) Relational thesauri in information retrieval *Journal of the American Society for Information Science* **36**(1), January, pp. 15-27.

Waterlow New Media (1999) 'The Waterlow New Media Product Database', available at: http://www.newmediainfo.com/products.

Waters S T (1986) Answerman, the expert information specialist: an expert system for retrieval of information from library reference books *Information Technology and Libraries* **5** pp. 208- 220.

Weckert J, Cooper C (1989) Libraries and expert systems *Riverina Library Review* **6** (4), pp. 323-328.

Weibel S, Miller E (1997) 'Dublin Core Metadata', at http://purl.oclc.org/metadata/dublin_core/.

Weiner N (1948) *Cybernetics: or, Control and Communication in the Animal and the Machine.*

Weinzerl H (1998) 'DVD: The Latest Evolutionary Link', available at http://www.layers.com/bpslay/layersos/dvd.htm.

Welstead S T (1994) *Neural Network and Fuzzy Logic Applications in C/C++* John Wiley, New York, 494 pp.

Winder D (1995) The secret life of e-mail *.net*, January, issue 2, pp. 37-43.

Winston P H (1984) *Artificial Intelligence* 2 ed., Addison-Wesley, Reading, MA, 524pp.

Wong S K M, Yao Y Y (1995) On modelling information retrieval with probabilistic inference *ACM Transactions on Information Systems* **13** (1), January, pp. 38-68.

Woods W A (1994) 'Understanding subsumption and taxonomy - a framework for progress' in *Principles of Semantic Networks* (ed) John F Sowa.

Wooldridge M, Jennings N R (1995) Intelligent agents: theory and practice *The Knowledge Engineering Review* **10** (2), pp. 115-152.

Wright J (1921) Correlation and causation *Journal of Agricultural Research* pp. 557-585.

Wu Z, Ramsden A, Zhao D (1995) 'The User Perspective of the ELINOR Electronic Library' in *Proceedings of 1st ELVIRA (Electronic Library and Visual Information Research) Conference* (eds M Collier & K Arnold) Milton Keynes, May 1994, Aslib, London, pp. 160-172.

WWW5 (1996) 'Fifth International World Wide Web Conference – Paper Sessions', 6-10 May, Paris, France, available at http://www5conf.inria.fr/fich-htm/paper-sessions.html.

WWW8 (1999) 'Eighth International World-Wide Web Conference', 11-14 May, Toronto, Canada, available at http://www8.org.

Wyle M F (1996) 'Effective Dissemination of WAN Information' *PhD Thesis* Lasalle University, available at http://vhdl.org/~wyle/.

XML (1999) 'XML – Extensible Mark-up Language', available at http://www.xml.com.

Yager R.R. and Larsen H.L. (1993) Retrieving information by fuzzification of queries *Journal of Intelligent Information Systems* **2** (4), November, pp. 421-442.

Yahoo (1998) 'Yahoo – home page, (Yet another hierarchical officious oracle!)', available at http://www.yahoo.com.

Yeates R (1998) 'NewsAgent for Libraries', at http://www.sbu.ac.uk/~litc/newsagent.

Zadeh L A (1964) 'Fuzzy Sets' *Memo ERL, No. 64-44* University of California, Berkeley, CA

Zadeh L. A. (1965) Fuzzy sets *Information Control* **8**, pp. 338-353.

Zadeh L A (1973) Outline of a new approach to the analysis of complex systems and decision processes *IEEE Transactions on Systems, Man and Cybernetics* **3** pp 28-44.

Zadeh L A (1978) Fuzzy Sets as a Basis for a Theory of Possibility *Fuzzy Sets and Systems* **1** pp. 3-28.

Zadeh, L.A. (1989) Knowledge Representation in Fuzzy Logic *IEEE Transactions on Knowledge and Data Engineering* **1** (1), pp. 89-100.

Zadeh, L. A. (1993) 'The Role of Fuzzy Logic and Soft Computing in the Conception and Design of Intelligent Systems' **in** *Proceedings of 8th Austrian AI Conference (FLAI '93)* E P Klement, W Slany (eds) *Lecture Notes in Artificial Intelligence* **695**, Springer, Berlin, pp.1-2.

Zadeh L A (1994) 'The Role of Fuzzy Logic in Modeling, Identification and Control' **in** *Selected Papers by Lofti Zadeh*, pp. 192-202.

Zadeh, L.A. (1996) Fuzzy Logic = Computing with Words *IEEE Transactions on Fuzzy Systems* **4** (2), pp. 103-111.

Zegher de I *et al.* (1992) 'IRHIS: Intelligent, adaptive information Retrieval system as Hospital Information System front end' **in** *Advances in Medical Informatics: Results of the AIM Exploration Action* J Noothoven van Goor, J P Christensen (eds), pp. 188-198.

Zell A *et al.* (1994) *SNNS - Stuttgart Neural Network Simulator* User Manual, v. 3.3, Report 3/94, Institute for Parallel and Distributed High Performance Systems (IPVR), University of Stuttgart, Stuttgart.

Zemankova M, Kacprzyk J (1993) The role of fuzzy logic and management of uncertainty in building intelligent information systems *Journal of Intelligent Information Systems* **2** (4), November, pp. 311-318.

Zhang X (1998) 'A Study of Effects of User Characteristics on Mental Models of Information Retrieval', available at http://valinor.purdy.wayne.edu/sherman.html.

Zink S D (1992) 'Toward More Critical Reviewing and Analysis of CD-ROM User Software Interfaces' **in** *The Best of CD-ROM Professional: Timeless Tips & Practical Advice* N Herther (ed.), Online Inc., Wilton, CT, pp. 239-245.

Appendix A

The following figure A.1 shows diagrammatically the default fuzzy membership functions used in the fuzzy user modelling process. These functions provide the basis upon which the process depends. A set of the functions shown is attached to each question in the questionnaires and the shape of the functions is modified by the answers to the questions.

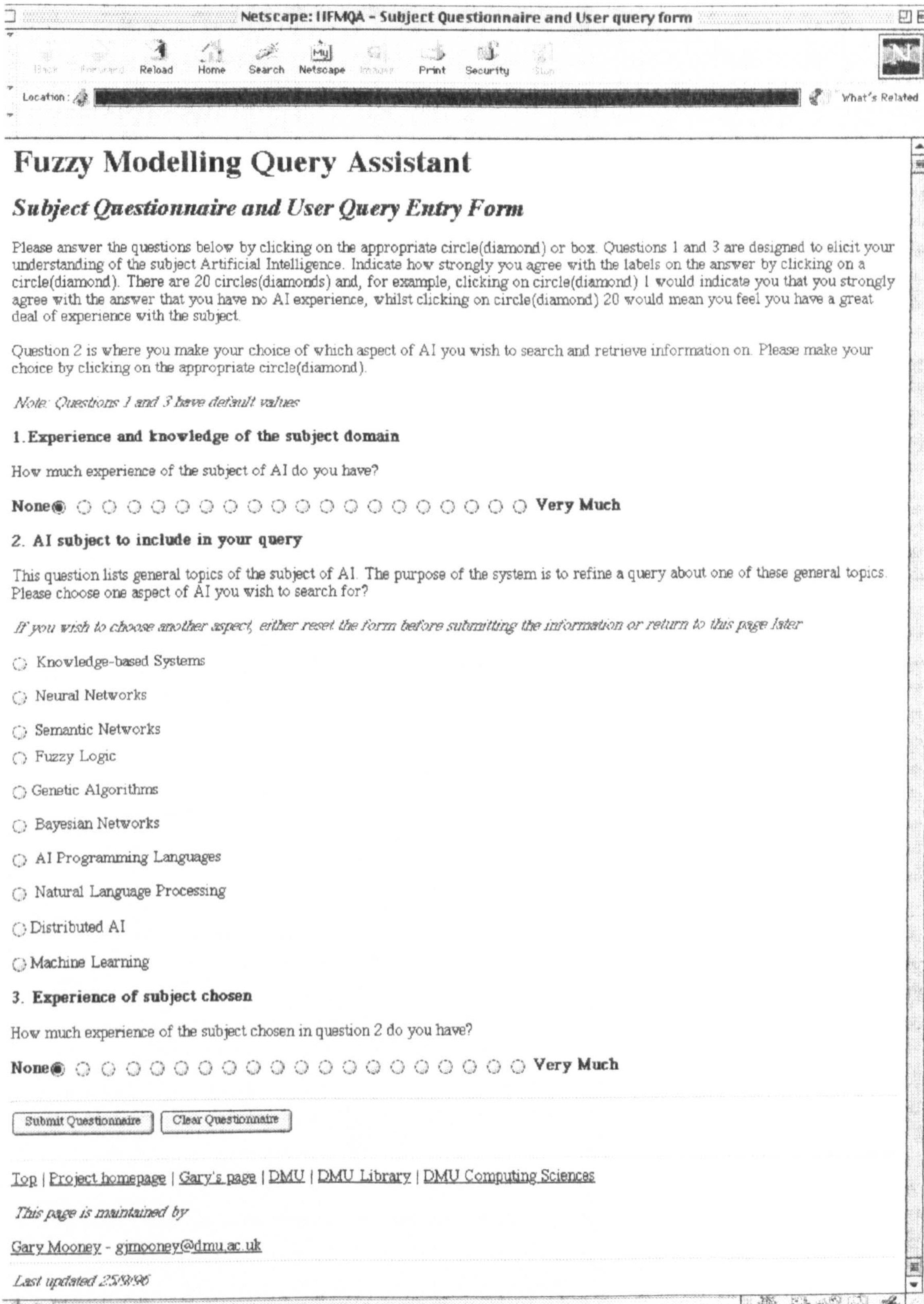
```
1.00.......                         *                        +++++++
0.90       ...                    ** **                   +++
0.80          ..               ***     ***              ++
0.70            ...          **           **         +++
0.60               ..     ***               ***    ++
0.50                 ..***                     **+++
0.40                 ** ..                     ++ **
0.30              ***     ...               +++    ***
0.20            **           ...         +++          **
0.10         ***                ..     ++               ***
0.00+++++++++++++++++++++++++++++++++++.....................*********
    0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
    0.00    2.5       5.0      7.5      10.0     12.50    15.0     17.50    20.0
```

. Fuzzy Set - Novice
Domained 0.00 to 20.0
* Fuzzy Set - Intermediate
Domained 0.00 to 20.0
+ Fuzzy Set - Expert
Domained 0.00 to 20.0

Figure A.1: Default Fuzzy Membership Functions.

Appendix B – The Questionnaires

As part of the fuzzy modelling process, the user is asked to complete two questionnaires designed to elicit information regarding their knowledge and experience of the subject domain and the Internet in general. There are also a number of background questions. As such, these two questionnaires are part of the FMQA and form part of the series of html pages which make up the user interface. The rest of the user interface is described in Appendix D – the following two figures show the questionnaires as they appeared to a user accessing the FMQA.


Netscape: IIFMQA - Subject Questionnaire and User query form

Back Forward Reload Home Search Netscape Images Print Security Stop

Location: [redacted] What's Related

Fuzzy Modelling Query Assistant

Subject Questionnaire and User Query Entry Form

Please answer the questions below by clicking on the appropriate circle(diamond) or box. Questions 1 and 3 are designed to elicit your understanding of the subject Artificial Intelligence. Indicate how strongly you agree with the labels on the answer by clicking on a circle(diamond). There are 20 circles(diamonds) and, for example, clicking on circle(diamond) 1 would indicate you that you strongly agree with the answer that you have no AI experience, whilst clicking on circle(diamond) 20 would mean you feel you have a great deal of experience with the subject.

Question 2 is where you make your choice of which aspect of AI you wish to search and retrieve information on. Please make your choice by clicking on the appropriate circle(diamond).

Note: Questions 1 and 3 have default values

1. Experience and knowledge of the subject domain

How much experience of the subject of AI do you have?

None ◉ ○ **Very Much**

2. AI subject to include in your query

This question lists general topics of the subject of AI. The purpose of the system is to refine a query about one of these general topics. Please choose one aspect of AI you wish to search for?

If you wish to choose another aspect, either reset the form before submitting the information or return to this page later

○ Knowledge-based Systems

○ Neural Networks

○ Semantic Networks

○ Fuzzy Logic

○ Genetic Algorithms

○ Bayesian Networks

○ AI Programming Languages

○ Natural Language Processing

○ Distributed AI

○ Machine Learning

3. Experience of subject chosen

How much experience of the subject chosen in question 2 do you have?

None ◉ ○ **Very Much**

[Submit Questionnaire] [Clear Questionnaire]

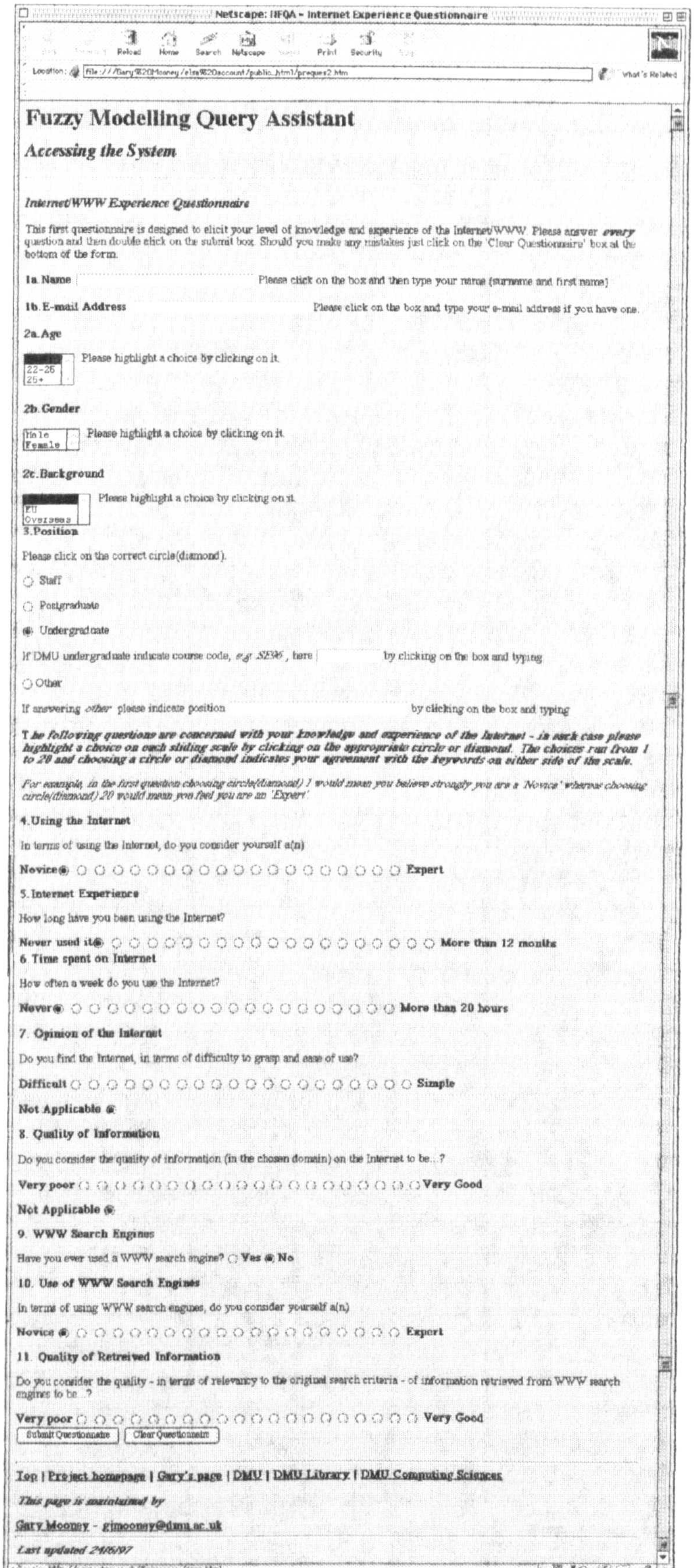
Top | Project homepage | Gary's page | DMU | DMU Library | DMU Computing Sciences

This page is maintained by

Gary Mooney - gjmooney@dmu.ac.uk

Last updated 25/8/96

Figure B.1: The Subject Domain Questionnaire.



Netscape: IIFQA - Internet Experience Questionnaire

Fuzzy Modelling Query Assistant

Accessing the System

Internet/WWW Experience Questionnaire

This first questionnaire is designed to elicit your level of knowledge and experience of the Internet/WWW. Please answer ***every*** question and then double click on the submit box. Should you make any mistakes just click on the 'Clear Questionnaire' box at the bottom of the form.

1a. Name Please click on the box and then type your name (surname and first name)

1b. E-mail Address Please click on the box and type your e-mail address if you have one.

2a. Age

22-25
25+
Please highlight a choice by clicking on it.

2b. Gender

Male
Female
Please highlight a choice by clicking on it.

2c. Background

EU
Overseas
Please highlight a choice by clicking on it.

3. Position

Please click on the correct circle(diamond).

Staff

Postgraduate

Undergraduate

If DMU undergraduate indicate course code, *e.g. SE96*, here by clicking on the box and typing

Other

If answering *other* please indicate position by clicking on the box and typing

The following questions are concerned with your knowledge and experience of the Internet - in each case please highlight a choice on each sliding scale by clicking on the appropriate circle or diamond. The choices run from 1 to 20 and choosing a circle or diamond indicates your agreement with the keywords on either side of the scale.

For example, in the first question choosing circle(diamond) 1 would mean you believe strongly you are a 'Novice' whereas choosing circle(diamond) 20 would mean you feel you are an 'Expert'.

