Modelling the IR task: supporting the user

Mark Ennis

Submitted for Examination of Doctor of Philosophy

Centre for Human-Computer Interface Design,
School of Informatics,
City University,
London.

October 1998

Table of Contents

Page

Thesis tables

The thesis has the following tables:

Thesis Figures

The thesis has the following figures:

Acknowledgements

I would like to thank my family, colleagues and friends for the support and assistance provided during this work. I would also like to thank my supervisor Professor Alistair Sutcliffe for his guidance throughout my PhD

Declaration

The author grants powers of discretion to the University Library to allow this thesis to be copied in whole or in part without further reference to the author.

Statement of Contribution

This disclaimer is to state that the research reported in this thesis is primarily the work of the author and was undertaken as part of his doctoral research. Referenced papers of which the student is not the sole author represent the role of the supervisor in the research, to direct the work and enhance the written style of these papers.

Abstract

The need for improved interface designs and targeted task support were highlighted based on performance data gathered in an empirical evaluation of a typical commercial database: MEDLINE[1]. The thesis outlines theoretical work addressing the cognitive activity underlying searchers behaviour in the information retrieval task and unifies this into a cognitive task model of their information seeking behaviour. The motivation for this work is to understand; explain and minimise the mismatch between information retrieval interfaces, and their associated functionality, and users' mental models of the search process. The cognitive model consists of IR processes, cognitive activities, correspondence rules, knowledge sources necessary to support the task, strategies, tactics and IR systems representations. The cognitive model is related to specific systems through taxonomies of possible functionality. The model is evaluated by comparing empirical observations of users' information retrieval against the models predictions of behaviour for a specific situation. Design implications attempt to link the model of human action to what users need to know and the task support provided by an IR system to indicate and justify design requirements. In this way the research attempts to define the facilities which support user process. A concept demonstrator is presented with associated usability studies. The effectiveness and limitations of this approach to the systematic design and evaluation of IR systems are discussed.

[1] **MEDLINE is the copyright of the U.S. National Library of Medicine and SilverPlatter**

Chapter 1

Overview

This chapter introduces the problem and gives an outline of the thesis

Chapter 1 : Overview

1.1 Introduction

Information retrieval systems provide users with the tools and facilities necessary to access information from large repositories of material. The advent of the World Wide Web has extended the scope and distributed nature of these information repositories and created a new need for configurable and adaptable support from systems. The broadening of the scope and quantity of information dealt with requires user retrieval to be effective and efficient if all the information needed is to be accessed within a reasonable time frame. If this is not achieved it will lead to user dissatisfaction and 'information overload' (Hiltz 1985). To enable this it is necessary to develop a greater understanding of the user if systems are to improve and support the tasks to be performed.

The need for more effective designs for information retrieval systems have been accentuated by the move away from professional search intermediaries towards end-user searching. At present the gap between the facilities and functionality of IR systems and users' needs is unacceptable (Dervin 1977 and Borgman 1985). The failure of IR systems to meet user requirements stems from the excessive gap between systems functionality and users' mental models of the search process. The outcome of this is that the searcher faces extra cognitive demands in expressing information needs causing them to be inefficient in the time and resources spent getting to the information required. The function of an IR system should be to reduce the user effort associated with the satisfaction of an information need, not to impose a greater work load on the user. The research of Marchionini and Liebscher (1991) has shown that users associate extra effort with using electronic media compared with paper based information. The result is that searchers often ignore the functionality and facilities offered by electronic systems in the search for the path of least cognitive resistance, rather than using the most effective method of searching and retrieval. It is therefore apparent that at some level search systems are failing users requirements.

- Do these issues impact on users' ability to retrieve relevant information?
- Can the effects caused by user differences be minimised by alterations in design and by tailoring design to support the activities of groups or categories of users?

- What are the optimal configurations of systems functionality and interfaces if user performance is to be optimised?

Innovative designs have been proposed for information searching user interfaces (e.g. Alhberg and Shneiderman 1994, Card et al 1991); however, few evaluations have been carried out to demonstrate the usability of such systems. Advanced information retrieval systems (e.g. OKAPI Robertson et al 1994) have shown improved performance measured in terms of recall[1] and precision[2] (Salton 1992), but these performance evaluations do not identify the contributions of different aspects of the design to improved retrieval (Marchionini 1995) or indicate how user performance can be improved. For instance, improved search algorithms or the task support facilities provided may be responsible; alternatively, user interface features such as different approaches to relevance feedback may be the major contributor to success. Hence there is a need for a more systematic approach to the design and evaluation of information retrieval systems and their user interfaces.

Theories of searcher behaviour available in the IR literature and discussed in detail in chapter two, focus on observable actions [Bates (1979, 1989); Markey and Atherton (1978); Marchionni (1995) and Belkin (1982, 1993)]; the categories of information used in search term selection (Allen 1991, 1994; Michel 1994); the communication of the information need to a professional search intermediary (Ingwersen 1982) or a cognitive viewpoint of IR (Kulthau 1988, 1993; Ingwersen 1996). These models offer insight into the complexities of retrieval activity but they are only able to model activity in a specific context and thus have limited applicability for IR design. The information system, the information presentation metaphor and the relevance of the results all influence the user's behaviour but unfortunately current models of the IR task do not account for this. The design of current systems is based on an algorithmic view of the search process rather than a user-centred focus (Ingwersen 1993, Marchionini 1995). Furthermore, many systems do not accommodate the different user information searching tasks such as browsing, goal directed searching, using rapid query refinement and high level evaluation judgements to adjust the results

[1] Recall reflects the proportion of relevant items retrieved in an answer to a search request
[2] Precision reflects the proportion of the retrieved items that are relevant

produced, sub-goaling etc. (Marchionini 1993); and there are no means of deciding how these different types of search are to be supported or determining which type of search a user will perform in a given context. These factors make current models of IR unsuitable as the basis for designing intelligent systems support, because although they highlight interesting issues about retrieval behaviour, they are incomplete and fail to predict or explain user behaviour in different search tasks.

HCI as a discipline has contributed extensively to improving the design of IR systems interfaces by investigating the usability of different query and results presentation metaphors. The development of the new interaction metaphors have focused on advances in computer technology (e.g. Thumbnails 3D browsing metaphor, Elvins 1998) rather than supporting the user. The visualisation of information structures (Furnas 1995, 1997; Pirolli et al 1996; Benyon et al 1997) aid user browsing but this type of search activity is only a narrow facet of the retrieval task. Alternatively the different interface approaches for querying such as dynamic filters (Ahlberg and Shneiderman 1994, Fiskin et al 1995) or query by sketching (Charles et al 1990) provide alternative access routes to information, but their applicability in different retrieval contexts has not been investigated. It is not known when a particular interaction technique or presentation metaphor enhances retrieval; since to establish this, a firm understanding of the IR task and the influence of the system on activity is necessary. HCI has not yet focused on the determination of the individual facets of user retrieval behaviour (i.e. process, strategies, tactics, and user-system dialogue) or how the effectiveness of these are influenced by the types of tasks performed. To understand why certain metaphors, functions and design features are more successful than others the effects of different design implementations on user success must be established. To provide adaptable, supportive and configurable systems which aid the activity of users, a detailed and comprehensive understanding of what triggers and determines the course of retrieval behaviour is required.

In summary the purpose of this thesis is to investigate the different strategies, tactics and activities requiring systems support for effective retrieval behaviour and the context in which these are applicable. The thesis attempts to link user behaviour to specific systems design decisions to predict the success of the interaction.

The remainder of this chapter reviews the components of the information seeking process, issues specifically relating to information retrieval systems design and the general use of models to enhance computer user interface designs. The chapter concludes with a summary of how this thesis attempts to extend the understanding of retrieval tasks through cognitive modelling and empirical investigations. This understanding should be used to specify the different types of advice and task support required to minimise the gap between users' mental models of systems and the functionality provided.

1.2 Information seeking

Information seeking occurs in many aspects of our lives in which information is interpreted, opinions and beliefs are formed about our environment, and information is used to make decisions e.g. when choosing which bank to use we may seek information concerning interest rates. The 'explosion' of information sources (Davidson 1996) provided by computer repositories and the diversity of information seekers has led to new issues relating to 'information overload' (Hiltz et al 1985, Iselin 1989, Quarterman 1990, Sharpe 1997). Information seeking has undergone a transition from a physical activity performed in a three dimensional environment, e.g. selecting a book from a library shelf and expanding the scope by glancing at adjacent books, to operations on a two dimensional computer interface e.g. using the INSPEC[3] CD-ROM to locate scientific information (Card et al 1991 information visualiser). It is in this computer related context information seeking is examined. Information seeking involves the use of tools to assist in the location of information. Obtaining high recall and precision from the information retrieved depends on the users' ability and the tools provided by computer systems to assist in the locating and filtering of information. Information seeking is often dynamic, iterative and reactive (Rasmussen 1996) with the information retrieved at different stages in the process, and its relevance to the user's need, determining the course of subsequent activity. Information seeking is a high level cognitive process including decisions that the user may not be conscious of (Marchionini 1995) and it has a diverse number of possible

[3] INSPEC is the copyright of IEEE and FULCRUM technologies Inc.

outcomes. The level and type of search activity performed in information seeking is a function of the users' knowledge (Allen 1991, Michel 1994) and context factors (Rouse and Rouse 1984) all of which feed into success or failure. Given that a user's knowledge and context varies the activities performed, consequently the system's support will need to be tailored appropriately.

There has been a transition in IR systems away from the professional intermediary locating information for a third party to end-user searching (Danilowicz 1993). This has occurred due to increases in the quantity of people seeking information from computer systems and the expansion of the number of access points to such systems. This transition has not resulted in a change to a user centred view in the design of IR systems (Marchioninni 1995) hence many systems fail to meet user task requirements. The move to end-user searching and the different situations in which IR systems are used has highlighted that multiple and varied strategies for retrieval success exist. Harter et al (1984) and Bates (1979) produced taxonomies of expert search tactics which can be applicable in different search contexts. This research highlights the extent of the strategic knowledge required for effective searching but unfortunately end-users rarely possess this knowledge. IR research has found that fundamentally different approaches can exist to effective retrieval as embodied in browsing (Herner 1971, Apted 1971) and goal directed searching. Either of these approaches may lead to success yet it is not clear when these approaches are most applicable or which would be most efficient for satisfying users information needs in a specific context. The applicability of tactics and search strategies depend not only on the user's abilities but also upon the facilities provided by the device. The strategies and tactics applicable in any given situation are affected by the information retrieved by previous searches and its relevance (Saracevic 1990, Su 1994, Beaulieu et al 1997). Efthamiades (1993) and Allen (1994) found that the display of information effects the user's perception of the search results and their reaction to this information. The system facilities and the presentation of the information retrieved must be an integral part of models of IR behaviour.

1.3 Information retrieval systems design

Early retrieval systems tended to be designed from a functional perspective providing powerful retrieval algorithms to extract relevant information from the ever increasing volumes of literature (Robertson et al 1994) rather than to support users instinctive retrieval activity. If the effectiveness of IR interfaces are to be improved then they must be designed from a user perspective and support the activities users are trying to perform. The interface and system can then be used to enhance retrieval activity by assisting the user in the construction of queries so expert performances in the retrieval of information may be attained. The end-users of a retrieval system can originate from a variety of educational and situational backgrounds; and so are often left out of the design process for IR systems until late in the software development cycle when decisions over the functionality, facilities and UI to provide have been made. Thus IR experts have designed retrieval systems they perceive as easy to use. However, the diversification of the user population means the probability of the user having the detailed knowledge of the necessary IR strategies and tactics required to make effective use of system functionality is low, and thus users tend to muddle through rather than locate information efficiently. The functionality focus in systems design may require excessive user effort to locate material (Marchionini and Liebscher 1991) and increases the probability much of the relevant information in the system is missed. This is an unacceptable state of affairs as IR systems should help a searcher in their quest for information not hinder it. This has led to a requirement for methods of improving the presentation of IR systems functionality and their access facilities for the variety of users who wish to exploit this technology by moving towards tailoring the advice and functionality offered based on the users knowledge and context (i.e. search history and need type).

Information retrieval generally centres around a reactive dialogue between the information seeker and the information retrieval system. The interaction metaphors and languages used for the retrieval vary with the type of database. SQL and Boolean languages for relational databases allow access through a fixed set of commands which provide control over the location and presentation of material. Bibliographic databases have been designed to support a set of access functions or common commands that are the result of the ANSI Z39.58 and ISO 8777 standards for interactive text searching (Common Command Language - CCL). These standards specify the commands used to access information and the form of responses produced by the system. Information systems based on a probabilistic retrieval model (Robertson and Sparck Jones 1994) use a list of keywords which are then ranked based on a weighting compiled from the following many measures including:

- Collection frequency - Terms which occur in a few documents are likely to be more useful than ones occurring in many. The collection frequency weight is a measure of this.
- Term frequency - The more frequently a term appears in a document the more important it is likely to be for that document. The term frequency is a measure of this.
- Document length - A term that occurs the same number of times in a short document as in a long one is likely to be more important to the short document than it is to the long one. The document length weight is a measure of this.

Each of these weights is combined to give a score for each term-document combination and then each term's score is combined to give a total score for each document which matches the query. The presentation of the different query protocols and the interfaces enable the technical details of these languages and how they operate to be abstracted away from the user, but this is rarely used in SQL interfaces. The specific interfaces provided by the device may hide these languages behind a GUI e.g. Query by example (Zloof 1977). It is unclear how decisions concerning what functionality to include, how the user-system dialogue should be structured and the interface metaphor to be used are made in the design of retrieval systems when they are to provide access to information for a variety of users and tasks. The different interfaces and search languages provide support for certain user strategies making them more effective e.g. rapid query refinement or browsing are supported more efficiently by different interfaces. A user may need to perform multiple strategies in information seeking but their behaviour may be constrained by the facilities provided, so it is important to establish the link between the activities the user tries to perform and the strategies supported by the design. In this way the optimal configuration of facilities can be determined to ensure users are able to satisfy information needs effectively. Lancaster (1977), Timbie et al (1969) and Lancaster et al (1972) found that for both experienced and inexperienced searchers the major problems occur due to ineffective search strategy not the mechanics of the specific system command language, but the command language can be a significant barrier for very inexperienced users. The standards and protocols provided by these query languages can be embedded within more user oriented dialogue enabling a more effective use of the powerful retrieval algorithms. However, this should improve the learnability, effectiveness and user confidence in such systems but issues relating to the facilities to provide to assist browsing and different visualisations of results should not be neglected as these are also fundamental to the success of the retrieval task.

1.4 Models in IR design

The design of interactive systems should embody a model of the user and support the activities pursued based on the tasks encountered (Carroll 1987a, 1987b). Effective information retrieval relies heavily upon a dialogue between the end-user and the system, or person, responsible for extracting and presenting that material. Ingwersen's (1982) communication models show how expert intermediaries (librarians) can act as translators between information retrieval systems and the end-user. The expert intermediary uses an acquisition dialogue to elicit the user's mental model of the information required and adapts this based on their own models of the strategies, tactics, system facilities and applicable information indexing policies. The librarian's knowledge of IR strategy, methods for query articulation and methods for eliciting complete, succinct and precise information needs from the end-user enables effective searching. The elimination of the librarian from the retrieval process owing to the distribution of information access and storage (e.g. World Wide Web and due to pressures on library resources) has caused a gap between the users' mental model of the information required and the system's model of the information and retrieval mechanisms (Marchionini and Liebscher 1991, Danilowicz 1993). Users have not, on the whole, acquired the expert knowledge of librarians in searching which often causes performance to be poor and sporadic. Information seeking often requires an interaction between the search activities which need to be performed to construct a well formed query and the concepts present in the users information need. Mediation is necessary if the information need is to be satisfied e.g. elaboration of concepts in the users need using multiple terms based on synonyms. The identification of differences in search activity for different types of user; and patterns of searching which could be elicited by expert intermediaries are necessary if the system is to take on the librarian's role in the negotiation process of IR. To achieve this mediation, predictive models of users and models of expertise need to be provided to IR designers if they are to ensure the interfaces and functionality incorporated in systems are supporting user activity. A number of models of IR behaviour exist in the literature and these are discussed in chapter 2 but these models have a number of pitfalls. The models tend to focus on narrow contexts for information seeking and their limited scope could cause problems if they were to be used as the basis for intelligent retrieval support. The second problem with the current models of the IR process arises because the models are not tightly linked to aspects of IR systems design and thus it is difficult to predict the effects of systems functions, or behaviour, on user activity or provide context sensitive assistance to users during the retrieval process. The need to include elements of the IR systems design within a user model of retrieval occurs because of the influence of the system dialogue and functionality on query articulation, activities performed and the strategy and tactics used.

User centred design as embodied in (ISO 13407) - Human Centred Design Process for Interactive Systems seems the most appropriate design methodology for information retrieval systems development owing to the key role of users in directing the course of the retrieval task. The approach places special emphasis on the need to develop software which is 'usable, effective, efficient and satisfying to use'. The standard specifies that to achieve this the essential design activities are:

1. understand and specify the context of use for the system. To achieve this the designer must have a detailed understanding of:
    - the characteristics of the intended users (knowledge, skill, experience, education, training etc.).
    - the different tasks the users will perform.
2. specify the user and organisational requirements.
3. produce designs and prototypes.
4. carry out user-based assessment

This design approach aims to increase user productivity (search recall), enhance the quality of work (effectiveness of retrieval activity and search precision), reduce support and training costs by assisting users to perform their tasks and improve user health and safety. An important aspect of user centred design is the determination of which aspects of the task should be performed by the computer or user. However in IR the distinction isn't static due to the adaptive and progressive nature of the retrieval task. The ISO standard provides the general process structure for the design of any system. IR needs to be treated slightly differently to other types of system due to the diversity in the characteristics of information seekers and the tasks they perform. To utilise user centred design techniques a thorough understanding of the variety of users, variety of tasks, variety of strategic approaches, differences in the activities used in the retrieval of information and the effects of the system designs on retrieval tasks are necessary. The interactive, reactive and dynamic nature of information retrieval means that the number of possible functionality and interface combinations to be tested on the different possible task and user types will be combinationally explosive. The necessity to extensively test different designs for the user-system dialogue originates from its dominant effect on the adaptation, structuring and the complexity of the retrieval task (Rasmussen 1996). User centred design methodologies require prototypes to be driven from a firm understanding of the IR tasks performed and user activity requiring support. Thus user centred design dictates a requirement for a thorough understanding of the IR task, the user and types of behaviour performed if it is to be effective; and current models of IR are unable to provide this for the variety of contexts in which information retrieval occurs. To assist in the design of more effective systems and to base design decisions on user process a more complete and adaptable theory of user retrieval behaviour is necessary.