4. Using the Internet

In terms of using the Internet, do you consider yourself a(n)

Novice ○ **Expert**

5. Internet Experience

How long have you been using the Internet?

Never used it ○ **More than 12 months**

6. Time spent on Internet

How often a week do you use the Internet?

Never ○ **More than 20 hours**

7. Opinion of the Internet

Do you find the Internet, in terms of difficulty to grasp and ease of use?

Difficult ○ **Simple**

Not Applicable ○

8. Quality of Information

Do you consider the quality of information (in the chosen domain) on the Internet to be...?

Very poor ○ **Very Good**

Not Applicable ○

9. WWW Search Engines

Have you ever used a WWW search engine? ○ **Yes** ○ **No**

10. Use of WWW Search Engines

In terms of using WWW search engines, do you consider yourself a(n)

Novice ○ **Expert**

11. Quality of Retreived Information

Do you consider the quality - in terms of relevancy to the original search criteria - of information retrieved from WWW search engines to be...?

Very poor ○ **Very Good**

[Submit Questionnaire] [Clear Questionnaire]

Top | Project homepage | Gary's page | DMU | DMU Library | DMU Computing Sciences

This page is maintained by

Gary Mooney - gjmooney@dmu.ac.uk

Last updated 24/6/97

Figure B.2: The Background and Internet Questionnaire.

Appendix C – An Example Logfile

During the modelling process, as the user interacts with the FMQA, the fuzzy rules are written and the fuzzy sets graphically displayed in the logfile. An example of a logfile is shown in figure C.1.

```
Policy 'AIEXPPOL' created.
Default Hedges installed in Policy 'AIEXPPOL'

aiexp
Domain [UofD]:       0.00 to      20.00


      1.00.......                          *                         +++++++
      0.90       ...                    ** **                     +++
      0.80          ..               ***     ***                ++
      0.70            ...          **           **           +++
      0.60               ..     ***               ***      ++
      0.50                 ..***                     **+++
      0.40                 ** ..                     ++ **
      0.30              ***     ...               +++      ***
      0.20            **           ...         +++            **
      0.10         ***                ..     ++                 ***
      0.00+++++++++++++++++++++++++++++++++++......................*********
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50    10.00   12.50   15.00   17.50   20.00
     .   FuzzySet:    LE3
         Description:
         Domained:          0.00 to      20.00
     *   FuzzySet:    ME3
         Description:
         Domained:          0.00 to      20.00
     +   FuzzySet:    HE3
         Description:
         Domained:          0.00 to      20.00

ans1
Domain [UofD]:       0.00 to      20.00


      1.00.......                          *                         +++++++
      0.90       ...                    ** **                     +++
      0.80          ..               ***     ***                ++
      0.70            ...          **           **           +++
      0.60               ..     ***               ***      ++
      0.50                 ..***                     **+++
      0.40                 ** ..                     ++ **
      0.30              ***     ...               +++      ***
      0.20            **           ...         +++            **
      0.10         ***                ..     ++                 ***
      0.00+++++++++++++++++++++++++++++++++++......................*********
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50    10.00   12.50   15.00   17.50   20.00
     .   FuzzySet:    LOW1
         Description:
         Domained:          0.00 to      20.00
     *   FuzzySet:    MED1
         Description:
         Domained:          0.00 to      20.00
     +   FuzzySet:    HIGH1
         Description:
         Domained:          0.00 to      20.00
```

```
ans2
Domain [UofD]:       0.00 to      20.00


      1.00.......                         *                         +++++++
      0.90       ...                   ** **                    +++
      0.80          ..              ***     ***               ++
      0.70            ...         **           **          +++
      0.60               ..    ***               ***     ++
      0.50                 ..***                   **+++
      0.40                 **  ..                  ++ **
      0.30              ***      ...            +++     ***
      0.20            **            ...      +++           **
      0.10         ***                 ..  ++                ***
      0.00+++++++++++++++++++++++++++++++++++.....................*********
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50    10.00   12.50   15.00   17.50   20.00
     .   FuzzySet:    LOW2
         Description:
         Domained:          0.00 to      20.00
     *   FuzzySet:    MED2
         Description:
         Domained:          0.00 to      20.00
     +   FuzzySet:    HIGH2
         Description:
         Domained:          0.00 to      20.00

Empty FuzzySet 'NULL' created.
Empty FuzzySet 'NULL4' created.
Empty FuzzySet 'NULL5' created.
Rules for subject questionnaire
Answer for general AI experience:       1.00
Answer for AI subjects:       1.00
The Rules:
R1   if ans1 is low and ans2 is low          then aiexp is very low
R2   if ans1 is low and ans2 is medium          then aiexp is medium
R3   if ans1 is low and ans2 is high          then aiexp is medium
R4   if ans1 is medium and ans2 is low          then aiexp is low
R5   if ans1 is medium and ans2 is medium          then aiexp is medium
R6   if ans1 is medium and ans2 is high          then aiexp is high
R7   if ans1 is high and ans2 is low          then aiexp is medium
R8   if ans1 is high and ans2 is medium          then aiexp is medium
R9   if ans1 is high and ans2 is high          then aiexp is very high
VL

Internet exp/time estimation begins
FMSNote(003): Fuzzy Work Area Initialized
Empty FuzzySet 'AIEXP' created.
FMSNote(005): Output Variable 'AIEXP' added to Fuzzy Model.
```

```
        FuzzySet:    AIEXP
        Description:
      1.00
      0.90
      0.80
      0.70
      0.60
      0.50
      0.40
      0.30
      0.20
      0.10
      0.00.................................................................
         0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
        0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:         0.00 to      20.00

Rule Execution....
Empty FuzzySet 'NULL3' created.
Empty FuzzySet 'NULL2' created.
R1   if ans1 is low and ans2 is low           then aiexp is very low
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       1.00
PremiseTruth2=       1.00
PremiseRes=      1.00
OK. Fuzzy set 'LE3' copied into Fuzzy set 'LE3'
Hedge 'very' applied to Fuzzy Set "LE3"

        FuzzySet:    LE3
        Description: very LE3
      1.00.......
      0.90       .
      0.80        ..
      0.70          .
      0.60           ..
      0.50             .
      0.40              ..
      0.30                ...
      0.20                   ..
      0.10                     ....
      0.00                         .......................................
         0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
        0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:         0.00 to      20.00
```

```
         FuzzySet:    AIEXP
         Description:
       1.00.......
       0.90        .
       0.80         ..
       0.70           .
       0.60            ..
       0.50              .
       0.40               ..
       0.30                 ...
       0.20                    ..
       0.10                      ....
       0.00                          .......................................
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
         Domained:         0.00 to      20.00

Empty FuzzySet 'NULL3' created.
R2   if ans1 is low and ans2 is medium          then aiexp is medium
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       1.00
PremiseTruth2=       0.00
PremiseRes=       0.00
OK. Fuzzy set 'ME3' copied into Fuzzy set 'ME3'

         FuzzySet:    ME3
         Description:
       1.00                                 .
       0.90                               .. ..
       0.80                            ...     ...
       0.70                          ..           ..
       0.60                       ...               ...
       0.50                    ...                     ...
       0.40                  ..                           ..
       0.30               ...                               ...
       0.20             ..                                     ..
       0.10          ...                                         ...
       0.00.........                                                .........
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
         Domained:         0.00 to      20.00
```

```
  FuzzySet:    AIEXP
  Description:
1.00.......
0.90       .
0.80        ..
0.70          .
0.60           ..
0.50             .
0.40              ..
0.30                ...
0.20                   ..
0.10                     ....
0.00                         ........................................
    0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
   0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
  Domained:          0.00 to      20.00
```

```
Empty FuzzySet 'NULL3' created.
R3   if ans1 is low and ans2 is high          then aiexp is medium
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       1.00
PremiseTruth2=       0.00
PremiseRes=       0.00
OK. Fuzzy set 'ME3' copied into Fuzzy set 'ME3'
```

```
  FuzzySet:    ME3
  Description:
1.00                                .
0.90                              .. ..
0.80                           ...     ...
0.70                         ..           ..
0.60                      ...               ...
0.50                   ...                     ...
0.40                 ..                           ..
0.30              ...                               ...
0.20            ..                                     ..
0.10         ...                                         ...
0.00.........                                               .........
    0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
   0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
  Domained:          0.00 to      20.00
```

```
        FuzzySet:    AIEXP
        Description:
      1.00.......
      0.90       .
      0.80        ..
      0.70          .
      0.60           ..
      0.50             .
      0.40              ..
      0.30                ...
      0.20                   ..
      0.10                     ....
      0.00                         ........................................
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:          0.00 to      20.00
```

```
Empty FuzzySet 'NULL3' created.
R4   if ans1 is medium and ans2 is low          then aiexp is low
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       0.00
PremiseTruth2=       1.00
PremiseRes=       0.00
OK. Fuzzy set 'LE3' copied into Fuzzy set 'LE3'
```

```
        FuzzySet:    LE3
        Description:
      1.00.......
      0.90       ...
      0.80          ..
      0.70            ...
      0.60               ..
      0.50                 ...
      0.40                    ..
      0.30                      ...
      0.20                         ...
      0.10                            ..
      0.00                              ..................................
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:          0.00 to      20.00
```

```
  FuzzySet:    AIEXP
  Description:
1.00.......
0.90       .
0.80        ..
0.70          .
0.60           ..
0.50             .
0.40              ..
0.30                ...
0.20                   ..
0.10                     ....
0.00                         ........................................
    0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
   0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
  Domained:         0.00 to      20.00
```

```
Empty FuzzySet 'NULL3' created.
R5   if ans1 is medium and ans2 is medium           then aiexp is medium
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       0.00
PremiseTruth2=       0.00
PremiseRes=      0.00
OK. Fuzzy set 'ME3' copied into Fuzzy set 'ME3'
```

```
  FuzzySet:    ME3
  Description:
1.00                                .
0.90                              .. ..
0.80                           ...     ...
0.70                         ..           ..
0.60                      ...               ...
0.50                   ...                     ...
0.40                 ..                           ..
0.30              ...                               ...
0.20            ..                                     ..
0.10         ...                                         ...
0.00.........                                               .........
    0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
   0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
  Domained:         0.00 to      20.00
```

```
         FuzzySet:    AIEXP
         Description:
       1.00.......
       0.90       .
       0.80        ..
       0.70          .
       0.60           ..
       0.50             .
       0.40              ..
       0.30                ...
       0.20                   ..
       0.10                     ....
       0.00                         ......................................
           0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:          0.00 to      20.00
```

```
Empty FuzzySet 'NULL3' created.
R6   if ans1 is medium and ans2 is high           then aiexp is high
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       0.00
PremiseTruth2=       0.00
PremiseRes=      0.00
OK. Fuzzy set 'HE3' copied into Fuzzy set 'HE3'
```

```
         FuzzySet:    HE3
         Description:
       1.00                                                              .......
       0.90                                                           ...
       0.80                                                         ..
       0.70                                                      ...
       0.60                                                    ..
       0.50                                                 ...
       0.40                                               ..
       0.30                                            ...
       0.20                                         ...
       0.10                                       ..
       0.00.................................
           0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:          0.00 to      20.00
```

```
     FuzzySet:    AIEXP
     Description:
   1.00.......
   0.90          .
   0.80           ..
   0.70             .
   0.60              ..
   0.50                .
   0.40                 ..
   0.30                   ...
   0.20                      ..
   0.10                        ....
   0.00                            ........................................
       0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
      0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
     Domained:         0.00 to      20.00
```

```
Empty FuzzySet 'NULL3' created.
R7   if ans1 is high and ans2 is low          then aiexp is medium
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       0.00
PremiseTruth2=       1.00
PremiseRes=       0.00
OK. Fuzzy set 'ME3' copied into Fuzzy set 'ME3'
```

```
     FuzzySet:    ME3
     Description:
   1.00                                 .
   0.90                               .. ..
   0.80                            ...     ...
   0.70                          ..           ..
   0.60                       ...               ...
   0.50                    ...                     ...
   0.40                  ..                           ..
   0.30               ...                               ...
   0.20             ..                                     ..
   0.10          ...                                         ...
   0.00.........                                                .........
       0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
      0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
     Domained:         0.00 to      20.00
```

```
        FuzzySet:    AIEXP
        Description:
      1.00.......
      0.90       .
      0.80        ..
      0.70          .
      0.60           ..
      0.50             .
      0.40              ..
      0.30                ...
      0.20                   ..
      0.10                     ....
      0.00                         ........................................
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:          0.00 to      20.00
```

```
Empty FuzzySet 'NULL3' created.
R8   if ans1 is high and ans2 is medium           then aiexp is medium
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=       0.00
PremiseTruth2=       0.00
PremiseRes=       0.00
OK. Fuzzy set 'ME3' copied into Fuzzy set 'ME3'
```

```
        FuzzySet:    ME3
        Description:
      1.00                                .
      0.90                              .. ..
      0.80                           ...     ...
      0.70                         ..           ..
      0.60                      ...               ...
      0.50                   ...                     ...
      0.40                 ..                           ..
      0.30              ...                               ...
      0.20            ..                                     ..
      0.10         ...                                         ...
      0.00.........                                               .........
          0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
         0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
        Domained:          0.00 to      20.00
```

```
    FuzzySet:     AIEXP
    Description:
  1.00.......
  0.90         .
  0.80          ..
  0.70            .
  0.60             ..
  0.50               .
  0.40                ..
  0.30                  ...
  0.20                     ..
  0.10                       ....
  0.00                           ........................................
      0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
    0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
    Domained:          0.00 to       20.00
```

```
Empty FuzzySet 'NULL3' created.
R9   if ans1 is high and ans2 is high          then aiexp is very high
Empty FuzzySet 'NULL2' created.
Empty FuzzySet 'NULL2' created.
PremiseTruth1=        0.00
PremiseTruth2=        0.00
PremiseRes=        0.00
OK. Fuzzy set 'HE3' copied into Fuzzy set 'HE3'
Hedge 'very' applied to Fuzzy Set "HE3"
```

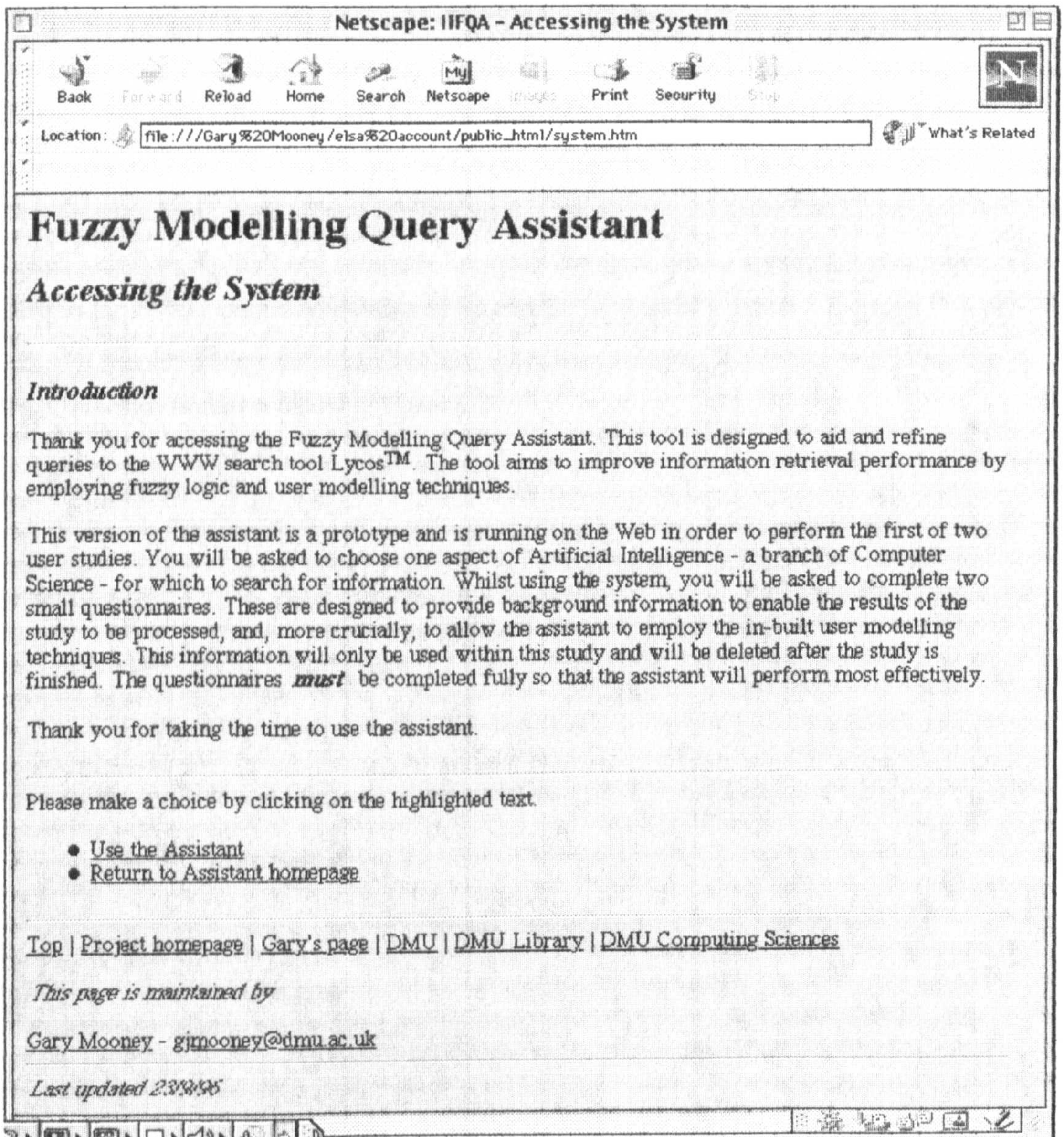
```
    FuzzySet:     HE3
    Description: very HE3
  1.00                                                           .......
  0.90                                                          .
  0.80                                                        ..
  0.70                                                       .
  0.60                                                     ..
  0.50                                                    .
  0.40                                                  ..
  0.30                                               ...
  0.20                                             ..
  0.10                                         ....
  0.00.......................................
      0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
    0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
    Domained:          0.00 to       20.00
```

```
    FuzzySet:    AIEXP
    Description:
  1.00.......
  0.90       .
  0.80        ..
  0.70          .
  0.60           ..
  0.50             .
  0.40              ..
  0.30                ...
  0.20                   ..
  0.10                     ....
  0.00                         ........................................
      0---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---0
     0.00    2.50    5.00    7.50   10.00   12.50   15.00   17.50   20.00
    Domained:          0.00 to      20.00

Empty FuzzySet 'NULL3' created.
'CENTROID'    defuzzification. Value:      2.656, [0.8329]
Model Solution:
 internet experience        =        2.66
  CompIdx     =       0.83
  SurfaceHght =       1.00
FMSNote(009): Fuzzy Work Area Closed.
```

Appendix D – The FMQA Interface

In chapter 6 of the main body of the thesis, the implementation of the prototype FMQA is described in detail. A major part of the FMQA is the html interface through which the user interacts with the system. The steps involved in this interaction are graphically displayed at a high-level in figure 6.1 (page 87). The user moves through a series of html pages – some containing forms – during an interaction. Two of these pages have already been described in Appendix B, in which the questionnaires were given. The other pages or screens are reproduced here below.



Netscape: IIFQA - Accessing the System

Back | Forward | Reload | Home | Search | Netscape | Images | Print | Security | Stop

Location: file:///Gary%20Mooney/elsa%20account/public_html/system.htm — What's Related

Fuzzy Modelling Query Assistant

Accessing the System

Introduction

Thank you for accessing the Fuzzy Modelling Query Assistant. This tool is designed to aid and refine queries to the WWW search tool Lycos™. The tool aims to improve information retrieval performance by employing fuzzy logic and user modelling techniques.

This version of the assistant is a prototype and is running on the Web in order to perform the first of two user studies. You will be asked to choose one aspect of Artificial Intelligence - a branch of Computer Science - for which to search for information. Whilst using the system, you will be asked to complete two small questionnaires. These are designed to provide background information to enable the results of the study to be processed, and, more crucially, to allow the assistant to employ the in-built user modelling techniques. This information will only be used within this study and will be deleted after the study is finished. The questionnaires ***must*** be completed fully so that the assistant will perform most effectively.

Thank you for taking the time to use the assistant.

Please make a choice by clicking on the highlighted text.

- Use the Assistant
- Return to Assistant homepage

Top | Project homepage | Gary's page | DMU | DMU Library | DMU Computing Sciences

This page is maintained by

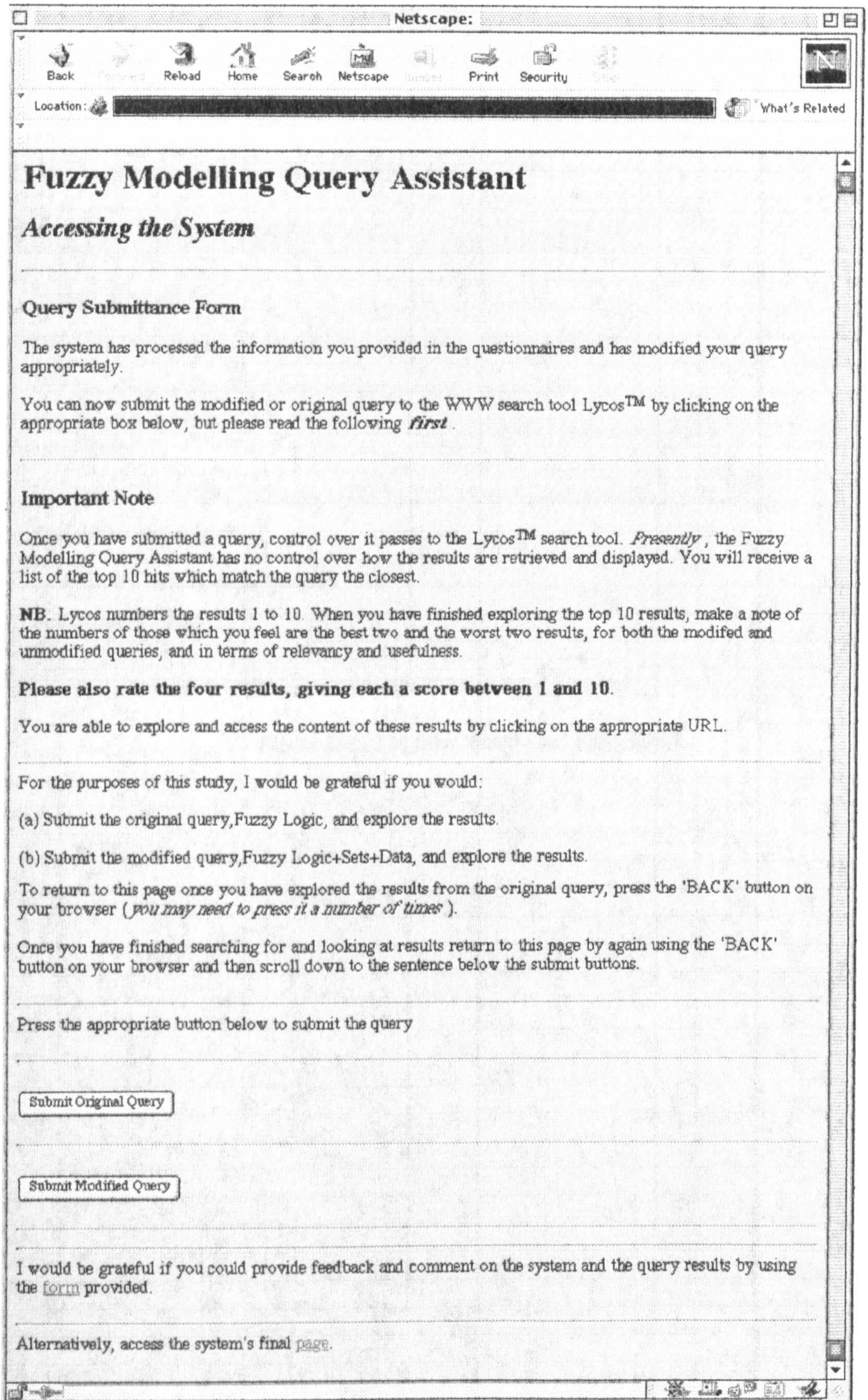
Gary Mooney - gjmooney@dmu.ac.uk

Last updated [illegible]

Figure D.1: The first page of the FMQA.

Clicking on the 'Use the Assistant' link in the page shown in figure D.1 would lead the user to the second page in the FMQA, which is the 'Background and Internet questionnaire' page given in Appendix B. On completing that questionnaire (which is and html form page) and submitting it, the user is presented with the 'Subject Questionnaire' page. Again this is an html form page and was given in Appendix B. Completing and submitting that means the user is presented with the 'Query Submit' page shown in figure D.2.

Submitting the query means that the user temporarily passes out of the realm of the FMQA as described in chapter 6. Control of the appearance and format of any list of results is handled by the Lycos web site (Lycos, 1997). In any interaction, a user may look at several results and use the functionality of the web browser, specifically the 'Back' and 'Forward' buttons to navigate between the list of IR results and the pages related to the links in the list. When a user is finished browsing and rating the results, they are asked to use the 'Back' button of the browser to return to the 'Query submit' page. There is a link on this page to the email feedback form and the user is asked to follow this link and complete the email feedback form. This html form, which is the next in the FMQA pages, was shown in its entirety in chapter 7, figure 7.1, pages 116-117. When the user has completed and submitted that form, they are presented with the last page in the FMQA which is shown below in figure D.3.



Netscape:

Back Forward Reload Home Search Netscape Images Print Security Stop

Location: [redacted] What's Related

Fuzzy Modelling Query Assistant

Accessing the System

Query Submittance Form

The system has processed the information you provided in the questionnaires and has modified your query appropriately.

You can now submit the modified or original query to the WWW search tool Lycos™ by clicking on the appropriate box below, but please read the following ***first***.

Important Note

Once you have submitted a query, control over it passes to the Lycos™ search tool. *Presently*, the Fuzzy Modelling Query Assistant has no control over how the results are retrieved and displayed. You will receive a list of the top 10 hits which match the query the closest.

NB. Lycos numbers the results 1 to 10. When you have finished exploring the top 10 results, make a note of the numbers of those which you feel are the best two and the worst two results, for both the modifed and unmodified queries, and in terms of relevancy and usefulness.

Please also rate the four results, giving each a score between 1 and 10.

You are able to explore and access the content of these results by clicking on the appropriate URL.

For the purposes of this study, I would be grateful if you would:

(a) Submit the original query,Fuzzy Logic, and explore the results.

(b) Submit the modified query,Fuzzy Logic+Sets+Data, and explore the results.

To return to this page once you have explored the results from the original query, press the 'BACK' button on your browser (*you may need to press it a number of times*).

Once you have finished searching for and looking at results return to this page by again using the 'BACK' button on your browser and then scroll down to the sentence below the submit buttons.

Press the appropriate button below to submit the query

Submit Original Query

Submit Modified Query

I would be grateful if you could provide feedback and comment on the system and the query results by using the form provided.

Alternatively, access the system's final page.

Figure D.2: The Query Submit Page.

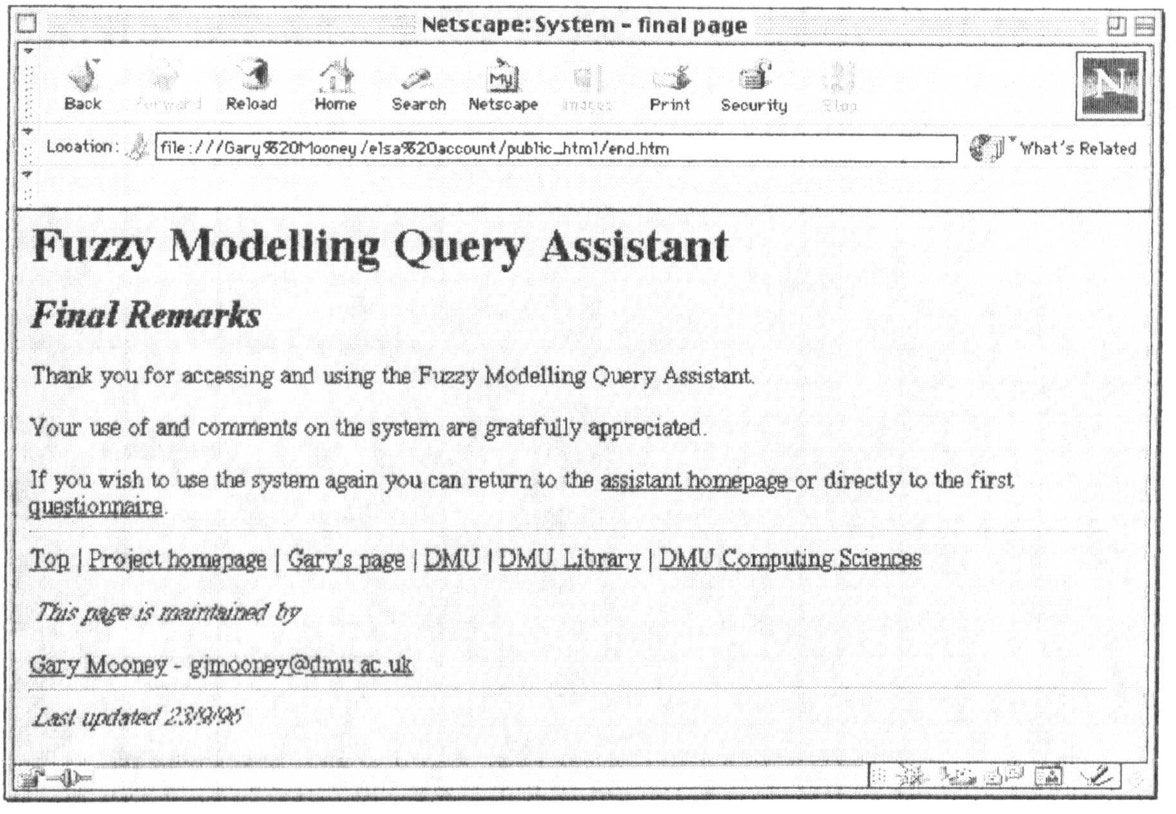


Figure D.3: The final 'thank-you' FMQA page.

Appendix E – User Studies Design Protocol

This appendix contains the original user studies design protocol. As noted in the main body of this thesis, and as is apparent from the document that follows, it was originally envisaged that there would be two studies. Lack of time and other resources prevented some of the aims of the original protocol from being fulfilled.

User Studies Design Protocol

Contents

1 Overview

This document details the design and implementation of two user studies. These were developed to evaluate research into information retrieval (IR) from the Internet. They are needed in order to test the research question. This is whether the results of IR from the World-Wide Web (WWW) can be improved (for novice and intermediate users) by the application of fuzzy logic and user modelling techniques.

The next section examines the background to the user studies in more detail, discussing how the research developed.

2 Background

Over the last 15 to 20 years there has been accelerating development in computing and telecommunications, boosted by the rise of the Internet and, recently, the part of it known as the WWW. The amount of information available globally is quite staggering – over 66 million documents as of September 1996 (Lycos, 1997) – and consequently a number of network search tools have been developed to facilitate search and retrieval.

Despite increasing sophistication (Alta Vista, 1999b), these tools are unable to capture effectively a user's information need and, additionally, the heterogeneous, dynamic, ever-changing nature of the Internet means that they suffer from the problems of irrelevance and redundancy.

It is the premise of this research that in order to remain effective, Internet IR systems need to become 'intelligent' in some way through the application of Artificial Intelligence (AI) techniques to the problem. Investigative and theoretical research revealed that the nature of the Internet precludes a number of previous and existing AI approaches to IR[1], but that a combination of fuzzy logic and user modelling techniques is a valid and fruitful approach. Being unable to rely on extensive knowledge of the retrieval database, as it is effectively unknowable, the focus of the research shifted to applying the techniques to the process of user query formulation, in order to test the research issue.

A prototype system, the Fuzzy Modelling Query Assistant (FMQA), was developed which intelligently assists the user. The concept being to refine the user's query before submitting it to an existing WWW search engine, Lycos (1997), in order to improve upon the results of using the search engine alone. Appendix C contains a diagram which schematically shows how the FMQA interacts with the user and the search tool[2].

To test the research question, two user studies have been devised. The rest of this document details their design and proposed implementation.

3 Aims of the Studies

The primary aim of the user studies is to test and answer the research question. In detail, this question – which is at the heart of this research – is:

> Whether the use of the FMQA will 'improve' upon the IR results obtained from the WWW by using the search tool alone in a certain branch of the subject domain AI, when used by 'novice' and 'intermediate' users?

Other main objectives of the studies are the following:

[1] See chapter 3 of the main body for a full discussion of this point.

[2] Excluded from the document here for clarity purposes, this diagram is given in full in chapter 6 of the main body (figure 6.6, page 97).

- to highlight any difference in the results in using the system, if any, for different types of user
- to acquire expert opinion on the performance of the system[3].

The words in quotes above are so for good reason because in order to answer this question effectively, these terms need to be understood in context. To encapsulate the concept of 'improvement', the results obtained from the FMQA and the search tool will be subject to a number of performance indicators. The relative values of these will allow qualified and accurate consideration of whether the FMQA results are indeed an 'improvement' upon those of the search tool alone. These performance indicators are discussed in detail in section 6 of this document. The terms 'novice' and 'intermediate' have special meaning as well. The user groups are discussed in more detail in the next section but they were chosen after initial research (Mooney, 1996) showed them to be novices or at an intermediate level in the categories of experience/knowledge of the Internet and experience/knowledge of AI. The FMQA is designed to decide through interaction with the user whether they are novice, intermediate or expert (to the greatest degree) in these two categories. One aim of this study is to examine whether there is any substantial difference for the different groups of users when using the system. Other background information collected, such as gender and place of origin, not only forms an integral part of the user model, but also will allow an analysis to show that any differences in the user groups reaction to the system is owing to their AI and Internet experience and not to other factors.
A recognised expert in the subject domain and the Internet will be employed to satisfy the last aim of the user studies. They will be asked to compare the sets of results of the FMQA and the search tool and comment upon them in terms of their usefulness and relevance to the original query, scoring each set of results.
All users of the system will be asked to provide feedback on the design and use of the FMQA (see Appendix A[4]). The aim of this is to acquire subjective opinion on the performance of the system and, especially between the first and second studies, to inform the adaptation of the FMQA to improve its performance, ease of use and usability.
Two user studies will take place with the same cohorts of users – the constituency of these is discussed in greater detail in the next section, whilst the detail of the actual implementation of the studies is described in section 5.

---

[3] Eventually discarded as one of the aims of the study that took place.
[4] Originally, appendix A of this document is the email feedback form that is part of the FMQA interface (see chapter 7, figure 7.1, page 117).