1.5 Summary

- Although models of IR exist they tend only to explain or describe user activity in a specific context. The models don't link the effects of system facilities or information retrieved on user behaviour and the strategies used.
- Information seeking can be performed on remote and distributed information repositories (e.g. WWW) which has seen increases in the number of information

access points and the removal of search intermediaries; causing retrieval to be end-user rather than expert driven. However, users have not acquired the necessary knowledge of retrieval techniques and systems to effectively replace the direction offered by the expert intermediary in the need articulation process.

- IR involves many different types of activities, strategies and tactics. The current views of the IR process are insufficient to specify the context in which these options are applicable or how decisions between the different behaviours are made.
- IR is dynamic, iterative and reactive to the behaviour of the system and the relevance of the information retrieved; and thus any view of the retrieval process must consider the effects of the IR design and system state on the behaviour, activities, strategies and tactics used. This produces a requirement for user centred design techniques to be used but these require a firm understanding of the type of tasks performed and the specification of the behaviour patterns expected from the different types of user given alternative system designs.

These factors lead to a need for a theory that combines a model of the user and the machine to predict successful interaction and to form the basis of context sensitive dialogue which assists effective retrieval.

1.6 Thesis rationale and overview

The many different user search activities requiring systems support have been neglected by IR system designers causing users to be dissatisfied with IR systems. This has culminating in users having a low confidence that the results sets produced will contain a high percentage of relevant information. There has been a failure to provide query assistance and help facilities at the differing levels required for the diversity of activities, strategies and tactics performed for the variety of tasks and user characteristics found in the different retrieval contexts.

The problems and issues addressed in this chapter have led to the following research aims which attempt to improve the quality of IR systems and specifically the perception of these systems within the user population. The research aims are:

- Decrease the gap between user's mental models of the IR task and the functionality provided by IR systems. To achieve this it is necessary to identify the extent of the mismatch between users models of the search process and query support offered by systems. This will enable the design of IR interfaces and the dialogues used to be brought into line with the mental models of the different users.

- Improve the understanding of the cognitive activity underlying observed search behaviour to enable the prediction of the technology support requirements.
- Increase the efficiency and effectiveness with which information needs are satisfied by tailoring system advice based on a mediation between models of expertise, models of user activity and the search context. To predict requirements for supporting information retrieval in a task and context sensitive manner.
- Decrease the cognitive effort users expend on the IR process, and specifically demands placed on the user by the device, by providing context and user sensitive systems support. This should result in reductions in the time and training required for effective searching.
- Improve search traceability and diagnosability for the IR process as a means of improving the quality of the information set retrieved.
- To predict how the services a system provides may be combined to support different information retrieval needs, users and strategy combinations.

The aims condense into the following main research hypotheses:

Hypothesis 1: IR systems interfaces will be improved by basing their design on a sound task model and guidelines derived from a theory of user information seeking behaviour.

Hypothesis 2: Patterns of user strategies, tactics and behaviour can be predicted, and operationalised as task support functions provided by a computerised assistant

Hypothesis 3: Computational model implementation will improve designs by suggesting how the facilities of IR systems may be tailored to specific types of user and types of task.

Hypothesis 4: Searching performance, user satisfaction and confidence in results sets can be improved by supportive dialogues, enhanced articulation methods and targeted system assistance driven from an understanding of user activity and their current context.

These research aims and hypotheses imply a requirement for a comprehensive model of the retrieval process from a searchers perspective and the specification of the functionality and interfaces required if user performance is to be enhanced.

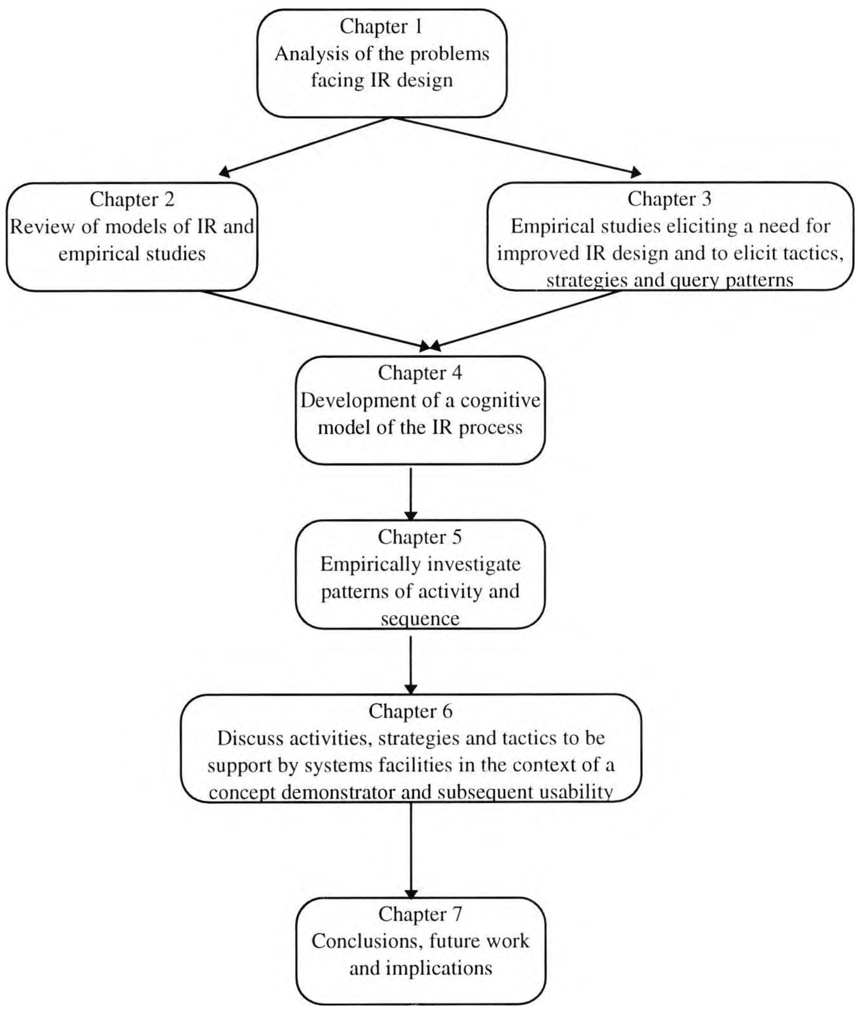

Figure 1.1 Overview of the thesis structure and rationale

Chapter 2 investigates the current literature in respect to modelling user behaviour for the retrieval task, empirical studies of the user characteristics affecting behaviour, strategies for information seeking and the components of information retrieval. The literature has been synthesised within a common framework in an attempt to look beyond a specific context or observed behaviour to understand why certain strategies and behaviour patterns are used. To conclude this chapter, techniques used in HCI to model user interaction are discussed.

Chapter 3 describes the results of an empirical study assessing performance, strategy use, query construction and facility use for a variety of user types and retrieval tasks. The study was performed on the MEDLINE database using the WinSpirs interface. The purpose of the study was to assess if a requirement for improved IR interfaces existed and to elicit the strategies and activities any cognitive model of the IR process must account for. The results of this study contributes to the development of the cognitive model presented in chapter 4 by indicating some of the activities, strategies and query patterns which need to be addressed. The study motivates the thesis as it establishes a requirement to view IR system design within the context in which it is to be used. The experiment performed produced a rich set of data and thus the analysis was split into a quantitative analysis (see chapter 3) and qualitative verbal protocol data (see chapter 5) for users retrieval sessions. The quantitative data included measures for performance, query characteristics, patterns of query development, term usage and analysis of the search facilities used. The qualitative verbal reports were not analysed until initial theory development was complete.

Chapter 4 presents a cognitive model of information retrieval based on a synergy of the literature on user models in IR (see chapter 2) and the results of the empirical study described in chapter 3. The cognitive model is presented within the context of the situations in which information is retrieved. This is necessary due to the iterative, dynamic and sometimes reactive nature of information retrieval. This allows the effects of the device on retrieval activity and the search strategies performed to be discussed. The model attempts to provide a more holistic view of the retrieval task within the context of the user, their task and the IR system.