4 User Groups

The prototype system, FMQA, is constructed with a WWW user interface and thus is open to use by anyone with access to a web browser. Thus, some users could access the system ad hoc and their monitored results and comment will provide anecdotal evidence of the performance of the system.

However, for the purpose of collecting valid data about the system's performance the two user studies will use the same cohort of undergraduate De Montfort University (DMU) students. These user groups are formed from two courses currently running in the School of Computing Sciences at DMU. These are Software Engineering and Computer Science. The two groups chosen are the first year intakes from these courses, SE1 and CS1.

Previous research, undertaken through an email questionnaire (Mooney, 1996 – see Appendix J[5]), showed that these groups will provide a suitable cohort for the testing of the system. The students are aware of the subject of AI and of the Internet, but have not been exposed to either to such a great extent. Therefore, they will provide a suitable number of subjects which fit the categories of 'novice' and 'intermediate' users.

Additionally, as they will, during their course, acquire more experience and knowledge of AI and of the Internet, they are suitable subjects for any follow-up studies that may take place with a later version of FMQA. As a separate but integral part of the user studies, a recognised expert will be asked to interact with the system and assess the performance of the FMQA versus the search tool alone.

5 Implementation

Two user studies will take place with the same cohort of subjects. The first study will be performed in October 1996 using a subset of the user groups described above, the second in November with a larger set. The aim is to use as many subjects of the user groups as possible in the second study.

Ideally, each subject should access the system at the same time on the same day to prevent any bias from when access took place influencing the results or subject responses. This is not practically feasible, so other steps will be taken to limit any bias.

During the implementation, the crucial aspect is not where the study takes place but ***when***. The FMQA is modifying a query before it is submitted to a search engine which resides on a computer in the USA. Thus, speed of access and retrieval of results is subject to the transatlantic link and the amount of network traffic.

Thus, it is important that the user studies take place before 12 p.m. when traffic is lighter ***and*** that each user access the system at a similar time during the day, so that their responses to the system's performance are not influenced unduly by this factor. More importantly, as the Internet is constantly changing, each user study must be performed quickly so that these changes do not introduce more bias into the results. The timing of the studies and the fact that they take place quickly is crucial. Accordingly, each study will take place over a week, everyday for 5 days between the hours of 9.30 a.m. and 12.00 p.m.

Where the study is performed is less important. There will be three machines running a web browser in the library set aside for a subject to access the system at the correct times. However, if a member of the user groups can access the system from a different location, and can do so at the correct times, then the results of such an access will be deemed to be suitable for inclusion in the study. The studies are designed so that a maximum of 60 subjects can be involved in the first study and up to 120 involved in the second. Appendix B[6] contains a more detailed description of the user studies timetable.

A sample of the results, from both the FMQA and the search tool alone, for both the novice and intermediate categories, and for different queries, will be randomised and given to the expert. The randomisation process will ensure that the expert cannot know which set of results they are considering. They will be asked to consider the sets of results and choose the best three sets, in terms of to the original query, whether they are useful, and whether they are suitable results for a novice or intermediate user. The expert will also be encouraged to access the system and provide comment upon its performance.

There will be a gap of three to four weeks between the studies, during which the results of the first will inform a review of the evaluation process itself and influence any modification of the system. After the completion of the second study, the results will be subject to a statistical analysis, based upon a series of performance indicators which are discussed in the next section, in order to meet the aims of the study.

6 Evaluation

Results from both the user studies as well as from the expert will form the basis of an evaluation of the system, subject to the following performance criteria:

[5] That is, Appendix J of the thesis.
[6] Reproduced here as figure E.1.

- number of hits (results) returned
- user group member scored rating of the FMQA hits versus the search tool hits
- expert scored rating of the FMQA hits versus the search tool hits

The user/expert will be asked to decide which are the best and worst two hits in terms of relevance, usefulness and precision of the hits with regard to the original query, and then give each a score between 1 and 10. The overall rating will be calculated by taking the average of the four scores.

Other more subjective performance criteria which will inform the evaluation are:

- a measure of what the uses have lost by only accessing the first 10 hits
- user group member/expert opinion of the relevance and usefulness of the results, FMQA versus search tool alone
- user group member/expert opinion of the overall design, ease of use and functionality of the system.

The first of these criteria will be satisfied by re-running the FMQA with the users' queries and accessing the 11^{th} to 20^{th} hits. Examination of these will provide a measure of what the users lost by only considering the first 10 hits.

The hits data will be collected automatically and the ratings and opinions solicited from the user/expert as part of a session using the system. The main aim of the evaluation is to answer the research question effectively and hopefully show that such an application of AI and user modelling techniques can 'improve' the process of IR from the WWW. Of course, the FMQA is a prototype and limited in its extent but any positive results obtained are generalisable. Additionally, collecting such data as described above will allow the other aims of the study to be met. Statistical analysis of the data will show if there are any differences between the different user groups in their use of and response to the FMQA.

7 Summary

The user studies detailed above are designed to test the performance of a prototype system and provide an answer to the research question. They will be completed and evaluated by the end of December 1996. The results of the evaluation will provide evidence to either support or contradict the premise of this research, and will form an integral part of the thesis.

DATE	ACTIVITY	MAXIMUM NUMBER OF USERS
21/10/96: a.m. p.m.	User access/comment results processing/'lost'*	15
22/10/96: a.m. p.m.	User access/comment results processing/'lost'*	15
23/10/96: a.m. p.m.	User access/comment results processing/'lost'*	15
24/10/96: a.m. p.m.	User access/comment results processing/'lost'*	15
25/10/96: a.m. p.m.	Randomisation of results for expert to analyse	
28/10/96 – 15/11/96	Review of evaluation and modification of system	
18/11/96: a.m. p.m.	User access/comment results processing/'lost'*	30
19/11/96: a.m. p.m.	User access/comment results processing/'lost'*	30
20/11/96: a.m. p.m.	User access/comment results processing/'lost'*	30
21/11/96: a.m. p.m.	User access/comment results processing/'lost'*	30
22/11/96: a.m. p.m.	Randomisation of results for expert to analyse	
25/11/96 – 20/12/96	Review of evaluation, statistical analysis of results and conclusions	

* *'Lost' refers to the criterion discussed in section 6.*

Figure E.1 User Studies Timetable (as originally envisaged).

N.B. The above document was written in September 1996. There were a number of changes between the user studies envisaged in the protocol and the actual user study that took place (see chapter 7 for full discussion). Principally, these were:

- Only one study took place (due to time and resource restrictions and initial subject apathy)
- The expert performance criteria were dropped (deemed impractical and not achievable in the timescale of the project)
- The 'number of hits' performance criteria was discarded (deemed to be irrelevant as only the most important hits, the first 10, were to be examined).

Appendix F – User Study Analysis Plan

After the user study was completed and the results gathered in a suitable format for analysis (see Appendix G for the results in spreadsheet form), a user study analysis plan was constructed. This plan informed the analysis reported in chapter 8 of the main body and is reproduced below.

Analysis Plan

The crux of the research can be expressed in the following question:

> Whether the application of the FMQA to IR from the WWW can improve upon the results returned by Lycos™, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics?

This question can be restated as a research hypothesis:

> The results obtained by the application of the FMQA to IR from the WWW improve upon those obtained by using Lycos™ alone, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

In order to analyse the results so that they can provide evidence to support the above hypothesis, it is better to rephrase the above so that it is a null hypothesis. A null hypothesis is one in which an assumption of null effect is stated. This is then statistically tested and either accepted as true or rejected as false (Reid & Boore, 1992). Restated as a null hypothesis, the research hypothesis becomes:

> There is no significant improvement in the results obtained by the application of the FMQA to IR from the WWW over those obtained by using Lycos™ alone, for novice and intermediate users, employing queries in one information domain chosen from a fixed list of topics.

To address this null hypothesis and thus provide evidence to support or reject the research hypothesis, and therefore answer the crux of the research, a number of sub-questions that need to be answered have been identified. These questions are :

(1) Is the FMQA working, is it making a difference to the IR process?

(2) Is there an overall improvement in the results through the use of the FMQA?

(3) Does the FMQA improve results for different categories of user, specifically for those users deemed novice and intermediate?

(4) Does the FMQA improve results over the range of the fixed different topics within the one information domain?

Initially, in order to answer these questions Summary Statistics will be produced for both the ad-hoc and controlled portions of the results. These summary statistics will be derived from the following:

- the data provided via answers to the two questionnaires (see chapter 6, pages 92 and 94, and Appendix G)
- the responses submitted by those users who completed the email feedback form
- the values produced by the fuzzy user modelling process and their interpretation.

The raw results data falls into two distinct parts according to the fact the user study had two distinct phases – ad-hoc and controlled. The two sets of data will be dealt with independently initially. The summary statistics will be produced by categorising the results. Some of these categories are obvious, as they are related directly to the questions asked in the study. For example, the category *Background* can have the values *UK, EU, Overseas* or *Other*. However, some categories and the values are derived during the process of fuzzification and defuzzification which the FMQA performs to modify the original query. Thus, the categories *Internet Model* and *Subject Model* can have the values *Low, Medium* and *High*. These values, as will become evident, represent different ranges into which the defuzzified numbers produced by the final fuzzy user models are placed[1].

When the summary statistics are completed, the four sub-questions will be answered by performing the following analyses.

(i) Is the FMQA working, is it making a difference to the IR process?
- this question will be addressed by analysing the summary statistics and examining the means of the scores given by the users.

(ii) Is there an overall improvement in the results through the use of the FMQA?
- this question will be also addressed by analysis of the summary statistics and examining the user responses in terms of the scores. Analysis of that data for both user groups (ad-hoc and controlled) will be performed using Wilcoxon's test (Reid & Boore, 1992).

Additionally, McNemar's test will be performed on the textual comments provided via the email feedback form (Reid & Boore, 1992).

(iii) Does the FMQA improve results for different categories of user, specifically for those users deemed novice and intermediate?
- using the summary statistics to identify the user scores data which pertains to the different categories of user and the different degrees of modification, the mean scores will be analysed.

(iv) Does the FMQA improve results over the range of fixed different topics within one information domain?
- using the summary statistics once more to identify which user scores data is associated with different categories of query, the mean scores will be analysed.

Answering these questions in these ways will hopefully allow the null hypothesis to be answered and thus the research question to be addressed.

---

[1] See chapter 8 for full discussion of this point and the presentation of the summary statistics.

Appendix G: Results – Collected Data

The following pages contain spreadsheets showing the collected data for both the ad-hoc and controlled stages of the user study. Where appropriate, notes are provided to ease understanding and clarification. A key to the meaning of the header rows in the spreadsheets is given below.

rec_id	number given to each set of user data
orig_id	original number given to each set user data as it was collected
age	age range of user
gender	gender of user
bckgrnd	background of user
position	position of user
other	other category (optional position)
course	course (if undergraduate user)
i_use	Internet use opinion
i_length	length of time using Internet
i_time	time spent on Internet per week
i_opinion	opinion of Internet in terms of difficulty
i_intqual	opinion of Internet in terms of quality of information
i_eng	use of search engine (yes/no)
i_enguse	opinion of amount of use of search engine
i_engqual	opinion of quality of search engines
i_userma	ip address of user's machine
time_acc	time user began session
s_aiuse	opinion of user's experience of AI
s_query	user choice of AI topic
s_queuse	opinion of user's experience of chosen topic
i_value	FUM derived value for Internet experience
s_value	FUM derived value for AI experience
que_mod	query modification factor
unmod_s	unmodified query string
mod_s	modified query string
f_time	elapsed time spent by user in session
bumr	number of result best scored (unmodified query)
2bumr	number of result 2nd best scored (unmodified query)
bmr	number of result best scored (modified query)
2bmr	number of result 2nd best scored (modified query)
wumr	number of result worst scored (unmodified query)
2wumr	number of result 2nd worst scored (unmodified query)
wmr	number of result worst scored (modified query)
2wmr	number of result 2nd worst scored (modified query)
bumr_sc	user rating of result best scored (unmodified query)
2bumr_sc	user rating of result 2nd best scored (unmodified query)
bmr_sc	user rating of result best scored (modified query)
2bmr_sc	user rating of result 2nd best scored (modified query)
wumr_sc	user rating of result worst scored (unmodified query)
2wumr_sc	user rating of result 2nd worst scored (unmodified query)
wmr_sc	user rating of result worst scored (modified query)
2wmr_sc	user rating of result 2nd worst scored (modified query)
rcomm	user written comment on results
scomm	user written comment on system
avumr_sc	average of the best scored results (unmodified query)
avmr_sc	average of the best scored results (modified query)
avwumr_sc	average of the worst scored results (unmodified query)
avwmr_sc	average of the worst scored results (modified query)
avumod	average of all four scored results (unmodified query)
avmod	average of all four scored results (modified query)

rec_id	orig_id	age	gender	bckgrnd	position
1	4	18-21	M	UK	3
2	5	18-21	M	UK	3
3	6	18-21	M	UK	3
4	7	18-21	M	UK	3
5	8	18-21	M	UK	3
6	10	18-21	M	UK	3
7	11	18-21	M	UK	3
8	12	18-21	M	UK	3
9	13	18-21	M	UK	3
10	14		M	UK	3
11	19	18-21	M	UK	3
12	20	18-21	M	UK	3
13	21	18-21	M	UK	3
14	22	18-21	M	UK	3
15	23	18-21	M	UK	3
16	24	18-21	M	UK	3
17	25	18-21	M	UK	3
18	26	18-21	M	UK	3
19	27	18-21	M	UK	3
20	28	18-21	M	UK	3
21	29	18-21	M	UK	3
22	30		M	UK	3
23	31	18-21	M	UK	3
24	32	18-21	M	UK	3
25	33	18-21	M	UK	3
26	35	18-21	M	UK	3
27	36	18-21	M	UK	3
28	37	18-21	M	UK	3
29	38	18-21	M	UK	3
30	39	18-21	F	UK	3
31	40	18-21	F	UK	3
32	41	18-21	M	UK	3
33	42	18-21	M	UK	3
34	43	18-21	M	UK	3
35	44	18-21	M	UK	3
36	47		M		3
37	48	18-21	M	UK	3
38	50	18-21	M	UK	3
39	52	22-25	M	UK	3

Table G.1: Data taken from the Controlled Phase of the User Study

rec_id	other	course	i_use	i_length	i_time	i_opinion	i_intqual	i_eng
1		SE96	10	3	4	15	10	1
2		CS96	1	15	8	15	11	1
3		SE96	10	8	8	16	11	1
4		SE96	4	7	9	12	16	1
5		SE96	17	11	14	8	13	1
6		SE96	15	20	10	14	12	1
7		CS96	2	2	2	9	12	1
8		CS96	3	3	3	11	9	2
9		CS96	7	5	5	11	15	1
10		SE96	18	20	15	17	14	1
11		CS96	17	16	6	19	10	1
12		SE96	13	6	9	18	12	1
13		SE96	10	6	8	20	14	1
14		SE96	7	14	9	13	8	1
15		SE96	18	6	20	20	16	1
16		SE95	11	20	13	13	1	1
17		SE96	4	2	6	13	10	2
18		SE95	14	20	7	17	7	1
19		SE96	15	7	14	20	20	1
20		SE96	10	20	18	1	19	1
21		SE96	7	5	7	14	14	1
22		SE95	11	15	10	1	9	1
23		SE95	16	20	17	10	16	1
24		SE95	15	20	10	20	5	1
25		CS96	11	8	8	17	11	1
26		SE96	18	19	13	17	17	1
27		SE95	16	17	19	17	17	1
28		CS96	2	4	14	16	15	1
29		CS96	11	16	5	20	15	2
30		CS96	1	3	3	9	7	1
31		CS96	1	1	3	9	10	2
32		SE96	16	14	16	18	16	1
33		SE95	14	20	10	19	7	1
34		SE96	2	2	3	9	20	2
35		SE96	16	16	18	19	19	1
36		SE95	15	20	13	17	1	1
37		SE95	15	20	10	18	15	1
38		SE96	7	16	6	11	17	1
39		MS93	11	20	5	14	6	1