Chapter 5 describes the results of a protocol study aimed at eliciting the search processes and mental activities performed for various types of user and task. The study investigates occurrence and sequence of activities. The study acts as a partial validation of the cognitive model described in chapter 4. The model presented in chapter 4 highlights a number of factors and contexts which impact upon behaviour. The model presented is complex due to the diversity of users and tasks in IR and thus only a partial validation is possible owing to the number of possible user, task and system combinations. This chapter investigates the activity reported by two user groups with differing device knowledge for a limited number of information need types. The chapter investigates activity occurrence and activity transitions as a trace of user behaviour through the retrieval process. The analysis and categorisation of verbal reports was undertaken to validate the theory. The subjects, tasks and experimental design reported in this chapter are the same as those found in chapter 3. It is only possible to perform the validation at this high level due to difficulties correlating verbal protocols to strategies and tactics and resource limitations given the time consuming nature of protocol analysis.

Chapter 6 discusses the user processes, strategies and search activities used in the development of queries requiring support from retrieval systems. The chapter discusses how this advice should be delivered for the different types of user, and the contexts in which it is appropriate to give this advice. The chapter takes a general view of the retrieval task, the various search stages and retrieval strategies in a discussion of how retrieval systems designs can enhance retrieval activity. The chapter proceeds to describe an exemplar web based retrieval system build for a specific category of user. The system embodies query development advice in terms of strategies and tactics based on the characteristics of the user, task and results. Two usability studies performed on the expert advisor are describe to conclude this chapter.

Chapter 7 summarises this research, discusses the findings and proposes future directions.

Chapter 2

Review of user models and empirical studies of information seeking and retrieval

This chapter provides background information for the thesis and details relevant research

Chapter 2: Review of user models and empirical studies of Information seeking and retrieval

2.1 Introduction

The chapter defines the components of the information retrieval domain before proceeding to discuss the relevant literature associated with modelling the behaviour of users. Current models of the IR process and explanations of search behaviour are considered as methods of solving some of the problems identified in chapter 1. This leads into discussion of the types of behaviour, or strategy types, present in the IR task. This is followed with a discussion of the literature concerning empirical work which assesses the situation factors affecting the outcome of user behaviour. The different HCI techniques for the modelling of users and interaction are reviewed. This is followed by a survey of previous research on expert systems that embodied intelligent strategies in an attempt to improve users retrieval activity. The chapter concludes by addressing the effectiveness of current views of the IR process and a summary of the situation factors which have the greatest effect on the queries and retrieval patterns produced.

2.2 Information retrieval a definition of its components

Traditionally the information retrieval domain has involved four components: the system, user, task and the librarian (Ingwersen 1982). The relevant components of information retrieval to this thesis are the system, user and the task; as the expansion and diversification of the users requiring information and the increased access to retrieval systems has minimised the librarians role in the retrieval process (Danolwitz 1993). The systems considered are document retrieval systems (e.g. bibliographic systems such as MEDLINE; or search engines Lycos, Webcrawler). The specific components of information retrieval systems are the user interface design, query dialogue, information presentation metaphors, tools and functionality. The thesis considers these components at a high level of abstraction, i.e. user requirements for services, as although the retrieval algorithm and computational technology underlying these facilities may alter the quality of results produced, i.e. systems recall and precision, they do not directly influence the users retrieval behaviour. The information content and document representations are modelled at the description level i.e.

quantity and quality of the results. Important aspects for modelling users' behaviour are the different situation contexts in which searching is performed, the knowledge held (Michel 1994, Allen 1991, 1994), the observed activity and the mental activity underlying the interaction with the retrieval system. The main questions to be answered are:

- Are the search techniques and behaviour patterns observed common to a category of user?
- How can the search activities of users be supported within the information retrieval systems design?
- To what extent do the task and the characteristics of the information required influence the retrieval activity performed?

2.3 Existing models of user activity in the search process

In this section several models of user behaviour within the retrieval process are discussed which provide high level observations of retrieval sessions in their description of user behaviour. The models are described along with the contributions they have made to increasing the understanding of the retrieval process and also aspects of the IR task they do not address. The observation of searcher behaviour is a starting point for modelling user behaviour but viewing the IR task at this granularity fails to provide detail concerning the mental activity performed or the underlying reasons for pursuing one particular course of action over another. Observed activity can be explained in terms of its underlying cognitive activity if it fulfils Neisser's (1966) requirements for a cognitive model (see figure 2.1). Thus given models of observed behaviour it is necessary to take the next step and extend these models by addresses and investigating the cognitive activity underlying the observations made. To gain a more comprehensive view of users retrieval behaviour the variations in observed search activity for differing search contexts must be explained within one model of the search process. To accomplish this any model of the search process must contain the general attributes required to explain behaviour and incorporate the adaptability necessary to allow the tracing or prediction of a searchers behaviour given the stimuli present.

Antecedent stimuli → Mediating cognitive event → Behaviour

Figure 2.1: General requirements of a cognitive model, Neisser 1966

2.3.1. States of information need transitions in IR

Taylor (1968) identified and categorised the four states through which the information need must be transformed in its development, as a search is specified, from its conception to communication to a third party:

1. the actual, but unexpressed need,
2. the conscious need,
3. the formalised need,
4. the compromised need.

Taylor identified these states from experiments into the communication of information needs to a librarian. Information needs aren't inherently verbal in nature and thus the user may encounter problems communicating the need to the retrieval system, due to the labelling effect (Ingwersen 1982), and this is one of the major causes of re-formulation. The last two stages of this information need transition are dominated by the librarian using their expert searching skill to formulate queries and respond to the results with appropriate strategies and tactics. The removal of the librarian from the search process will cause articulation and search strategy problems for users especially when determining how to react to the results produced.

2.3.2 Pearl growing

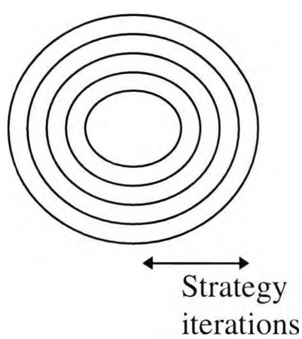

Figure 2.2 Pearl growing or onion peeling behaviour in search progression

Figure 2.2 shows Markey and Atherton's pearl growing model (1978). This acknowledges the iterative nature of many searches and implicitly addresses the dynamics of interactive searching activity. The model requires that the search system enables easy search iteration, otherwise effort will be duplicated, and that a searcher makes relevance judgements as the search progresses which effect the nature of query development. The searcher starts with a query on a specific set of items; reviews the retrieval set associated with this and refines the query in subsequent retrievals based on the results produced. The search is a multi-layered and cumulative process whereby the initial search request is narrowed (or broadened); focused and refined by altering the granularity of search concepts or terms used in the query. The focusing of a query can be either additive (add more terms to further constrain the results) or deductive (remove terms to reduce constraints on the results) depending on the granularity of the information retrieved (i.e. reaction based on the quantity and quality of the results set). The importance, or impact, of a search item is not limited to those which successfully contribute to the solution set for the need but also those which discount alternatives and thus narrow the space of possibilities within which the solution exists. The inter-twined nature of the search process emphasises the iteration in the search and the importance of feedback on strategy reformulation. The pearl growing model is a high level description of observable actions users perform in the search process but fails to sufficiently break it down into the detailed activities users perform as the information problem is refined, or to account for the different types of retrieval strategy e.g. directed or semi-directed browsing (Herner 1970, Apted 1971). The model was not designed to predict or explain problems searchers have in the retrieval of information. This view of IR lacks a system model which is a dominant factor as it affects how the query pattern and strategy develops.

2.3.3 Information search tactics

Bates' (1979) account of the search process deals with the overall planning of the search through the use of tactics for the implementation of short-term goals and manoeuvres to facilitate the solution of an information need. Every move towards finding the desired information is a tactic and the search strategy is a plan for the whole search. The model acts as a facilitation and teaching tool for improving the interaction with the system, but not as a prescription of search behaviour. The model suggests the tactics searchers should apply to their problem situation in an effort to improve the efficiency and effectiveness of the search process. The research offers four types of search tactics: monitoring tactics, file structure tactics, search formulation tactics and term tactics. Monitoring tactics aim to keep the search on track and make sure its efficient:

- *Check*: to compare the current request with the original request for compatibility.
- *Weigh*: to make a cost-benefit assessment for search actions
- *Pattern*: to be aware of a particular search pattern and alter if not efficient
- *Correct*: spot and correct mistakes
- *Record*: to keep a track of options followed or ignore during the search

File structure tactics provide techniques for threading through the file structure of the information facility:

- *Bibble*: to look for a bibliography already prepared and to check if the search planned has been done in a usable form by others.
- *Select*: to break complex problems down into sub-problems
- *Survey*: to review the available options
- *Cut*: to use searches which cut out the largest part of the search domain at once
- *Stretch*: to use a source for other than its intended purpose
- *Scaffold*: to design a indirect route through the information files and resources to the desired information
- *Cleave*: to employ binary searching in locating an item in an ordered file

Search formulation tactics enable the construction or re-construction of the search formulation:

- *Specify*: to search on terms that are as specific as the information desired
- *Exhaust*: to include most or all of the elements of the query in the initial search formulation
- *Reduce*: to minimise the number of elements of the query in the initial search formulation
- *Parallel*: to make the search formulation broader by including synonyms
- *Pinpoint*: to make the search formulation precise by minimising the number of synonyms used
- *Block*: to reject items containing certain terms

Term tactics enable the selection and revision of the specific terms within the search formulation:

- *Super*: to use a broader term
- *Sub*: to use a narrower term
- *Relate*: to use a co-ordinate term
- *Neighbour*: to seek additional terms by looking at neighbouring terms
- *Trace*: to examine information already found in the search to find additional terms
- *Vary*: to substitute search terms
- *Fix*: to try alternative affixes
- *Rearrange*: to reverse or rearrange the words in search terms
- *Contrary*: to use a logically opposite term
- *Respell*: to search under a different spelling
- *Respace*: to try spacing variants

The model aims to alter the users behaviour towards the practises of efficient IR rather than to provide support tools to enhance instinctive retrieval behaviour. The focus on user adaptation over system support causes the tactics to have limited use in the improvement of the IR process apart from indicating how models of expertise may be developed as the majority of end-users are not aware of these tactics.