Table G.1 continued

rec_id	i_enguse	i_engqual	i_userma	time_acc
1	7	6	elsa-kim.dmu.ac.uk	Mon Oct 21 12:10:25 1996
2	10	14	kl-res17.dmu.ac.uk	Tue Oct 22 10:16:42 1996
3	5	5	elsa-kim.dmu.ac.uk	Tue Oct 22 10:17:17 1996
4	6	11	elsa.dmu.ac.uk	Tue Oct 22 10:18:45 1996
5	14	13	gim.dmu.ac.uk	Tue Oct 22 10:28:13 1996
6	10	16	elsa-kim.dmu.ac.uk	Wed Oct 23 11:31:18 1996
7	2	4	elsa-kim.dmu.ac.uk	Thurs Oct 24 10:17:57 1996
8	1	10	elsa-kim.dmu.ac.uk	Thurs Oct 24 11:34:04 1996
9	7	7	gim.dmu.ac.uk	Thurs Oct 24 11:37:11 1996
10	18	19	eagle.cms.dmu.ac.uk	Thurs Oct 24 12:22:29 1996
11	14	8	elsa-kim.dmu.ac.uk	Tue Oct 29 10:38:501996
12	15	15	elsa-kim.dmu.ac.uk	Wed Oct 30 09:11:55 1996
13	14	5	elsa-kim.dmu.ac.uk	Wed Oct 30 11:01:17 1996
14	10	8	orff.cms.dmu.ac.uk	Tue Nov 5 13:28:08 1996
15	19	15	oriole.cms.dmu.ac.uk	Tue Nov 5 13:39:22 1996
16	12	14	bessel1.cms.dmu.ac.uk	Tue Nov 5 14:26:34 1996
17	1	1	aether.cms.dmu.ac.uk	Tue Nov 5 15:34:05 1996
18	17	5	harrier.cms.dmu.ac.uk	Tue Nov 5 15:57:02 1996
19	6	1	isidore.cms.dmu.ac.uk	Tue Nov 5 17:08:22 1996
20	12	15	aether.cms.dmu.ac.uk	Wed Nov 6 11:19:38 1996
21	12	15	kodaly.cms.dmu.ac.uk	Wed Nov 6 16:24:50 1996
22	16	20	freya.dmu.ac.uk	Fri Nov 8 10:53:32 1996
23	16	16	fermat.cms.dmu.ac.uk	Mon Nov 11 08:31:32 1996
24	12	2	egate.lut.ac.uk	Mon Nov 11 11:31:29 1996
25	10	8	jw36-2.cms.dmu.ac.uk	Tue Nov 12 10:14:15 1996
26	14	15	elsa-kim.dmu.ac.uk	Wed Nov 13 14:54:32 1996
27	17	8	bittern.cms.dmu.ac.uk	Thurs Nov 14 10:42:45 1996
28	15	12	egate.lut.ac.uk	Thurs Nov 14 11:15:35 1996
29	1	1	elsa-kim.dmu.ac.uk	Fri Nov 15 14:06:57 1996
30	2	7	gim.dmu.ac.uk	Fri Nov 15 14:11:52 1996
31	1	11	elsa.dmu.ac.uk	Fri Nov 15 14:12:39 1996
32	13	14	oughtred.cms.dmu.ac.uk	Thurs Nov 21 09:28:13 1996
33	6	9	curlew.cms.dmu.ac.uk	Thurs Nov 21 12:53:56 1996
34	1	1	gim.dmu.ac.uk	Tue Nov 26 10:11:53 1996
35	17	14	euler.cms.dmu.ac.uk	Thurs Nov 28 12:51:13 1996
36	8	1	freya.dmu.ac.uk	Mon Dec 2 15:18:07 1996
37	17	15	aether.cms.dmu.ac.uk	Tue Dec 3 13:22:15 1996
38	6	16	elsa-kim.dmu.ac.uk	Tue Dec 10 11:01:55 1996
39	15	6	wwwcache.foobar.net	Thu Dec 19 19:47:06 1996
			Table G.1 continued	

rec_id	s_aiuse	s_query	s_queuse	i_value	s_value	que_mod
1	5	Neural Networks	5	14.77	5	E
2	2	AI Programming Languages	2	16.56	2.66	S
3	4	Natural Language Processing	4	14.22	5.31	S
4	1	Fuzzy Logic	1	16.56	2.66	S
5	15	Neural Networks	12	16.56	11.09	E
6	4	Fuzzy Logic	3	16.56	4.14	S
7	2	Neural Networks	1	12.66	2.66	C
8	2	Machine Learning	1	14.84	2.66	S
9	3	Knowledge-based Systems	2	16.02	2.66	S
10						
11	5	Knowledge-based Systems	6	16.56	7.42	E
12	6	Genetic Algorithms	1	16.56	3.36	S
13	1	Fuzzy Logic	1	16.56	2.66	S
14				15.47		
15	7	Fuzzy Logic	6	16.56	7.34	E
16	1	Machine Learning	1	11.88	2.66	C
17				8.83		
18	3	Machine Learning	3	15	3.98	S
19	1	Machine Learning	1	13.98	2.66	C
20	1	Machine Learning	1	16.56	2.66	S
21	1	Fuzzy Logic	1	16.56	2.66	S
22	5	Machine Learning	3	16.56	4.45	S
23	10	Neural Networks	12	16.56	10.86	E
24	7	Machine Learning	9	13.44	9.38	E
25	7	Neural Networks	1	16.48	3.36	S
26	1	Neural Networks	1	16.56	2.66	S
27	15	Neural Networks	11	16.56	10.47	E
28	1	Machine Learning	1	15.94	2.66	S
29	3	AI Programming Languages	2	14.38	2.66	S
30	2	Machine Learning	1	13.13	2.66	C
31	8	Knowledge-based Systems	7	13.67	8.28	E
32	14	Fuzzy Logic	11	16.56	10.63	E
33	1	Machine Learning	1	14.84	2.66	S
34	2	Knowledge-based Systems	7	12.27	8.36	E
35	12	Fuzzy Logic	2	16.56	5.16	S
36	5	Machine Learning	2	11.17	2.97	C
37	7	AI Programming Languages	3	16.56	4.61	S
38	6	Natural Language Processing	6	16.56	7.34	E
39	10	Neural Networks	12	13.75	10.86	E
		Table G.1 continued				

rec_id	unmod_s	mod_s
1	Neural Networks	
2	AI Programming Languages	
3	Natural Language Processing	
4	Fuzzy Logic	
5	Neural Networks	
6	Fuzzy Logic	
7	Neural Networks	
8	Machine Learning	
9	Knowledge-based Systems	
10		
11	Knowledge-base Systrems	
12	Genetic Algorithms	
13	Fuzzy Logic	
14		
15	Fuzzy Logic	
16	Machine Learning	AI Machine
17		
18	Machine Learning	
19	Machine Learning	AI Machine
20	Machine Learning	
21	Fuzzy Logic	
22	Machine Learning	
23	Neural Networks	
24	Machine Learning	
25	Neural Networks	
26	Neural Networks	
27	Neural Networks	
28	Machine Learning	
29	AI Programming Languages	
30	Machine Learning	
31	Knowledge-based Systems	
32	Fuzzy Logic	
33	Machine Learning	
34	Knowledge-based Systems	
35	Fuzzy Logic	
36	Machine Learning	AI Machine
37	AI Programming Languages	
38	Natural Language Processing	
39	Neural Networks	
		Table G.1 continued

rec_id	f_time	bumr	2bumr	bmr	2bmr
1	None				
2	Tue Oct 22 11:26:47 1996	5	10	1	9
3	Tue Oct 22 11:27:06 1996	2		2	
4	Tue Oct 22 11:26:57 1996	3	4	1	3
5	Tue Oct 22 11:27:15 1996	1	2	3	7
6	Wed Oct 23 12:26:58 1996	5		8	
7	Thu Oct 24 11:09:17 1996	2	4	1	9
8	Thu Oct 24 12:33:31 1996	5	8	1	2
9	Thu Oct 24 12:33:03 1996	1	4	1	3
10	None				
11	Tue Oct 29 11:07:09 1996	1	9	1	8
12	Wed Oct 30 10:10:05 1996	1	9	3	9
13	Wed Oct 30 11:52:39 1996	1	2	1	2
14	None				
15	None				
16	None				
17	None				
18	None				
19	None				
20	None				
21	Wed Nov 6 17:21:25 1996	2	4	2	3
22	None				
23	None				
24	None				
25	None				
26	Wed Nov 13 15:42:00 1996	1	4	3	4
27	None				
28	Thurs Nov 14 11:46:15 1996	3	10	1	3
29	Fri Nov 15 14:58:11 1996	5	10	5	6
30	Fri Nov 15 14:58:20 1996	4	10	1	5
31	Fri Nov 15 14:58:29 1996	1	6	1	2
32	Thurs Nov 21 09:44:00 1996	1	2	1	8
33	None				
34	Tue Nov 26 10:57:59 1996	5		6	
35	Thurs Nov 28 13:38:19 1996	1	2	2	4
36	Mon Dec 2 15:38:32 1996	3	8	1	6
37	None				
38	Tue Dec 10 11:55:45 1996	5		5	
39	None				

Table G.1 continued

rec_id	wumr	2wumr	wmr	2wmr	bumr_sc	2bumr_sc	bmr_sc	2bmr_sc
1								
2	9		4		7	6	9	9
3			10		6	6	6	6
4	2	5	2	5	9	8	10	7
5	6	8	8	10	8	7	8	8
6	3		6		5	5	9	8
7	3	5	2	7	8	8	9	9
8	1	4	3	8	6	5	8	6
9	6	8	4	6	9	7	9	8
10								
11	6	10	5	10	9	7	10	10
12	4	7	1	4	9	8	10	9
13	6	7	3	5	9	8	9	9
14								
15								
16								
17								
18								
19								
20								
21	7	8	8	9	8	7	9	9
22								
23								
24								
25								
26	8	9	2	6	9	7	7	6
27								
28	7	8	8	9	7	6	9	8
29	6	7	1	9	5	10	6	5
30	1	3	2	3	7	5	7	6
31	2	10	4	6	8	3	9	7
32	3	6	5	6	9	7	9	7
33								
34	2		1		4	6	7	6
35	8	10	7	10	8	9	8	8
36	4	5	4	5	4	4	7	6
37								
38	7		5		5	4	3	3
39								

Table G.1 continued

rec_id	wumr_sc	2wumr_sc	wmr_sc	2wmr_sc	rcomm
1					
2	1	2	5	6	Modified Results Better
3	6	6	6	6	Similar Results
4	1	1	1	1	Modified Results Better
5	3	2	2	2	Modified Results Better
6	3	1	6	6	Modified Results Better
7	2	3	3	4	Modified Results Better
8	2	3	4	4	Modified Results Better
9	2	1	4	3	Modified Results Better
10					
11	3	1	1	1	Modified Results Better
12	2	3	1	4	Response Unclear
13	1	2	4	4	Modified Results Better
14					
15					
16					
17					
18					
19					
20					
21	1	2	1	2	Modified Results Better
22					
23					
24					
25					
26	4	3	4	3	Unmodified Results Better
27					
28	3	3	1	1	Modified Results Better
29	7	6	1	9	Modified Results Better
30	2	1	3	2	Modified Results Better
31	1	1	4	5	Modified Results Better
32	7	9	8	6	Modified Results Better
33					
34	3	3	4	5	Modified Results Better
35	1	3	2	4	Modified Results Better
36	2	1	1	1	Response Unclear
37					
38	5	4	3	4	Modified Results Better
39					

Table G.1 continued

scomm	rec_id
	1
Well designed, easy to use	2
Well designed, easy to use	3
Well designed, easy to use	4
Well designed, easy to use	5
Well designed, easy to use	6
Well designed, easy to use	7
Well designed, easy to use	8
Well designed, easy to use	9
	10
Well designed, easy to use	11
Well designed, easy to use	12
Well designed, easy to use	13
	14
	15
	16
	17
	18
	19
	20
Well designed, easy to use	21
	22
	23
	24
	25
Well designed, easy to use	26
	27
Well designed, easy to use	28
Well designed, easy to use	29
Well designed, easy to use	30
Well designed, easy to use	31
Well designed, easy to use	32
	33
Room for improvement	34
Well designed, easy to use	35
Well designed, easy to use	36
	37
Well designed, easy to use	38
	39

Table G.1 continued

rec_id	bumr_sc	bmr_sc	diff	ranked diffs	rank		totals	
1				(zeroes excluded				
2				and signs ignored)			positive	negative
3	7	9	2	1	4		4	-10
4	6	6	0	1	4		4	-10
5	9	10	1	1	4		4	
6	8	8	0	1	4		4	
7	5	9	4	1	4		4	
8	8	9	1	1	4		4	
9	6	8	2	1	4		4	
10	9	9	0	2	10		10	
11				2	10		10	
12	9	10	1	2	10		10	
13	9	10	1	-2	-10		13.5	
14	9	9	0	-2	-10		13.5	
15				3	13.5		15	
16				3	13.5			
17				4	15		90	-20
18								
19								
20					T	20		
21								
22	8	9	1		dof	15	(22-7)	
23								
24					T (5%)	25		
25								
26								
27	9	7	-2					
28								
29	7	9	2					
30	5	6	1					
31	7	7	0					
32	8	9	1					
33	9	9	0					
34								
35	4	7	3					
36	8	8	0					
37	4	7	3					
38								
39	5	3	-2					

Table G.2: Calculation of the Wilcoxon Test forBest Scored Results Paired Data

rec_id	2bumr_sc	2bmr_sc	diff	ranked diffs	rank		totals	
1				(zeroes excluded				
2	6	9	3	and signs ignored)			positive	negative
3	6	6	0					
4	8	7	-1	1	6		6	-6
5	7	8	1	1	6		6	-6
6	5	8	3	1	6		6	-6
7	8	9	1	1	6		6	-6
8	5	6	1	1	6		6	
9	7	8	1	1	6		6	
10				1	6		6	
11	7	10	3	-1	-6		13	
12	8	9	1	-1	-6		13	
13	8	9	1	-1	-6		13	
14				-1	-6		16	
15				2	13		16	
16				2	13		16	
17				2	13		18	
18				3	16			
19				3	16			
20				3	16			
21	7	9	2	4	18		137	-24
22								
23					T	24		
24								
25					dof	18	(22-4)	
26	7	6	-1					
27					T (5%)	40		
28	6	8	2					
29	10	5	error = 0					
30	5	6	1					
31	3	7	4					
32	7	7	0					
33								
34	6	6	0					
35	9	8	-1					
36	4	6	2					
37								
38	4	3	-1					
39								

Table G.3: Calculation of the Wilcoxon test for the 2nd Best Scored Results Paired Data

rec_id	wumr_sc	wmr_sc	diff	ranked diffs	rank		totals	
1				(zeroes excluded				
2	1	5	4	and signs ignored)			positive	negative
3	6	6	0					
4	1	1	1	1	5		5	-5
5	3	2	-1	1	5		5	-5
6	3	6	3	1	5		5	-5
7	2	3	1	1	5		5	-12
8	2	4	2	1	5		5	-12
9	2	4	2	1	5		5	-12
10				-1	-5		12	
11	3	1	-2	-1	-5		12	
12	2	1	-1	-1	-5		16	
13	1	4	3	2	12		16	
14				2	12		16	
15				-2	-12		18	
16				-2	-12			
17				-2	-12			
18				3	16			
19				3	16			
20				3	16			
21	1	1	0	4	18		120	-41
22								
23					T	41		
24								
25					dof	18	(22-4)	
26	4	4	0					
27					T (5%)	40		
28	3	1	-2					
29	7	1	error=0					
30	2	3	1					
31	1	4	3					
32	7	8	1					
33								
34	3	4	1					
35	1	2	1					
36	2	1	-1					
37								
38	5	3	-2					
39								

Table G4: Calculation of the Wilcoxon Test for the Worst Scored Results Paired Data

rec_id	2wumr_sc	2wmr_sc	diff	ranked diffs	rank		totals	
1				(zeroes excluded				
2	2	6	4	and signs ignored)			positive	negative
3	6	6	0					
4	1	1	0	1	3.5		3.5	-8
5	2	2	0	1	3.5		3.5	
6	1	6	5	1	3.5		3.5	
7	3	4	1	1	3.5		3.5	
8	3	4	1	1	3.5		3.5	
9	1	3	1	1	3.5		3.5	
10				2	8		8	
11	1	1	0	2	8		8	
12	3	4	1	-2	-8		10.5	
13	2	4	2	4	10.5		10.5	
14				4	10.5		12	
15				5	12			
16							70	-8
17								
18					T	8		
19								
20					dof	12	(22-10)	
21	2	2	0					
22					T (5%)	13		
23								
24								
25								
26	3	3	0					
27								
28	3	1	-2					
29	6	9	error=0					
30	1	2	1					
31	1	5	4					
32	9	6	error=0					
33								
34	3	5	2					
35	3	4	1					
36	1	1	0					
37								
38	4	4	0					
39								

Table G.5: Calculation of the Wilcoxon Test for the 2nd Worst Scored Results Paired Data

rec_id	bumr_sc	2bumr_sc	bmr_sc	2bmr_sc		avumr_sc	avmr_sc	diff
1								
2	7	6	9	9		6.5	9	2.5
3	6	6	6	6		6	6	0
4	9	8	10	7		8.5	8.5	0
5	8	7	8	8		7.5	8	0.5
6	5	5	9	8		5	8.5	3.5
7	8	8	9	9		8	9	1
8	6	5	8	6		5.5	7	1.5
9	9	7	9	8		8	8.5	0.5
10								
11	9	7	10	10		8	10	2
12	9	8	10	9		8.5	9.5	1
13	9	8	9	9		8.5	9	0.5
14								
15								
16								
17								
18								
19								
20								
21	8	7	9	9		7.5	9	1.5
22								
23								
24								
25								
26	9	7	7	6		8	6.5	-1.5
27								
28	7	6	9	8		6.5	7.5	1
29	5	10	6	5		error	error	error=0
30	7	5	7	6		6	6.5	0.5
31	8	3	9	7		5.5	8	2.5
32	9	7	9	7		8	8	0
33								
34	4	6	7	6		5	6.5	1.5
35	8	9	8	8		8.5	8	-0.5
36	4	4	7	6		4	6.5	2.5
37								
38	5	4	3	3		4.5	3	-1.5
39								

rec_id	ranked diffs	rank		totals	
1	(zeroes excluded				
2	and signs ignored)			positive	negative
3					
4	0.5	3		3	-3
5	0.5	3		3	-11
6	0.5	3		3	-11
7	0.5	3		3	
8	-0.5	-3		7	
9	1	7		7	
10	1	7		7	
11	1	7		11	
12	1.5	11		11	
13	1.5	11		11	
14	1.5	11		14	
15	-1.5	-11		16.5	
16	-1.5	-11		16.5	
17	2	14		16.5	
18	2.5	16.5		16.5	
19	2.5	16.5			
20	2.5	16.5			
21	3.5	16.5		146	-25
22					
23					
24		**T**	25		
25					
26		**dof**	18	(22-4)	
27					
28		**T (5%)**	40		
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					

Table G.6: Calculation of the Wilcoxon Test for the Average Best Scored Results Paired Data

rec_id	wumr_sc	2wumr_sc	wmr_sc	2wmr_sc	avwumr_sc	avwmr_sc	diff	
1								
2	1	2	5	6	1.5	5.5	4	
3	6	6	6	6	6	6	0	
4	1	1	1	1	1	1	0	
5	3	2	2	2	2.5	2	-0.5	
6								
7								
8								
9	3	1	6	6	2	6	4	
10	2	3	3	4	2.5	3.5	1.5	
11	2	3	4	4	2.5	4	1.5	
12	2	1	4	3	1.5	3.5	2	
13								
14	3	1	1	1	2	1	-1	
15	2	3	1	4	2.5	2.5	0	
16	1	2	4	4	1.5	4	2.5	
17								
18								
19								
20								
21								
22								
23								
24	1	2	1	2	1.5	1.5	0	
25								
26								
27								
28								
29	4	3	4	3	3.5	3.5	0	
30								
31	3	3	1	1	3	1	-2	
32	7	6	1	9	error	error	error=0	
33	2	1	3	2	1.5	2.5	1	
34	1	1	4	5	1	4.5	3.5	
35	7	9	8	6	error	error	error=0	
36								
37	3	3	4	5	3	4.5	1.5	
38	1	3	2	4	2	3	1	
39	2	1	1	1	1.5	1	-0.5	
	5	4	3	4	4.5	3.5	-1	

rec_id	ranked diffs	rank		totals	
1	(zeroes excluded				
2	and signs ignored)			positive	negative
3					
4	-0.5	-1.5		5.5	-1.5
5	-0.5	-1.5		5.5	-1.5
6	1	5.5		8	-5.5
7	1	5.5		8	-5.5
8	-1	-5.5		8	-10.5
9	-1	-5.5		10.5	
10	1.5	8		12	
11	1.5	8		13	
12	1.5	8		14.5	
13	2	10.5		14.5	
14	-2	-10.5			
15	2.5	12			
16	3.5	13			
17	4	14.5			
18	4	14.5			
19				99.5	-24.5
20		T	24.5		
21					
22		dof	15	(22-7)	
23					
24		T (5%)	25		
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					

Table G.7: Calculation of the Wilcoxon Test for the Average Worst Scored Results Paired Data

rec_id	avumod	avmod	diff	ranked diffs	rank		totals	
1				(zeroes excluded				
2	4	7.25	3.25	and signs ignored)			positive	negative
3	6	6	0					
4	4.75	4.75	0	0.25	1		1	-4
5	5	5	0	0.5	2		2	-8.5
6	3.5	7.25	3.75	0.75	4		4	
7	5.25	6.25	1	0.75	4		4	
8	4	5.5	1.5	-0.75	-4		6.5	
9	4.75	6	1.25	1	6.5		6.5	
10				1	6.5		8.5	
11	5	5.5	1.5	1.25	8.5		11.5	
12	5.5	6	0.5	-1.25	-8.5		11.5	
13	5	6.5	1.5	1.5	11.5		11.5	
14				1.5	11.5		11.5	
15				1.5	11.5		14	
16				1.5	11.5		15	
17				3	14		16	
18				3.25	15			
19				3.75	16		123.5	-12.5
20								
21	4.5	5.25	0.75		T	12.5		
22								
23					dof	16	(22-6)	
24								
25					T (5%)	29		
26	5.75	5	-0.75					
27								
28	4.75	4.75	0					
29	error	error	error=0					
30	3.75	4.5	0.75					
31	3.25	6.25	3					
32	error	error	error=0					
33								
34	4	5.5	1.5					
35	5.25	5.5	0.25					
36	2.75	3.75	1					
37								
38	4.5	3.25	-1.25					
39								

Table G.8: Calculation of the Wilcoxon Test for the Average of the Four Scored Results Paired Data

unmodified query	av_bum_s	av_bm_s	av_wum_s	av_wm_s		
Genetic Algorithms	8.5	9.5	2.5	2.5		
Natural Language Processing	6	6	6	6		
	4.5	3	4.5	3.5		
AI Programming Languages	6.5	9	1.5	5.5		
	error	5.5	6.5	error		
Neural Networks	7.5	8	2.5	2		
	8	9	2.5	3.5		
	8	6.5	3.5	3.5		
Machine Learning	5.5	7	2.5	4		
	6.5	7.5	3	1		
	6	6.5	1.5	2.5		
	4	6.5	1.5	1		
Knowledge-based Systems	8	8.5	1.5	3.5		
	8	10	2	1		
	5.5	8	1	4.5		
	5	6.5	3	4.5		
Fuzzy Logic	8.5	8.5	1	1		
	5	7.5	2	6		
	8.5	9	1.5	4		
	7.5	9	1.5	1.5		
	8	8	8	7		
	8.5	8	2	3		

Table G.9: Mean Scores versus Original Query (Controlled Phase)

Mod_cat	unmodified query	av_bum_s	av_bm_s	av_wum_s	av_wm_s	que_mod
Med-Low	Neural Networks	8	9	2.5	3.5	C
	Machine Learning	6	6.5	1.5	2.5	C
	Machine Learning	4	6.5	1.5	1	C
Med-Med	Knowledge-based Systems	5.5	8	1	4.5	E
	Knowledge-based Systems	5	6.5	3	4.5	E
High-Low	Genetic Algorithms	8.5	9.5	2.5	2.5	S
	Natural Language Processing	6	6	6	6	S
	AI Programming Languages	6.5	9	1.5	5.5	S
	AI Programming Languages	error	5.5	6.5	error	S
	Neural Networks	8	6.5	3.5	3.5	S
	Machine Learning	5.5	7	2.5	4	S
	Machine Learning	6.5	7.5	3	1	S
	Knowledge-based Systems	8	8.5	1.5	3.5	S
	Fuzzy Logic	8.5	8.5	1	1	S
	Fuzzy Logic	5	7.5	2	6	S
	Fuzzy Logic	8.5	9	1.5	4	S
	Fuzzy Logic	7.5	9	1.5	1.5	S
	Fuzzy Logic	8.5	8	2	3	S
	(Means of Means)	7.25	8	2.375	3.458333	
High-Med	Natural Language Processing	4.5	3	4.5	3.5	E
	Neural Networks	7.5	8	2.5	2	E
	Knowledge-based Systems	8	10	2	1	E
	Fuzzy Logic	8	8	8	7	E
	(Means of Means)	6.416667	7.25	3.5	3.75	

Table G.10: Mean Scores Versus Query Modification Category (Controlled Phase) - Also Gives Means of Means

rec_id	orig_id	age	gender	bckgrnd	position
1	1	25+	M	UK	1
2	2	25+	F	UK	1
3	3	25+	F	UK	1
4	9	25+	F	UK	1
5	15	25+	F	UK	1
6	17	25+	M	UK	4
7	18	25+	F	UK	1
8	45	25+	M	UK	1
9	46	25+	M	Overseas	4
10	51	25+	M	UK	3
11	53	18-21	M	UK	1
12	55	22-25	M	UK	4
13	56	25+	M	UK	1
14	57	22-25	M	EU	4
15	58	25+	M	UK	1
16	59	25+	F	Overseas	4
17	60	22-25	M	UK	4
18	61	22-25	F	UK	2
19	62	25+	M	UK	4
20	63	18-21	M	UK	3
21	64	25+	M	UK	3
22	65	25+	M	UK	4
23	66	25+	M	EU	4
24	67	22-25	M	EU	2
25	68	22-25	M	UK	2
26	69	22-25	M	EU	1
27	70	25+	F	Overseas	4
28	71	25+	M	Overseas	2
29	72	22-25	M	EU	2
30	73	25+	M	Overseas	4
31	75	25+	M	UK	1
32	76	25+	M	Overseas	4
33	77	25+	M	Overseas	2
34	78	18-21	M	Overseas	4
35	79	22-25	M	Overseas	4
36	80	25+	M	Overseas	2
37	81	22-25	M	EU	2
38	82	22-25	M	EU	3
39	83	22-25	M	EU	3
40	84	22-25	F	EU	4
41	85	25+	M	Overseas	2
42	86	25+	M	EU	1
43	87	25+	M	UK	2
44	88	25+	M	Overseas	1
45	89	25+	M	EU	2
46	90	25+	M	Overseas	2
47	91	25+	M	Overseas	2
48	92	25+	M	Overseas	2
49	93	25+	M	EU	1
50	94	25+	M	UK	1
51	95	25+	M	UK	4

Table G.11: Collected Data – Ad-hoc Phase

rec_id	orig_id	age	gender	bckgrnd	position
52	96	25+	M	UK	4
53	97	22-25	M	UK	1
54	98	22-25	M	UK	1
55	99	25+	M	Overseas	4
56	100	25+	M	Overseas	4
57	101	22-25	F	Overseas	2
58	102	22-25	M	Overseas	4
59	103	25+	M	UK	4
60	104	25+	F	Overseas	2
61	105	25+	M	Overseas	2
62	106	22-25	M	Overseas	2
63	107	22-25	M	Overseas	2
64	108	25+	M	Overseas	1
65	109	25+	M	Overseas	1
66	110	25+	M	Overseas	1
67	111	25+	M	UK	2
68	112	25+	M	EU	1
69	113	22-25	M	Overseas	4
70	114	25+	M	Overseas	1
71	115	25+	M	Overseas	1
72	116	22-25	M	EU	2
73	117	22-25	F	EU	4
74	118	22-25	M	EU	3
75	119	25+	M	UK	4

Table G.11 - continued

rec_id	other	course	i_use	i_length	i_time	i_opinion
1	research		10	10	13	11
2			16	20	16	15
3			17	20	5	20
4			1	1	1	20
5			1	3	2	15
6	IT Man.		17	20	6	20
7			11	20	4	12
8	Prin.Lect		19	20	11	20
9			3	7	20	20
10			8	15	2	8
11			9	6	5	12
12	Ph.D		10	20	6	16
13	Lect.		11	6	6	1
14	Ph.D		12	20	10	16
15			18	20	20	20
16	Sen.Lib		14	20	12	20
17	Lib.Ass		18	20	5	20
18			9	20	7	6
19	Comp.Man.		16	20	10	20
20		BIS95	14	20	13	16
21		BIS95	7	10	3	5
22	Grad stud.	LIS	11	20	15	15
23	Company		16	20	17	20
24			13	20	20	20
25			18	20	11	13
26			19	20	14	19
27			18	20	20	18
28			20	20	20	15
29			20	20	20	20
30	Inves,NASA		20	20	20	16
31			20	20	20	9
32	US CompW		18	20	20	19
33			19	20	20	19
34			17	20	20	20
35	Engineer		16	20	19	16
36			18	20	20	16
37			16	15	15	9
38			13	20	11	20
39			18	20	11	20
40	PhD		20	20	20	20
41			17	20	5	16
42			20	20	20	20
43			17	20	20	20
44			19	20	18	18
45			18	20	16	9
46			20	20	20	11
47			19	20	20	18
48			19	20	20	10
49			11	20	5	13
50			18	17	11	15
51	BT Manager		18	20	17	19

Table G.11 - continued

rec_id	other	course	i_use	i_length	i_time	i_opinion
52	Ass. Prof.		17	20	14	20
53			20	20	16	20
54			20	20	16	20
55	Post Doc		16	20	11	15
56	PhD		17	20	20	18
57	Research		14	20	9	19
58			12	17	10	20
59	Research		16	20	4	15
60			14	20	20	10
61			16	20	14	20
62			14	16	9	8
63			16	20	14	15
64			20	20	20	10
65			18	20	16	13
66			17	20	9	20
67			20	20	20	19
68			15	20	8	11
69			1	1	1	1
70			20	20	20	20
71			10	18	13	11
72			16	20	8	16
73	PhD		20	20	20	20
74			16	20	9	16
75	Res. Prof		15	20	12	20
	Table G.11 -	*continued*				

rec_id	i_intqual	i_eng	i_enguse	i_engqual	i_userma
1	5	1	12	3	germane.cms.dmu.ac.uk
2	11	1	12	5	mkdldst3.mk.dmu.ac.uk
3	17	1	18	10	freya.dmu.ac.uk
4	20	2	1	1	freya.dmu.ac.uk
5	10	2	1	1	freya.dmu.ac.uk
6	16	1	18	13	npcip5.demon.co.uk
7	1	1	10	5	caraway.hensa.ac.uk
8	15	1	15	10	femto.cms.dmu.ac.uk
9	15	1	17	15	207.50.127.34
10	12	1	11	13	a207a_04.ntu.ac.uk
11	15	1	10	6	aztec.cms.dmu.ac.uk
12	17	1	8	10	miro.cms.dmu.ac.uk
13	1	1	1	1	klein.cms.dmu.ac.uk
14	9	1	14	8	mkcsst18.mk.dmu.ac.uk
15	19	1	20	15	freya.dmu.ac.uk
16	9	1	14	10	134.7.145.138
17	12	1	14	3	wombat.dmu.ac.uk
18	8	1	9	8	kl-res17.dmu.ac.uk
19	8	1	17	11	npcip5.demon.ac.uk
20	10	1	15	8	dunlin.cms.dmu.ac.uk
21	9	1	10	6	bartok.cms.dmu.ac.uk
22	6	1	12	5	206.155.180.216
23	13	1	14	11	194.78.53.198
24	18	1	17	12	interzone.ucc.ie
25	15	1	20	15	iti-ven-036.salford.ac.uk
26	11	1	12	7	odyssey.ucc.ie
27	11	1	18	8	citecuh.citec.qld.gov.au
28	15	1	5	11	ultra2.src.ncu.edu.tw
29	16	1	20	14	gvaona1.cns.hp.com
30	14	1	20	8	jhujsak.cts.com
31	7	1	20	8	dial-in01.dmu.ac.uk
32	12	1	15	14	lpwa.com
33	11	1	12	11	ehdup-d3-3.rmt.net.pitt.edu
34	13	1	15	14	annex12-36.dialumd.edu
35	18	1	14	13	harding.reticular.com
36	20	1	18	16	163.121.239.69
37	10	1	19	8	lilie.eurecom.fr
38	15	1	18	17	pi0124.kub.nl
39	10	1	14	14	marcello.limsi.fr
40	7	1	17	5	armor.ccett.fr
41	10	1	3	9	lamprea.ait.uvigo.es
42	2	1	20	3	hze028.elca.ch
43	20	1	16	11	cax024.gcal.ac.uk
44	13	1	16	6	vantomme.bbn.com
45	7	1	18	11	agent.gmd.de
46	11	1	20	13	100.orlando-003.fl.dial-access.att.net
47	14	1	14	8	bob.coe.uga.edu
48	15	1	18	11	spirit.kaist.ac.kr
49	1	1	10	11	freya.dmu.ac.uk
50	1	1	14	15	freya.dmu.ac.uk
51	17	1	15	13	btwms.info.bt.co.uk

Table G.11 - continued

rec_id	i_intqual	i_eng	i_enguse	i_engqual	i_userma
52	9	1	15	5	ka.informatik.tu-chemnitz.de
53	10	1	20	13	158.234.138.138
54	16	1	13	14	jx5-4-7.jde.aca.mmu.ac.uk
55	11	1	16	10	neuron.usask.ca
56	20	1	18	18	gate.cs.auckland.ac.nz
57	13	1	12	9	sun.rcanaria.es
58	10	1	13	11	mondrian.tau.ac.il
59	8	1	16	3	fw1.bre.co.uk
60	9	1	15	13	tduncan4.shore.net
61	14	1	17	4	cs506.gw.ulg.ac.be
62	9	1	18	11	lilie.eurecom.fr
63	10	1	16	11	squirrel.dur.ac.uk
64	6	1	20	7	trol.info-science.uiowa.edu
65	15	1	18	13	jplbdw2.jpl.nasa.gov
66	4	1	17	6	newgate.mitel.com
67	16	1	20	15	machine3.hhcl.com
68	13	1	7	13	alkmr1-p.33.worldonline.nl
69	12	1	11	13	156.153.255.218
70	1	2	1	9	sf-pm3-2-34.dialup.slip.net
71	9	2	11	11	1cust41.max7.boston.ma.ms.uu.net
72	12	1	13	10	soling.cs.vu.nl
73	7	1	20	7	armor.ccett.fr
74	5	1	13	7	wi6a86.informatik.uni-wuerzburg.de
75	6	1	12	12	hayley.info.bt.co.uk