2.3.4 Models of communication in IR

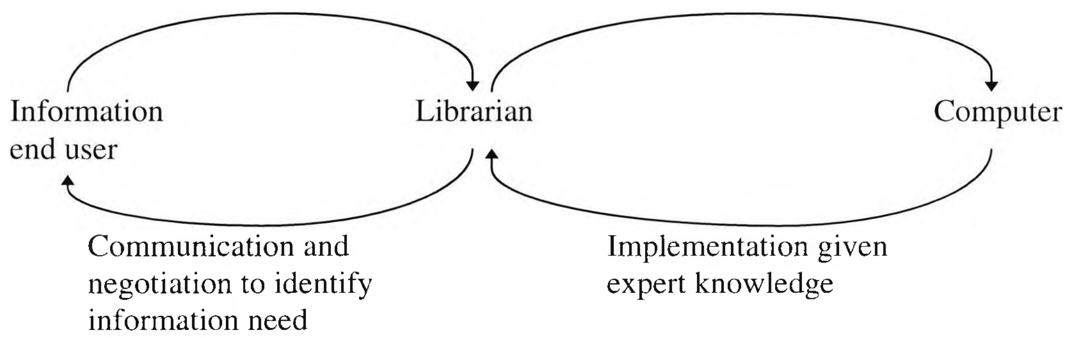

Figure2.3: Communication models of the negotiation in need definition and query formulation.

Ingwersen (1982) produced two models of different aspects of the information retrieval task; one containing the formal steps in the retrieval process and the other is a cognitive model of information retrieval. This research is based on the view that users of information negotiate with an expert intermediary, librarian, who then implements a search on the user's behalf. This view of the retrieval process is largely outdated as increased access to IR systems has forced the user into the driving seat in the search process and now most searches are conducted by the end-user of the information (Danolwitz 1993). Ingwersen's models are geared around two functional phases:

1. Human-human interaction between a user and a librarian;
2. Human computer interaction between the librarian and the IR system.

The models focus on the communication between these three agents in the retrieval process. The applicability of such models in current IR practises is limited as the search participants and their roles have changed. Ingwersen's view of IR has a fundamental flaw as the users original information requirement is filtered through a increased number of agents (i.e. information user, expert intermediary and information system) this has the potential to increase the chance that perceptual differences occur between the end-users conceptual need and the query implemented by the librarian. The models operate at a high level of granularity and do not address the search methods, strategies, effects of systems facilities and end-users personal experience on the behaviour produced.

Ingwersen's model describes the formal steps in the retrieval process as follows:

1. Information need of user (deriving from a problem situation)
2. The formulated information need of the user
3. User-librarian negotiation
4. Developing the search profile - topic analysis
5. Choice of tools
6. Looking up. Systematic or alphabetic
7. Judgements based on index (terms)
8. Judgements based on descriptions, abstracts, titles
9. Evaluation of the documents themselves

The formal steps are based on the activities performed in the retrieval process as an expert intermediary attempts to elicit and satisfy an end-users information need. The model fails to consider the effects of diversity in the characteristics of the users of the information and the diversity in the types of information needs. It would be expected that if the structure of the retrieval process is controlled by an expert intermediary then it will conform to a formal structure for the IR process. The average end-user requiring the information is unlikely to have been formally trained and thus their IR behaviour is unlikely to conform to a formal IR approach if they are not directed by an expert intermediary. The model needs to be extended to represent the new user population and thus must address the cognitive processes underlying strategy development for novice searchers. This model acknowledges the existence of feedback loops in the retrieval process but fails to formalise the search situations in which these loops are applicable e.g. the reaction to low results. The cyclic, iterative and reactive nature of IR generates a need for a greater understanding of the cognitive processes which occur in the search process to determine why specific courses of action are taken. The model needs to account for the influence of the system; its facilities, results feedback and dialogue, on the outcome and structure of the retrieval process.

Ingwersen's cognitive model for IR focuses on the transformations that are performed on knowledge states in the IR process, but fails to address the methods used to facilitate these transformations. The information transformations are at a high level of granularity and don't determine the causes of a particular users search actions. This high level overview of the retrieval process fails to show the iterative nature of searches or explain IR behaviour other than direct query generation e.g. browsing. Ingwersen's cognitive models focuses on the knowledge transition angle of IR using the classification of knowledge states in the retrieval process of Taylor (1968) and Derr (1983). Ingwersens model does not address the causes of search activity underlying the different states of user knowledge. The model has limited applicability in the general design of IR systems because the effects of personal experience and the granularity of user knowledge on the activity, strategies and tactics performed are not explored and because it models the cognitive behaviour of librarians not the average end-user. The model is too general and expert specific for it to act as a base upon which to generate intelligent IR advice or determine the facilities and mechanisms required. The model doesn't lend itself to the prediction of the type of user interface and user-system dialogue which would enhance users retrieval activities.

2.3.5 On-line retrieval heuristics

Harter and Peters (1985) attempted to identify and classify the heuristics used in the IR process as methods and procedures. The search heuristics are categorised in the following typology:

1. *philosophical attitudes and overall approach*
2. *language of problem description*
3. *record and file structure*
4. *concept formulation and reformulation*
5. *recall and precision*
6. *cost/ efficiency*

These categories are further decomposed into general heuristics. To illustrate the nature of heuristics those associated with philosophical attitudes and overall approach are described as follows:

- *Stay loose, be flexible, look at a search in more than one way.*
- *Serendipity is important for effective retrieval. Browse.*
- *Good searching is heuristic, interactive.*
- *Search formulations are best viewed as hypotheses, as in scientific enquiry.*
- *Anticipate methods for narrowing or broadening a search prior to its conduct.*
- *Save search output to analyse if problems occur.*
- *Be sceptical of systems responses.*
- *Always question null sets.*
- *Sample retrieved citations before printing final bibliography.*
- *Browse intermediate results of a search to assess relevancy of output.*
- *Keep current with online searching literature and with manuals of search services and database producers.*
- *Be familiar with your terminal.*
- *If several paths are available for searching, pick the one with the smallest domain.*

The work follows on from Bates (1979) search tactics for efficient search behaviour. The research is interesting as it explores some general heuristics for improving on-line retrieval. It is recognised that IR has a complex and adaptive nature; and is sometimes a trial and error problem solving process in which heuristics are used to reduce the number of possibilities considered. Bate's research specifies the methods most appropriate for a variety of problem situations if a searcher is to retrieve results quickly. The onus of this research is to modify user behaviour by educating them of the search tactics and general retrieval principles for producing more effective performance. This approach is possibly justifiable if expert intermediaries are to be trained more effectively but it is an unacceptable demand for the average end-user. The focus of attention, if the effectiveness of retrieval activity is to be improved, should concentrate on the interface and facilities provided by the system. The interfaces must provide assistance to a searcher if they are to attain the performance of a retrieval expert without the training effort associated with becoming an expert. The system should be designed to match, and support, the cognitive processes and strategies a user performs rather than an attempt to adapt their behaviour to the system.

2.3.6 Mental models of IR

Borgman's (1985) work aimed at enabling searchers to be trained to develop a more complete mental model of a system. This improved mental representation and understanding of the device can be used to improve the interaction with the system. The work tried to establish that a searcher trained with a conceptual model of an IR system would out perform those trained procedurally, and assess whether there were any differences based on the complexity of the task performed. The results of her experiments suggest that no differences in user performance occurred on simple tasks but that searchers trained with a conceptual model out-performed those trained with a procedural model on complex tasks. Furthermore, Borgman found that patterns of user actions and systems responses were significantly different on complex tasks. This work tries to improve the quality of the search process and systems interaction based on different types of user education. The research has implications for this thesis if the effectiveness of the retrieval process is to be improved by promoting an improved match between the system's functionality and the users' mental model of the search process. Intelligent help, information presentation methods and search guidance should take into account the differences Borgman found concerning improvements in systems interaction and its relationship to task complexity.

2.3.7 Human information seeking

Rouse and Rouse (1984) reviewed the context factors effecting information seeking behaviour. They recognise that the search process is dynamic and has a number of different variables which can effect the methods and information sources used. Their paper reviews models of information selection highlighting many problems with these models and their failure to represent the information seeking task. The factors effecting the search process are identified as: payoffs or costs, resources available, expected value of gambles, information update rates, amount of information sought, diagnosticity of data, distributed characteristics of data and conflicts between information sources. The paper is useful as it highlights the many factors that a model of the search process must address, but views the IR problem at too high a level of granularity failing to break it down into its constituent parts.

2.3.8 ASK's in IR

Belkin (1987) produced an account of the IR process in terms of anomalous states of knowledge, when a searcher is often faced with a poorly specified and incomplete information need. The searcher uses 'anomalous states of knowledge' (ASK's) to aid in the understanding of the problem space with which they are faced. The searcher explores and clarifies the problem through an iterative and adaptive search process. The model considers the searching activity for a particular type of task and user, but does not address the cognitive processes underlying searcher behaviour as they explore the problem or the relationship between the design and facilities of the system and the retrieval behaviour observed.

2.3.9 Berry picking

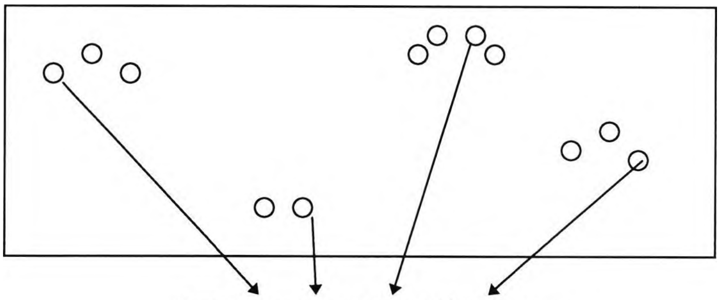

Figure 2.4: Berry-picking view of the search process

Bates (1989) introduced the idea that the search process was not a simple straight forward location of the required material. The Berry-picking model describes the searcher's use of the search space during the formulation of a query. The searcher selects relevant information from the information space and uses this to enable the production of more effective queries. The model shows how browsing activity reduces the cognitive load associated with complex IR tasks and describes how some searchers produce a retrieval set from partial searches. The model conceptualises the search problem as a number of search goals which are solved as retrieval progresses. This moves away from the idea that the searcher knows specifically what information they want and how to locate it. This is in tune with a diversification of the knowledge held by the end-users of retrieval systems. The information space is seen as a berry bush and the searcher follows various lines of attack, or branches, to satisfy parts of their information problem. The berries can be either articles or terms, which allow the searcher to fulfil their search requirements. The search is seen as an evolving process as paths are explored and search specific knowledge is accumulated. The searcher uses querying not only to locate the articles they require, but also to explore the terms and tactics to be used in subsequent queries. Iterations of the retrieval process can be used to reconcile ambiguities and gaps in the searcher's knowledge as the problem is formalised and satisfied through interaction with the system. The model acknowledges that the information problem is not always solved by a single retrieval set but can consist of a series of stages in a refinement cycle. The model indicates the different strategies used and the observable actions associated with the search process. The model raises interesting issues associated with the sub-division of problems and the use of sub-goals in IR. It can be assumed that, due to the nature of the search type identified, searchers activities vary with user knowledge. However, the model does not address the cognitive activity underlying the observable activities identified or the effects of knowledge on the different search processes and strategies. The model also does not account for the effects of the system design and the dialogue used on observed search behaviour.

2.3.10 Information seeking process

Kulthau (1988, 1993) recognised that searchers share common experiences in their interaction with information retrieval systems. Her model (ISP) comprises of six stages: initiation, selection, exploration, formulation, collection and presentation. The 'initiation' describes the user's recognition that a need for information exists; 'selection' enables the selection of the general topic to be investigated and the approach to be pursued; 'exploration' enables the exploration of information about the general topic and is used to a develop the users understanding of the information required. 'Formulation' enables the formation of a focus from the information encountered in 'exploration.' The 'collection' phase enables the user to gather information relating to the focus topic; while the 'presentation' stage enables the search to be completed, the problem to be resolved and the preparation to use the findings. The research recognises that the search process is iterative and can be non-linear (divide and conquer rather than simple query progression) but does not explore the situations in which the different steps are applicable or the cognitive activity underlying observable actions. Kulthau's (1991) empirical studies reveal that user behaviour is a more recursive and iterative process than described within her model. The failure to address the underlying cognitive processes and individual differences means that the model is unable to specify or predict the pathways the user will take through the ISP steps. Furthermore, the model does not address the different strategy types observed in IR research: browsing, sub-goaling etc., and does not account for the influence of the system features and facilities on the retrieval activity performed.

2.3.11 Electronic information seeking

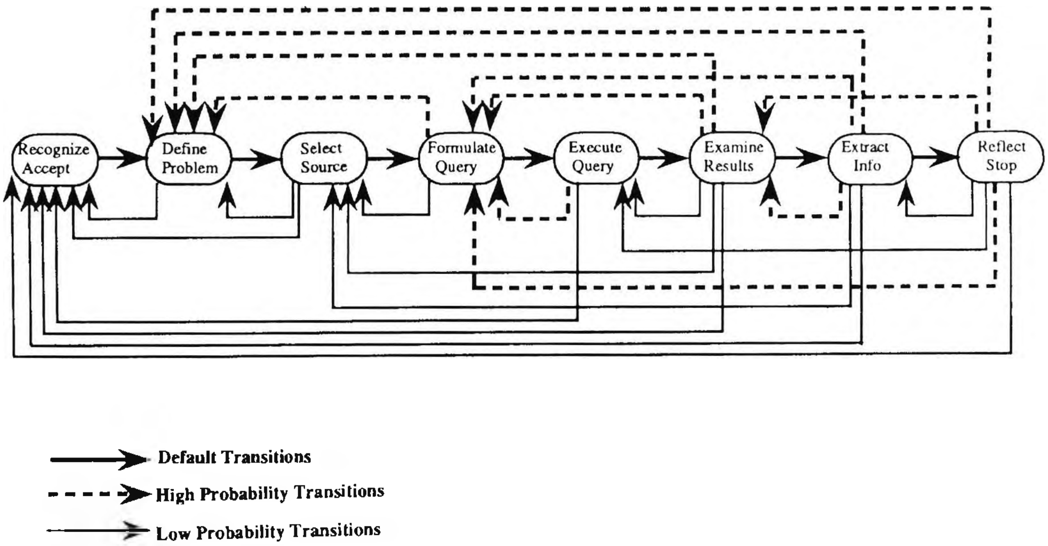

Figure 2.5: Model of the information seeking process

Marchionini's (1995) model of the information seeking process defines a number of steps a searcher is believed to perform in the search process. The model aims to make explicit a distinction between the different activities a searcher performs and their possible inter-relationships. The model starts with the recognition that an information problem exists and the acceptance that it should be solved now. The definition of the problem and the development of an understanding of the information required is used to constrict and determine the form and scope of the search. The searcher chooses the search system within which relevant results are expected; based upon their knowledge of the domain and previous experiences with the possible devices. A query is formulated by matching the current understanding of the task against their model of the system to be searched. This matching is composed of a semantic mapping between the system and the user's vocabulary and an action mapping based on the set of possible actions given the search system's limitations e.g. translation of a users need into a Boolean expression. Search execution communicates the query to the system. The examination of search results is used to define the extent to which the results produced match the solution sought. Information contained in the results is evaluated based upon its characteristics and its relevance to the information seeking task. The searcher finally reflects on the completeness of the information found and determines the subsequent course of action i.e. further iteration or search termination. The model operates at a high level of granularity and does not make explicit how these processes are performed; their direct influence on the behaviour of subsequent processes or how to determine the different routes through the model given users with different levels of knowledge, varying task characteristics and differences in the characteristics of the results retrieved. The model does not consider the effects of different strategy types and knowledge, domain and device, on the search process and omits situation factors such as task complexity (Borgman 1985, Large et al 1994) and the different types of knowledge used at the different stages in the search process (Ingwersen 1982; Allen 1991 and Michel 1994). The model breaks search activity into sub-processes and recognises that there are complex interactions between these processes as a search is produced, but does not explore the complex activities associated with executing the processes or inter-relationships which may exist between them. The main limitation of the model is that it does not account for the effects of the retrieval system on the strategies and processes performed. The effects of the retrieval system are not confined to the results produced but must also consider the facilities and functionality present and the level of interaction required for their use.

2.3.12 Dialogue scripts for information seeking

Belkin (1987) proposes a list of 16 information seeking strategies (ISS) based on a interaction between four situation dimensions: method of interaction, goal of interaction, mode of interaction and the resources considered. The method of interaction is categorised as scanning for interesting items from a collection of information or searching for a known item. The goal of interaction is categorised as learning or selecting items for retrieval. The mode of retrieval is identified as either the specification of the items required or retrieval by association. The resources considered refer to the inspection of the information in the database or meta-data describing the structure and content of the database. An information seeking strategy is composed of the users situation described in terms of these situation dimensions e.g. a search may require a known item which is to be selected from the database if it is associated with another search item by consulting the system meta-data.