Table G.11 – continued

rec_id	time_acc	s_aiuse	s_query	s_queuse
1	Thurs Oct 17 11:46:28 1996	9	Knowledge-based Systems	1
2	Thurs Oct 17 12:18:38 1996	11	Knowledge-based Systems	5
3	Thurs Oct 17 12:34:33 1996	4	Semantic Networks	1
4	Tue Oct 22 13:47:07 1996	1	Fuzzy Logic	1
5	Thurs Oct 24 17:04:52 1996			
6	Fri Oct 25 09:21:06 1996	10	Knowledge-based Systems	16
7	Fri Oct 25 20:20:06 1996	2	Knowledge-based Systems	2
8	Fri Nov 29 14:14:14 1996	16	Fuzzy Logic	15
9	Sun Dec 1 17:16:59 1996	1	Knowledge-based Systems	1
10	Wed Dec 11 16:31:31 1996	1	Neural Networks	2
11	Tue Jan 7 12:38:32 1997	None	None	None
12	Tue Jan 717:35:26 1997	1	Bayesian Networks	1
13	Tue Jan 7 18:00:45 1997	1	Knowledge-based Systems	1
14	Fri Jan 31 10:37:42 1997	11	Knowledge-based Systems	6
15	Wed Feb 5 15:31:01 1997	14	Genetic Algorithms	17
16	Thu Feb 6 00:30:36 1997	3	Fuzzy Logic	1
17	Thu Feb 6 09:33:41 1997	8	Neural Networks	14
18	Fri Feb 7 10:20:23 1997	4	Fuzzy Logic	4
19	Fri Feb 7 17:07:03 1997	None	None	None
20	Mon Mar 3 15:29:21 1997	1	Neural Networks	1
21	Thu Mar 13 10:09:32 1997	3	Fuzzy Logic	1
22	Sun Mar 23 18:53:42 1997	2	Fuzzy Logic	1
23	Mon Mar 24 14:14:53 1997	15	Knowledge-based Systems	16
24	Mon Mar 24 17:27:19 1997	15	Distributed AI	16
25	Mon Mar 24 17:50:13 1997	11	Knowledge-based Systems	10
26	Mon Mar 24 18:33:31 1997	16	Semantic Networks	17
27	Tue Mar 25 03:01:49 1997	4	Neural Networks	2
28	Tue Mar 25 06:24:03 1997	18	Natural Language Processing	6
29	Tue Mar 25 07:44:13 1997	14	Genetic Algorithms	6
30	Tue Mar 25 07:51:10 1997	20	Knowledge-based Systems	20
31	Tue Aug 5 00:12:08 1997	12	Bayesian Networks	1
32	Thu Aug 7 19:40:00 1997	9	Machine Learning	6
33	Thu Aug 7 22:10:06 1997	20	Natural Language Processing	20
34	Thu Aug 7 23:42:21 1997	5	Neural Networks	6
35	Fri Aug 8 00:51:38 1997	16	Machine Learning	11
36	Fri Aug 8 06:16:09 1997	None	None	None
37	Fri Aug 8 07:21:32 1997	12	Distributed AI	12
38	Fri Aug 8 07:52:49 1997	3	AI Programming Languages	1
39	Fri Aug 8 08:36:07 1997	20	Distributed AI	14
40	Fri Aug 8 09:09:02 1997	13	Distributed AI	17
41	Fri Aug 8 09:56:13 1997	10	Distributed AI	10
42	Fri Aug 8 10:40:21 1997	20	Natural Language Processing	19
43	Fri Aug 8 13:00:46 1997	13	Distributed AI	16
44	Fri Aug 8 14:45:44 1997	11	AI Programming Languages	14
45	Fri Aug 8 15:38:55 1997	15	Natural Language Processsing	13
46	Sat Aug 9 19:05:10 1997	18	Neural Networks	17
47	Sun Aug 10 22:56:17 1997	12	Machine Learning	7
48	Mon Aug 11 05:12:28 1997	19	Neural Networks	17
49	Mon Aug 11 13:33:16 1997	1	Knowledge-based Systems	1
50	Tue Aug 19 15:48:47 1997	3	Natural Language Processing	3
51	Wed Aug 20 10:05:09 1997	15	Fuzzy Logic	18

Table G.11 - continued

rec_id	time_acc	s_aiuse	s_query	s_queuse
52	Wed Aug 27 16:16:22 1997	20	Distributed AI	20
53	Fri Aug 29 09:52:58 1997	9	Genetic Algorithms	3
54	Tue Sep 2 18:08:26 1997	14	Knowledge-based Systems	12
55	Thu Sep 4 16:35:53 1997	18	Neural Networks	18
56	Thu Sep 4 23:35:01 1997	18	Neural Networks	18
57	Fri Sep 5 09:12:53 1997	2	Neural Networks	3
58	Fri Sep 5 15:35:57 1997	1	Natural Language Processing	1
59	Fri Sep 5 16:54:55 1997	20	Knowledge-based systems	18
60	Fri Sep 5 17:06:00 1997	3	Neural Networks	3
61	Fri Sep 5 17:06:34 1997	3	Neural Networks	3
62	Fri Sep 5 18:09:22 1997	12	Distributed AI	14
63	Fri Sep 5 18:13:45 1997	11	Genetic Algorithms	6
64	Fri Sep 5 19:36:32 1997	15	Knowledge-based Systems	16
65	Fri Sep 5 20:04:03 1997	15	Distributed AI	8
66	Fri Sep 5 21:09:59 1997	12	Distributed AI	15
67	Fri Sep 5 22:14:26 1997	16	Natural Language Processing	13
68	Sat Sep 6 16:49:36 1997	17	Distributed AI	17
69	Sun Sep 7 13:38:09 1997	13	Neural Networks	8
70	Sun Sep 7 23:30:57 1997	1	Knowledge-based Systems	7
71	Mon Sep 8 04:42:14 1997	16	Machine Learning	15
72	Mon Sep 8 08:10:35 1997	18	Distributed AI	16
73	Mon Sep 8 08:30:42 1997	17	Distibuted AI	18
74	Mon Sep 8 09:26:55 1997	14	Distributed AI	12
75	Mon Sep 8 15:46:16 1997	14	Machine Learning	13

Table G.11 - continued

rec_id	i_value	s_value	que_mod	unmod_s
1	12.34	3.36	C	Knowledge-based Systems
2	15.78	9.61	E	Knowledge-based Systems
3	16.56	2.73	S	Semantic Networks
4	12.19	2.66	C	Fuzzy Logic
5	8.75	2.66	C	Fuzzy Logic
6	16.56	14.53	G	Knowledge-based Systems
7	8.28	2.66	C	Knowledge-based Systems
8	16.56	13.44	E	Fuzzy Logic
9	16.56	2.66	S	Knowledge-based Systems
10	15.55	2.66	S	Neural Networks
11	None	None	None	None
12	16.56	2.66	S	Bayesian Networks
13	7.58	2.66	C	Knowledge-based Systems
14	16.56	7.34	E	Knowledge-based Systems
15	16.56	15	G	Genetic Algorithms
16	16.56	2.66	S	Fuzzy Logic
17	16.48	12.5	E	Neural Networks
18	15.47	5.31	S	Fuzzy Logic
19	None	None	None	None
20	16.56	2.66	S	Neural Networks
21	13.59	2.66	C	Fuzzy Logic
22	13.28	2.66	C	Fuzzy Logic
23	16.56	14.22	G	Knowledge-based Systems
24	16.56	14.22	G	Distributed AI
25	16.56	9.92	E	Knowledge-based Systems
26	16.56	15.63	G	Semantic Networks
27	16.56	2.73	E	Neural Networks
28	16.56	9.92	E	Natural Language Processing
29	16.56	7.34	E	Genetic Algorithms
30	16.56	17.27	G	Knowledge-based Systems
31	16.56	5.16	S	Bayesian Networks
32	16.56	7.34	E	Machine Learning
33	16.56	17.27	G	Natural Language Processing
34	16.56	7.42	E	Neural Networks
35	16.56	10.39	E	Machine Learning
36	None	None	None	None
37	16.56	10.86	E	Distributed AI
38	16.56	2.66	S	AI Programming Languages
39	16.56	12.42	E	Distributed AI
40	15.7	15.23	G	Distributed AI
41	16.33	9.92	E	Distributed AI
42	12.73	17.27	G	Natural Language Processing
43	16.56	14.3	G	Distibuted AI
44	16.56	12.5	E	AI Programming Languages
45	16.56	11.64	E	Natural Language Processsing
46	16.56	15.86	G	Neural Networks
47	16.56	8.28	E	Machine Learning
48	16.56	15.86	G	Neural Networks
49	8.52	2.66	C	Knowledge-based Systems
50	12.73	3.98	C	Natural Language Processing
51	16.56	16.88	G	Fuzzy Logic

Table G.11 - continued

rec_id	i_value	s_value	que_mod	unmod_s
52	16.17	17.27	G	Distributed AI
53	16.56	4.22	S	Genetic Algorithms
54	16.56	11.33	E	Knowledge-based Systems
55	16.56	17.19	G	Neural Networks
56	16.56	17.19	G	Neural Networks
57	16.56	3.98	S	Neural Networks
58	16.56	2.66	S	Natural Language Processing
59	15.08	17.19	G	Knowledge-based systems
60	16.56	3.98	S	Neural Networks
61	16.56	3.98	S	Neural Networks
62	16.56	12.5	E	Distributed AI
63	16.56	7.34	E	Genetic Algorithms
64	16.17	14.22	G	Knowledge-based Systems
65	16.56	8.83	E	Distributed AI
66	14:14	13:59	E	Distributed AI
67	16.56	11.48	E	Natural Language Processing
68	16.56	15.86	G	Distributed AI
69	8.28	8.83	E	Neural Networks
70	12.66	8.36	E	Knowledge-based Systems
71	16.17	13.44	E	Machine Learning
72	16.56	14.53	G	Distributed AI
73	16.56	17.19	G	Distibuted AI
74	11.8	11.33	E	Distributed AI
75	14.77	11.95	E	Machine Learning

Table G.11 - continued

rec_id	mod_s	f_time	bumr	2bumr
1	Expert Systems	None		
2		None		
3		None		
4	Fuzzy Sets	None		
5	Fuzzy Sets	None		
6		None		
7	Expert Systems	None		
8		Fri Nov 29 15:55:02 1996	4	10
9		None		
10	Error in software	None		
11	Error in software?			
12		Tue Jan 7 18:08:59 1997	2	7
13	Expert Systems	None		
14		None		
15		None		
16		None		
17		Thu Feb 6 13:28:12 1997	6	8
18		Fri Feb 7 11:37:15 1997	4	10
19	None	None		
20		None		
21	Fuzzy Sets	None		
22	Fuzzy Sets	None		
23		None		
24		None		
25		None		
26		None		
27		Unclear Feedback		
28		None		
29		None		
30		Tue Mar 25 08:04:50 1997		
31		Tue Aug 5 00:33:31 1997	1	6
32		None		
33		Thu Aug 7 22:16:40 1997	9	
34		None		
35		None		
36	None	None		
37		None		
38		None		
39		Fri Aug 8 08:59:11 1997	6	8
40		None		
41		Unclear feedback		
42		None		
43		Fri Aug 8 13:14:13 1997	1	5
44		Fri Aug 8 15:16:57 1997	1	10
45		Unclear Feedback		
46		None		
47		Sun Aug 10 23:10:44 1997	6	
48		None		
49	Expert Systems	None		
50	Natural Language Systems	None		
51		None		

Table G.11 – continued

rec_id	mod_s	f_time	bumr	2bumr
52		None		
53		None		
54				
55		None		
56		Thu 4 Sep 23:41:17 1997	8	
57		None		
58		None		
59		None		
60		None		
61		None		
62		None		
63		Fri Sep 5 18:49:48 1997	9	
64		None		
65		None		
66		None		
67		None		
68				
69		None		
70		None		
71		None		
72		None		
73			5	10
74		None		
75		Mon Sep 8 16:14:03 1997	2	4

Table G.11 - continued

rec_id	bmr	2bmr	wumr	2wumr	wmr	2wmr	bumr_sc	2bumr_sc
1								
2								
3								
4								
5								
6								
7								
8	2	10	3	5	6	9	3	5
9								
10								
11								
12	1	2	3	4	3	4	9	8
13								
14								
15								
16								
17	5	7	1	7	2	3	6	2
18	1	4	1	9	6	7	8	7
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31	4	6	4	5	7	8	8	6
32								
33	5		2		1		8	8
34								
35								
36								
37								
38								
39	2	8	1	4	3	5	7	4
40								
41								
42								
43	3	6	3	7	2	4	7	4
44	1	2	3	6	7	10	10	9
45								
46								
47	7		2		3		6	3
48								
49								
50								
51								

Table G.11 – continued

rec_id	bmr	2bmr	wumr	2wumr	wmr	2wmr	bumr_sc	2bumr_sc
52								
53								
54								
55								
56	9						9	9
57								
58								
59								
60								
61								
62								
63	9		2		1	2	9	9
64								
65								
66								
67								
68								
69								
70								
71								
72								
73	7	8	3	7	5	9	10	5
74								
75	2	6	1	3	1	3	8	7

Table G.11 - continued

rec_id	bmr_sc	2bmr_sc	wumr_sc	2wumr_sc	wmr_sc	2wmr_sc
1						
2						
3						
4						
5						
6						
7						
8	5	8	1	1	1	1
9						
10						
11						
12	9	7	5	5	3	5
13						
14						
15						
16						
17	9	3	1	1	1	1
18	10	9	3	2	5	3
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31	8	6	2	1	1	1
32						
33	5	5	1	1	1	1
34						
35						
36						
37						
38						
39	6	5	2	1	3	1
40						
41						
42						
43	7	4	1	1	1	2
44	10	10	6	6	7	7
45						
46						
47	8	6	1	2	2	2
48						
49						
50						
51						

Table G.11 – continued

rec_id	bmr_sc	2bmr_sc	wumr_sc	2wumr_sc	wmr_sc	2wmr_sc
52						
53						
54						
55						
56	10	10				
57						
58						
59						
60						
61						
62						
63	9	9	6	6	7	7
64						
65						
66						
67						
68						
69						
70						
71						
72						
73	8	7	3	7	5	9
74						
75	7	7	1	1	3	2

Table G.11 – continued

rec_id	rcomm	scomm
1		
2		
3		
4		
5		
6		
7		
8	Modified Results Better	Easy to use, suggests changes
9		
10		
11		
12	Unmodified Results Better	Expresses no opinion
13		
14		
15		
16		
17	Modified Results Better	Well designed, Context crucial
18	Modified Results Better	Raises Network problems
19		
20		
21		
22		
23		
24		
25		
26		
27	Modified Results Better	Suggests changes to design
28		
29		
30	Modified Results Better	Well designed, suggests changes
31	Unmodified Results Better	Raises points about function
32		
33	None	None
34		
35		
36		
37		
38		
39	None	None
40		
41	Response Unclear	Response Unclear
42		
43	Modified Results Better	None
44	Modified Results Better	Comments on 20-point scale
45	Similar Results	Easy to Use, too in control
46		
47	Modified Results Better	Easy to Use, in part unclear
48		
49		
50		
51		

Table G.11 - continued

rec_id	rcomm	scomm
52		
53		
54		
55		
56	Modified Results better	Excellent
57		
58		
59		
60		
61		
62		
63	results pretty good for both	Nice, suggests running off-line
64		
65		
66		
67		
68		
69		
70		
71		
72		
73	no set better	not easy to use, suggests changes
74		
75	not sure which results better	will serve as aid to novices

Table G.11 - continued

Mod_cat	av_bum	av_bm	av_wum	av_wm	unmod_s
High-Low	8.5	8	5	4	Bayesian Networks
	7.5	9.5	2.5	4	Fuzzy Logic
	7	7	1.5	1	Bayesian Networks
	7.66667	8.1667	3	3	**(Means of averages)**
High-Med	4	4.5	1	1	Fuzzy Logic
	4	6	1	1	Neural Networks
	5.5	5.5	1.5	2	Distributed AI
	9.5	10	6	7	AI Programming Languages
	4.5	7	1.5	2	Machine Learning
	9	9	6	7	Genetic Algorithms
	7.5	7	1	2.5	Machine Learning
	6.28571	7	2.571429	3.2143	**(Means of averages)**
High-High	8	5	1	1	Natural Language Processing
	5.5	5.5	1	1.5	Distibuted AI
	9	10	none	none	Neural Networks
	7.5	7.5	5	7	Distibuted AI
	7.5	7	2.333333	3.1667	**(Means of averages)**

Table G.12: Mean Scores Versus Modification Category (Ad-hoc Phase) - Also Gives Means of Means

unmod_s	av_bum	av_bm	av_wum	av_wm	Mod_cat
AI Programming Languages	9.5	10	6	7	High-Med
Natural Language Processing	8	5	1	1	High-High
Genetic Algorithms	9	9	6	7	High-Med
Machine Learning	4.5	7	1.5	2	High-Med
	7.5	7	1	2.5	High-Med
Neural Networks	4	6	1	1	High-Med
	9	10	none	none	High-High
Fuzzy Logic	7.5	9.5	2.5	4	High-Low
	4	4.5	1	1	High-Med
Bayesian Networks	8.5	8	5	4	High-Low
	7	7	1.5	1	High-Low
Distributed AI	5.5	5.5	1.5	2	High-Med
	5.5	5.5	1	1.5	High-High
	7.5	7.5	5	7	High-High

Table G.13: Means of Scores Versus Original Query (Ad-hoc Phase)

rec_id	orig_id	que_mod	av_bum	av_bm	av_wum	av_wm	
12	55	S	8.5	8	5	4	
18	61	S	7.5	9.5	2.5	4	
31	75	S	7	7	1.5	1	
			7.666667	8.166667	3	3	**(Means of Averages)**
8	45	E	4	4.5	1	1	
17	60	E	4	6	1	1	
39	83	E	5.5	5.5	1.5	2	
44	88	E	9.5	10	6	7	
47	91	E	4.5	7	1.5	2	
63	107	E	9	9	6	7	
75	119	E	7.5	7	1	2.5	
			6.285714	7	2.571429	3.214286	**(Means of Averages)**
33	77	G	8	5	1	1	
43	87	G	5.5	5.5	1	1.5	
56	100	G	9	10	none	none	
73	117	G	7.5	7.5	5	7	
			7.5	7	2.333333	3.166667	**(Means of Averages)**

Table G.14: Means of Scores Versus Query Modification Category (Ad-Hoc Phase) - Also Gives Means of Means.