Belkin et al (1993) propose that searchers use scripts (Schank 1977) to determine the activities to be performed as the search progresses, based on these information seeking strategies (ISS). The search process is composed of a number of different scripts which specify the actions and goals which must be executed if the search is to be completed. The scripts are triggered by the situation with which the searcher is faced as defined by the 16 information seeking strategies in the model. The scripts refer to interaction patterns or plans of query dialogue which should be optimal given a set of information triggers. The scripts provide a structure to enable an assignment of a hierarchy for the goals to be satisfied. This account of IR does not attempt to model the searchers' behaviour or decision process in the production of these scripts. A script is a optimal dialogue, or procedure, based on the constraints of the system. Belkin has currently only specified a sample of scripts associated with the different ISS and acknowledges that the ISSs and their associated scripts still require empirical validation. The use of scripts for the interaction process could improve the information sets retrieved but are unlikely to improve searcher satisfaction with IR systems.

2.3.13 Elements of a cognitive theory of IR

Ingwersen (1996) discusses the component elements of a cognitive IR theory. He views IR as the communication of information between a user's world model and that held by the system. A users world model consists of highly dynamic and inter-changeable cognitive structures which control perception and the processing of external information. Cognitive structures are defined by the users experience and may change based on the information encountered i.e. users understanding may alter based on the information encountered in retrieval. Ingwersen uses the notion of poly-representation to represent the user's information need, problem, knowledge states and domain work task in a structure of causality and indicates that a variety of IR techniques should be applied to information objects based upon these different viewpoint representations. The users context and cognitive state are categorised within a matrix of intrinsic information needs (see figure 2.6) which are based on two categories: well-defined or ill-defined; stable or variable. The behaviour observed is linked to the categorisation of the information need in respect of these contexts. The presence of an information need acts as a trigger for information seeking. The process of learning and reactions to search outcomes enable transitions between the categories of information need. Ingwersen concludes that the matrix of information needs indicates an approximate classification and that more complex intermediate forms may occur in retrieval.

	Well-defined	(external behaviour: search loops) **Ill-defined**
Stable	1 **Rich, variable, cognitive state** Conceptual 'throwness' Limited uncertainty Topical relevance assessment : Yes Curiosity: low Confined Navigation	4 **Weak, variable, cognitive state** Conceptual 'breakdown' High uncertainty Topical relevance assessment : No Curiosity: low Matching on the spot (dead ends)
Variable	2 **Rich, variable, cognitive state** Conceptual 'throwness' Controlled uncertainty Topical relevance assessment : Yes Curiosity: high Exploratory navigation	3 **Weak, variable, cognitive state** Conceptual 'breakdowns' High uncertainty Topical relevance assessment : No Curiosity: high Browsing

(external behaviour:
Berry-picking)

Figure 2.6 Matrix of intrinsic information needs

System world models consist of pre-defined and fixed structures (for example algorithms, indexes, thesaurus) which may interact with each other but do not change each other. The processing of information occurs at the linguistic level not a cognitive level. Computers thus interpret information based on pre-defined principles whereas humans may form a deeper cognitive understanding which may change the interpretation of information. Poly-representation of the system occurs because a number of different agents produce the components of information systems: human indexers, thesaurus constructors, information producers, database designers and retrieval engine designers. Each of these agents may interpret and structure information through different representations which are encompassed in the retrieval system.

Ingwersen's model advances the understanding of IR behaviour through its dynamic view of the information need, problem states, knowledge representations and the recognition that multiple context variables influence activity. The categorisation of information needs and the acknowledgement that this dictates the strategy performed is an important and significant aspect of this research. The interaction between the users' model of the world and the system's model of the world are introduced but the complex interactions of one model on the other are not explored. The research does not explore how variations in the systems model effects users need development and thus the strategies used. The model acknowledges the effects of relevance and feedback on the reformulation of information needs but does not explore these at a lower level of detail e.g. Bate's (1979) search tactics. However, the model does focus on a cognitive viewpoint for IR in terms of the different and sometimes conflicting information representations present in retrieval situations but does not offer a view of user process as information is translated between these representations.

2.4 Types of search activity and tactics identified with associated definitions

The different types of search activity are not mutually exclusive and thus the behaviour patterns observed can be a function of multiple strategy types. Any model of IR which attempts to simulate the retrieval process must explain the following strategies.

- Sub-goaling (divide and conquer): The searcher breaks the problem into information parts which are solved individually. For example 'lets look at this first then we'll go back to that before bringing it all together'. This type of activity has been demonstrated by the observation of berry picking (Bates 1989).

- Information seeking may occur in situations in which the target is not known and needs to be defined in the process of retrieval. Even in situations in which the search target is known, people browse to add context information. Kwasnik (1992) found that the goals of a search may emerge as a function of performing browsing strategies. The activity performed is informal, opportunistic and reactive based on information encountered during the search process. Three categories of browsing activity have emerged from the IR literature:

- Directed (Herner 1970) or specific (Apted 1971) browsing in which activity is directed by a specific and known goal. This type of browsing is systematic and focused. The searcher selects information elements as they navigate the information space structures until locating the desired information.
- Semi-directed and predictive (Herner 1970) or generally purposeful (Apted 1971) browsing in which activity is directed by a less definite target and thus the navigation activity is less systematic and structured. An example of this type of browsing is the use of general queries and article skimming. The searcher explicitly consults the information structures, or help facilities, of the system in a quest for information. For example 'I'll just look in the thesaurus to see if there are different ways of saying that'. Ellis (1984) states that browsing is semi-structured searching.
- Undirected (Herner 1970) or general (Apted 1971) browsing which is characterised as having very little focus and no real goal. However, Kwasnik (1992) highlighted that people impose structure on their activity very quickly and thus even casual browsing soon becomes purposeful.

- Direct query composition and rapid prototyping ('quick and dirty'): The searcher tries queries in succession looking for the one that gets the results. This behaviour is seen as an expansion or adaptation of the query in respect to the previous results retrieved. The distinction between direct query composition and rapid prototyping occurs due to the emphasis placed on activities in the search process. Direct query composition focuses on the solving of search problems pre-implementation and the production of more complete and well articulated queries. Rapidly prototyping activity focuses on the solving of search problems through interaction with the systems and a 'trial and error' approach which is a more reactive strategy as queries are adjusted based on the results.

It is evident that multiple different approaches exist in a retrieval session and these strategies will each have situations in which they are the most efficient and effective.

User knowledge may effect the type of strategy used in the same search context e.g. novices may try and reduce interaction by limiting the number and type of queries submitted whereas experts may prefer to specify the information need more precisely.

2.5 Expert IR systems as computer intermediaries in the search process

Marcus (1983) conducted experiments to compare the effectiveness of a computer search intermediary (CONNIT) with human search intermediaries. CONNIT enabled searchers to query multiple distributed databases each with their own query and results presentation protocols through a common language and interface. The system translates the users requests in the common command language of the CONNIT system into the appropriate commands necessary to interrogate whichever system is to be used. The results were inconclusive as sometimes the computer and sometimes the human intermediary were superior based on measures of recall and search times. However, in general with the use of the computer intermediary recall was higher but so were session times. The computer intermediary could improve recall but the advice wasn't as focused as the human intermediary who could get to the result quicker. Marcus concluded that as more comprehensive models and techniques for on-line searching became available that these should be incorporated into computer intermediary systems to make them more effective.

Vikery et al (1987) developed the PLEXUS system as an expert intermediary in the search process for a gardening domain. The system attempts to generate a formal problem statement applicable to traditional Boolean retrieval systems from an informal user input. If the user input does not match the systems dictionary the user is required to locate the term within semantic categories of the domain. The PLEXUS system evaluates the results of a search in terms of the number of article matches and reformulates the strategy and query accordingly. The system takes control of querying until it determines that an acceptable number of results have been retrieved. For example if too few results are retrieved the search statement is reformulated by successively:

1. dropping terms with zero articles;

2. dropping terms that the system judges to be of minor importance;
3. introducing broader terms from the systems domain classification;
4. dropping further terms to try and identify more general sources within the database.

Vickery et al found that the specificity of user input could not be matched to the more general terms used in the index to the database. The success of the system in automatically adjusting queries is dependent on the quality of the systems domain representation and it was apparent that search modification techniques do not have the flexibility required. The system bases search amendments solely on the number of articles retrieved whereas users are likely to wish to amend search requests based on the quality of the information within articles as well as based on then number of articles. The number of results required in any search situation is dependent on the user and the type of information need they have which is not acknowledged in the system. The third strategy in the example indicates that the system is making importance judgements on term inclusion based on a domain model whereas in reality importance of particular concepts to a need will depend on the situation for which this information is required not solely based on a domain model.