Appendix H - Wilcoxon tests

The following test results are the outcomes of applying Wilcoxon's test (see chapter 8, page 148) to the paired datasets for both the ad-hoc and controlled phases of the user study. Recall that the paired datasets are derived from comparing the best, 2nd best, worst and 2nd worst scores for the modified and unmodified results as rated by the users. The number of the test refers to its position in the list of tests performed upon the results, and is not significant.

Test 8: Wilcoxon's test on the best scores for the modified and unmodified query results from the controlled phase.

Description: See chapter 8, page 146.

Result: T = 20. From tables, the degrees of freedom (dof) are 15 and the T-value, at the level of 5% significance is, **T (5%)** = 25.

Therefore, in this case the paired data are significantly different to the level of 5%.

Test 9: Wilcoxon's test on the 2nd best scores for the modified and unmodified query results from the controlled phase

Description: Identical test as test 8.

Result: T = 24. In this case the dof are 18 and the T-value from tables is **T (5%)** = 40.

In this the case the paired data are significantly different to the level of 5%.

Test 10: Wilcoxon's test on the worst scores for the modified and unmodified query results from the controlled phase

Description: Identical test as test 8.

Result: T = 41. The dof are 18 and therefore **T (5%)** = 40.

Therefore for this set of paired data, there is no significance at the 5% level.

Test 11: Wilcoxon's test on the 2nd worst scores for the modified and unmodified query results from the controlled phase

Description: Identical test as test 8.

Result: T = 8. The dof are 12 and from tables **T (5%)** = 13.

This result means that this data exhibits significance at the 5% level.

Test 12: Wilcoxon's test on the best scores for the modified and unmodified query results from the ad-hoc phase

Description: Identical test as test 8

Result: T = 20. From tables, the degrees of freedom (dof) are 9 and the T-value, at the level of 5% significance is, **T (5%)** = 5.

Therefore, there is no significant difference between the paired data.

Test 13: Wilcoxon's test on the 2nd best scores for the modified and unmodified query rsults from the ad-hoc phase

Description: Identical test as test 8.

Result: T = 12. In this case the dof are 10 and the T-value from tables is **T (5%)** = 8.

Again, the result is not significant.

Test 14: Wilcoxon's test on the worst scores for the modified and unmodified query results from the ad-hoc phase

Description: Identical test as test 8.

Result: **T** = 3. The dof are 8 and therefore **T (5%)** = 3.

Therefore for this set of paired data, there is significance at the 5% level.

Test 15: Wilcoxon's test on the 2nd worst scores for the modified and unmodified query results from the ad-hoc phase

Description: Identical test as test 8.

Result: **T** = 5.5. The dof are 6 and from tables **T (5%)** = 0.

This result means that this data exhibits no significance.

Appendix I - The development and implementation of the FMQA software

I.1 Overview

In chapter 6 (sections 6.2,6.3 and 6.4) the development of the software which drove the FMQA is discussed. In this appendix, the FMQA software is examined in more detail. As stated in chapter 6 software was needed to provide various parts of the functionality of the FMQA, namely:

- the creation and representation of the fuzzy set models and the FKB
- the capture and use of the questionnaire data to modify the fuzzy sets
- the capture and representation of the user query
- the defuzzification of the adapted fuzzy sets and the modification of the user query
- the display of any information to the user during interaction with the prototype
- the submission of the query to Lycos™.

The implementation of software to provide this functionality involved the combined use of HTML (simple and form-driven), cgi-scripts, hybrid C/C++ programs and the use of proprietary scripts. Each part of the software is discussed in turn below but figure I.1 on the next page shows how the software is employed in the different combinations to deliver the different parts of the FMQA shown in figure 6.1.

The rest of this appendix concentrates mainly upon examining the development of the C/C++ programs which drive the FUM process as the FMQA HTML interface is covered in chapter 6 and Appendix D, though the interface is briefly discussed next in section I.2.

I.2 The HTML Interface

As is discussed in some detail in section 6.3 in chapter 6, the FMQA interface was developed using HTML so that it could be positioned easily on the WWW and thus easily accessible to potential users. Another reason for creating a WWW interface was to avoid any potential bias which might be associated with the user having to learn a new type of interface. The interface screens are shown in chapter 6, and Appendices B and D. A special type of HTML page was employed within the FMQA in order to capture information via the user interaction – this a form page

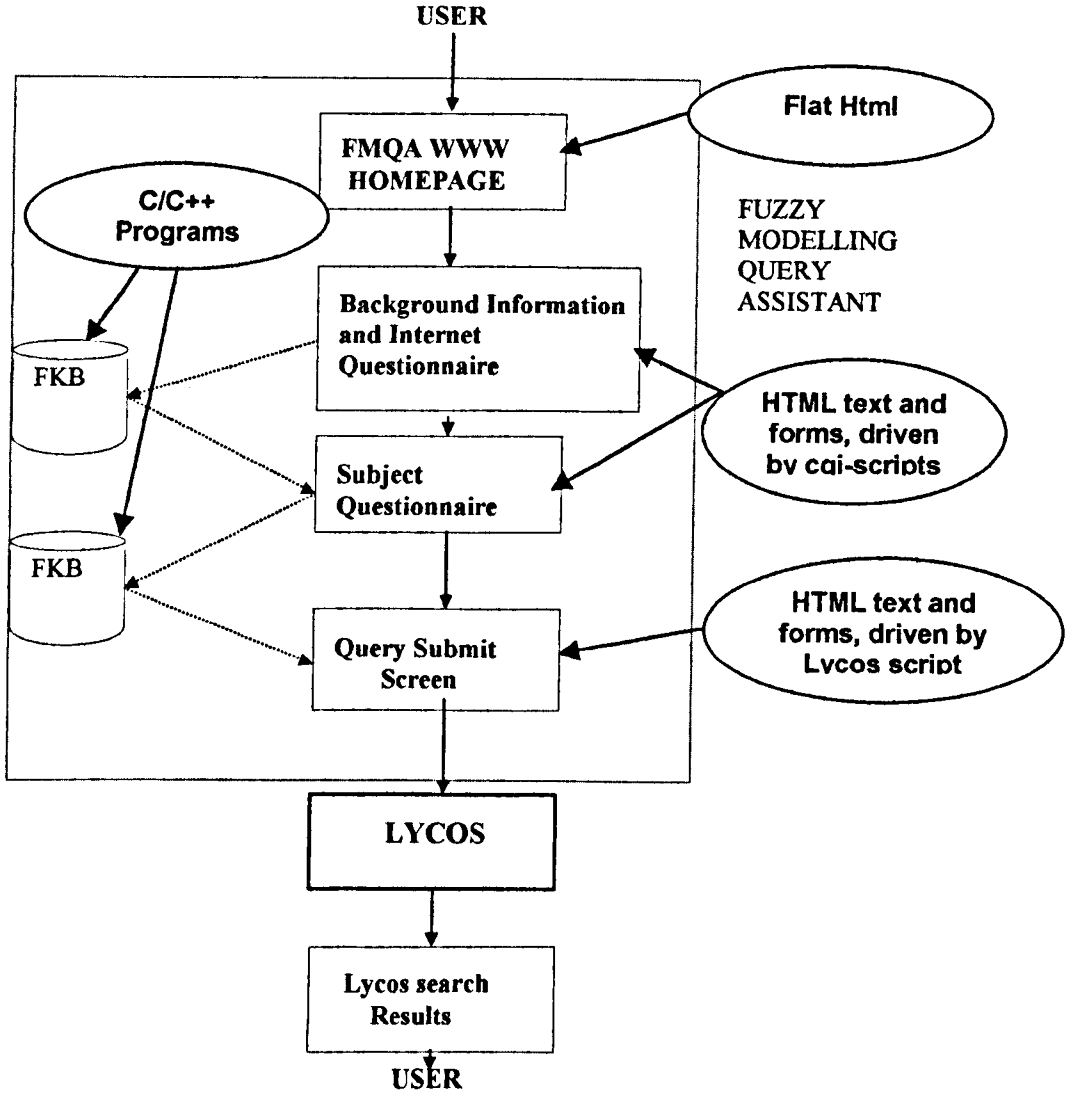


Figure I.1 High-level functional diagram of the prototype FMQA (showing software used).

A HTML form is basically the same as a paper form but instead of filling in boxes or answering questions on paper, a user is able to do this on-line whilst interacting with the system. The interface form-based pages call small programs called cgi-scripts to capture and process the information entered, in the case of the FMQA, in the two questionnaires. These scripts can be written so that other programs are executed, which was the case with the FMQA, as the scripts used by the forms not only stored the questionnaire answers in a logfile but also passed the information to the C/C++ programs which performed the FUM process. These programs are discussed next in section I.3.

I.3 The FUM Software

In chapter 6, section 6.2, the development of the software to implement the FUM approach is discussed. In that section, it is noted that the software developed was based on the C/C++ programs developed by Cox (1994) to implement fuzzy systems. Also described in that chapter is how a membership function is represented using Cox's approach. Basing the implementation of the FUM approach on Cox's software was ideal because he provides a complete set of tools for building fuzzy systems. His programs concentrate upon the construction of fuzzy IF..THEN rules and the dealing with linguistic variables, both aspects making them suitable for use in the development of the FMQA. However, though chapter 6 touches on the development of the FKB, it does not emphasise just how big an extension and reengineering of Cox's software is represented by the development of the FMQA. In this section of Appendix I, this expansion and implementation is focused upon.

The initial problem to overcome in using Cox's programs was to reengineer them so that would work on another platform. Cox had only developed and tested his software using Borland C and thus a DOS-based operating system (OS) platform. The programs which drive the FMQA are written in UNIX-based C/C++ and were implemented upon a UNIX OS platform. This reengineering and reimplementation took considerable resources in terms of time and author effort. Cox's software is modular in design and the reengineering involved testing not only that each different module would work on the new platform but then slowly grouping modules together and testing the groups before building any fuzzy sets and membership functions with the reengineered software.

Once the basic kernel of Cox's software was implemented upon the UNIX platform and the modules worked, the development of the FMQA could be begin. As described in section 6.2 of chapter 6, the Fuzzy Associative Memory (FAM) approach (Kosko, 1990) was adopted to develop the rules which constitute the FKB. An example of a FAM was given in chapter 5 of the thesis and the figure is repeated below in figure I.2. In the FAM technique, there are two premise variables and one consequent or solution variable. The premise variables ranges are represented in the FAM by the horizontal and vertical axes, the solution variable possible values by the intersection points. In this way, the FKB rules which combine the answers to the questions were constructed.

ans5 \ ans 6	L	M	H
L	VL	M	M
M	L	M	H
H	M	M	VH

Figure I.2 Fuzzy Associative Memory (FAM) matrix for questions 5 and 6 of the WWW questionnaire. (reproduction of figure 5.6).

Although Cox discusses the FAM approach to developing fuzzy systems he does not provide any example fuzzy systems using the approach in the software and only discusses the approach briefly. In fact, in the example fuzzy systems offered there exist a maximum of 5 fuzzy rules. In the FMQA implementation, using the FAM approach, over 100 rules were developed. A portion of these rules, showing the nine rules produced by the FAM given in figure I.2, is shown in figure I.3.

Thus, the 100+ rules which constitute the FKB represent a significant extended implementation and development of Cox's software and a major development of a fuzzy system using the FAM approach.

Amongst Cox's software suite, there is a module which provides for standard defuzzification techniques to be applied to any final fuzzy membership functions. This software was employed to defuzzify the two final fuzzy membership functions produced by the applying the FKB to the questionnaires answers, but in order to interpret the answers and use them to modify the original query, author-written software was needed.

```
R1    if ans5 is low and ans6 is low
      then exp/time is very low
R2    if ans5 is low and ans6 is medium
      then exp/time is medium
R3    if ans5 is low and ans6 is high
      then exp/time is medium.
R4    if ans5 is medium and ans6 is low
      then exp/time is low
R5    if ans5 is medium and ans6 is medium
      then exp/time is medium
R6    if ans5 is medium and ans6 is high
      then exp/time is high.
R7    if ans5 is high and ans6 is low
      then exp/time is medium
R8    if ans5 is high and ans6 is medium
      then exp/time is medium
R9    if ans5 is high and ans6 is high
      then exp/time is very high.
```

Figure I.3: The Nine Rules in the FKB which Combine the answers to questions 5 and 6 of the Internet Questionnaire

The first program written was a small knowledge base which categorised the defuzzified values within a range and then produced the query modification category by combining the two range values. The defuzzified values are categorised as being in a ***Low, Medium*** or ***High*** range and then two ranges can be combined to produce the four different query modification categories ***Contracted, Slight, Extra*** or ***Great***. The rules of this knowledge base are reproduced in figure I.4. These rules are encoded within a C program which then uses the query modification category in conjunction with the user query to produce the modified query. The modification extent depends upon the choice of original query and the query modification category as discussed in chapter 6. The FMQA user is oblivious to the software which drives the system and after completing the questionnaires is next presented with the query submission page. The software behind this page is discussed next in section I.4.

```
        if subject_value =< 6.00 then
              subject_range is LOW
        if subject_value > 6.00 or < 14.00
              subject_range is MEDIUM
        if subject_value >= 14.00 then
              subject_range is HIGH
        if internet_value =< 6.00 then
              internet_range is LOW
        if internet_value > 6.00 or < 14.00
              internet_range is MEDIUM
        if internet_value >= 14.00 then
              internet_range is HIGH
  if subject_range is HIGH and internet_range is HIGH
              then query_cat is G
  if subject_range is HIGH and internet_range is MEDIUM
              then query_cat is G
  if subject_range is HIGH and internet_range is LOW
              then query_cat is E
  if subject_range is MEDIUM and internet_range is HIGH
              then query_cat is E
  if subject_range is MEDIUM and internet_range is MEDIUM
              then query_cat is E
  if subject_range is MEDIUM and internet_range is LOW
              then query_cat is S
  if subject_range is LOW and internet_range is HIGH
              then query_cat is S
  if subject_range is LOW and internet_range is MEDIUM
              then query_cat is C
  if subject_range is LOW and internet_range is LOW
              then query_cat is C
```

Figure I.4: The Knowledge-base Rules Used to Interpret the Defuzzified Numbers

I.4 Query Submission

After the FUM process, the defuzzification and the interpretation of the values, the next step in the software implementation is the query submission process. This is fairly easy as the final output from the software programs described in the previous section is two query strings. These are presented to the user as hidden values in a HTML form-based WWW page (see Appendix D for a reproduction of an example). The user is able to submit either the original or modified query by submitting the form.

On submitting the form, the query string is passed to a proprietary Lycos search script. Control of the script and the format of the IR results returned is outside the scope of the software of the FMQA. The user is asked to return to the query submission page when they have finished browsing and rating the results so that they can access the email response form. This HTML form and the sofware behind it is discussed in section I.5.

I.5 Email Response Form

The last element of the FMQA software is the email response form. This is again a HTML form-based WWW page which allows the use to enter the scores for the IR results and make comments about the performance of the FMQA. The users were asked to do this at the end of their interaction with the FMQA. The form interface was given in chapter 7, figure 7.1, pages 116-117. When the user submits the form, the information contained is passed to a special type of cgi-script which formats the information and sends it to a specified email address. This address is specified in advance and was in the author's university address. In this manner the feedback and user study results were collected.

The final section of this appendix summarises the FMQA software and development.

I.6 Conclusions

The software developed and implemented to build the FMQA was a major undertaking required in order to test the research hypothesis and satisfy the aims and goals of project. Though based on existing software, the actual implementation represented a significant extension and reengineering of said software, involving new OS platforms and many more fuzzy rules than had been provided in the software suite example systems. The FMQA software also included author written programs to interpret and use the defuzzified values from the FUM process and a number of cgi-scripts to capture and use information gained via the FMQA interface. The interface itself was an integrated part of the FMQA software.

The FMQA software was designed and implemented to be as transparent and seamless as possible so as to avoid any bias in the study results from any user-perceived notion of software functionality. The implementation took considerable time and effort but in terms of performing the FUM process as envisaged in chapter 5, and collecting user feedback as described in chapter 7, it was a success.

Appendix J – The Email Questionnaire

As part of the process of finding suitable subjects to take part in the user study, an email questionnaire was sent to prospective subjects – undergraduate students studying within the School of Computing and Mathematical Sciences – to elicit whether certain cohorts of students would have the correct level of experience/knowledge of the Internet and the AI domain. The questionnaire is reproduced below.

Dear Student

I am a postgraduate student within the School of Computing Science researching information retrieval from the WWW. As part of this research I need to elicit the experience and knowledge of the Internet and the subject domain of the AI of students within the School. I would be very grateful if you could complete the short questionnaire below and return it to me at gjmooney @dmu.ac.uk. You can do this by pressing reply within your email software and then completing the questionnaire by inserting an 'X:' in the appropriate place. Thank you for your help.

yours sincerely

Gabrielle Mooney

Questionnaire follows

****

1. Have you ever used the Internet before? Yes [ ] No []

2. If answer to question 1 is Yes, how long have you been using the Internet?
less than 3 months [ ] 3-12 months [ ] more than 12 months [ ]

3. How often per week do you use the Internet?
never [ ] less than 5 hours [ ] 5-15 hours [ ] 15 hours+ [ ]

4 Do you find the Internet, in terms of difficulty to grasp and use
simple [ ] moderate [ ] easy [ ] n/a [ ]

5. Do you have any experience of the subject of Artificial Intelligence (AI)? Yes [ ] No [ ]

6.. If answer to question 5 is yes, how much experience of AI do you have?

very little [ ] a little [ ] some [ ] a lot [ ]

****