Crofts' (1989) I^3R system uses expert systems techniques to help users apply powerful statistical retrieval techniques and browsable displays. I^3R provides a variety of facilities and system experts to assist in three phases of IR: query formulation and reformulation, search and user evaluation. System experts contain rules composed of a condition part and a action part. For example

- *Condition*: request model contains unevaluated documents
- *Action:* obtain user evaluations of documents and update the request model

The experts implemented in I^3R are

1. User model builder which collects information about the user to enable this to be matched against a category of user. This categorisation is used to determine the style of interaction, session goals, and other information.
2. Request model builder enabling the formalisation of a model of the user's information need.

3. Domain knowledge expert enables the inference of concepts related to those in the initial query based on the user model and systems knowledge base.
4. Search controller selects and executes formal search strategies based on the probabilistic model and clustering.
5. Browsing expert enables the informal navigation of information by the user.
6. Explainer provides explanations of systems actions in response to user demands.

Good performance was reported with the I^3R system but it required significant computational resources and can place more demands on the user than conventional systems (Croft 1989). The aim of the system was to enable the acquisition of more detailed specifications of the user's information need to improve retrieval performance. The approach followed in this system seems to be effective in improving recall scores but domain knowledge is expensive and difficult to acquire hence the power of such intelligent query refiner systems depends on the quality of the domain knowledge base.

The RABBIT system provides an interface which aids in the formulation and re-formulation of a query (Williams and Tou 1982, Tou et al 1982). Users interactively refine partial descriptions of the target by critiquing successive examples that satisfy the query description. These critiquing actions result in RABBIT's reformulation of the query. RABBIT provides six critiquing options:

- REQUIRE: add the attribute to the query
- PROHIBIT: applies the relative complement of the attribute to the query
- ALTERNATIVES: suggests a series of alternative values and inserts those selected into the query
- DESCRIBE: allows the recursive specification of an embedded query
- SPECIALISE: refines the generic category selected by allowing a user to select one or more sub-concepts
- PREDICATE: allows the user to apply a predicate to the value of an attribute.

The critiquing options displayed depend on the characteristics of a specific element of the results to the current query chosen to be critiqued by the user, rather than a more general view of the current search context e.g. number of results retrieved. The system automatically reformulates the query based on the critiquing option selected. RABBIT facilitates and assists users with only a vague idea of the target information or casual users with limited knowledge of a given database to achieve more effective recall. The RABBIT system and reformulation clarification operates on a relational database not bibliographic databases.

2.6 Factors affecting information retrieval

Previous empirical studies of searcher behaviour have drawn attention to a wide variety of factors that affect performance; for instance, the display of retrieved results can alter search strategies (Efthaniades 1993, Allen 1994), the information need type influences search behaviour, (Elkerton et al 1984, Marchionini 1995); while task complexity, reflected in the information need, can affect user's search behaviour (Borgman 1985, Large et al 1994). Furthermore, information source selection (Bassilli 1977), and the user's model of the system and domain impact on the search process (Allen 1991, 1994 and Michel 1994); while motivation (Jacobsen et al 1992) and the importance of the information need (Irwin et al 1957, Wendt 1969) influence search duration and the effort a user will employ. In this section these factors and others associated with user interaction are discussed.

2.6.1 Effect of knowledge on the interaction process

The effects of user knowledge on the quality and efficiency of the interaction observed was highlighted by Rasmussen (1983) and Solomon (1993). Rasmussen classified the interaction related knowledge of users into one of three types: skill based, rule based or knowledge based. The speed of interaction, attention required to interact and the choices between the facilities to use vary given categories and level of user knowledge. Skill based interaction consists of smooth integrated and automated patterns of interaction which do not need to be attended to consciously. The user has instance level experience of how the device should behave in various situations. Rule based interaction is driven by partially compiled procedures of the required activities and are consciously attended too. The user has knowledge of general principles about how the device should behave. Knowledge based interaction is driven by the current state of the interface and interaction from first principles (trial and error); and thus attention focuses on problem solving rather than proceduralised interaction. The extent of the effects of interaction related knowledge on information retrieval behaviour varies based on the complexity and granularity of the facilities and functionality offered by the device. Lancaster (1977), Timbie et al 1969 and Lancaster et al (1972) found that for both experienced and inexperienced searchers the major problems occur due to ineffective search strategies rather than the mechanics of the specific system command language but the command language can be a significant barrier for very inexperienced users. However, Sewell (1986) found that users' knowledge of the device effected the relationships used in the expression of information needs. Novice searchers only use 'AND' relationships submitting multiple parallel searches rather than more complex query representations using 'OR' relationships. Martin (1973) and Oldroyd and Citroen (1977) noted that even experienced searchers sometimes fail to use obvious synonyms and do not always pursue useful strategies. This indicates that even experts require assistance and advice if they are to perform effectively and consistently.

2.6.2 Types of knowledge involved in the IR process

The knowledge a searcher recruits in retrieval activity and in the selection of terms to use in queries has been extensively researched by Ingwersen (1982), Allen (1991, 1994), Michel (1994). The differing types, levels and sources of knowledge have implications for the structure of search activity and the behaviour observed. This highlights that the support provided should be matched to experience. The major categories of knowledge effecting activity are domain, IR strategy and device knowledge as discussed by Allen (1991, 1994) and Michel (1994).

Fidel (1991) in a study of 47 professional searchers found that search term selection is related to three factors

1. Request related factors, such as specific terms that allow the information need to be formalised or are key terms which have a high recall in relation to the topic of the search, accounted for 32% of decisions.

2. Database related factors. These factors are the constraints of the functionality of the system and thus limit the search by the relevance of the term to the system, such as there was no thesaurus or a need to search several databases, accounted for 48% of decisions.
3. Searcher related factors, such as individual preferences accounted for 20% of decisions.

Factors influencing searcher's preference for search terms were:

1. searching styles of the searcher
2. subject area
3. number of databases per search
4. quality and availability of thesaurus and indexing

It appears that even within a group of expert searchers, search term selection and thus the progress of the search is determined by knowledge related factors as well the user device knowledge. Unfortunately this study didn't look at the term choice decisions of novice searchers and expert term selection may be driven by knowledge of retrieval strategies and tactics not just domain knowledge.

2.6.3 Motivation

Jacobson et al (1992) found the enjoyment users associated with different methods of interaction or search facilities effected method choice rather than it being directed by the most appropriate facility for the task. The motivation of a searcher effects the persistence of search behaviour and the tasks selected. The motivation of the user effects the evaluation criteria used to compare the retrieved information with the desired information, for example the level of motivation will affect the reaction to a results set being marginally greater than expected. Motivation does not effect the pattern of the cognitive tasks underlying behaviour but it alters the time spent on and completeness of the different search activities.

2.6.4 Payoffs, costs, value of information and the critical nature of information

The effects of the utility of the information sought on the search process are highlighted by Wendt et al (1969) and Irwin et al (1957). The effects of the utility of the information need on the search process is to alter the time and resources, both monetary and cognitive, that searchers are willing to expend on retrieval.

2.6.5 Information types and task types

Elkerton et al (1984) found that the information types frames the user's interaction techniques and strategies used. The work highlighted the effect of the systems information space in determining the interaction techniques and strategies used. Ingwersen (1996) introduced a 2*2 matrix of intrinsic information needs which was used to define the type of strategy used i.e search loops to Berry-picking. The nature of the information need can evolve within the search process given the systems influence on user behaviour i.e. relevance feedback (Saracevic et al 1990, Su 1994).

2.6.6 Task structure and complexity

The structure, expression and order of the information present in an external task may frame search behaviour, as observed in decision analysis and problem solving literature (French 1984).

Large et al (1994) and Borgman (1985) found task complexity effected the interaction methods and search facilities chosen. It seems intuitive that task complexity will effect the level of search activity and the type of strategy used due to the cognitive load associated with more complex tasks. Large et al (1994) found that given the same information task and systems access paths to information, users choose common information cues to extract the information they require. However, the experiments were in a constrained domain in which the structure of the tasks may have given clues to the strategies which should be used.

2.6.7 User intention

Searle (1984) and Derr (1983) state that a desire for information cannot exist without a reason or underlying 'information purpose' i.e. an intention or goal. Solomon (1993) found different user intentions, e.g. fact retrieval and location of material, effect the methods chosen to execute a search. User intention only effects behaviour if the information need is initiated by the end user not a third party, e.g. information to answer an essay question as part of academic studies.

2.6.8 Physical constraints and user expectations

Case (1991) found that physical characteristics of both the office (spatial constraints and layout) and documents (physical form e.g. hardbacks, paperbacks, journals, articles) are given priority over topic in determining storage location or the subject classification of information. Importantly physical embodiment plays a crucial role in retrieval as it provides a rich set of visual and physical cues. Users have expectations about the form of the material they expect to retrieve. This was emphasised by Large et al's (1994) findings that the search process and the choice of access path is a function of the users perception of the information gathering task. This was confirmed by Saracevic and Kanton's (1986) findings that the information seeking process is affected by the users' perception of the information need and its nature. Efthimiades (1993) and Allen (1994) found that the structure and order of the information in a display plays a significant role in a person's perception and interpretation. The effect of task on the retrieval process indicates a need to look at the categorisation of tasks and their effects on information retrieval behaviour.

2.6.9 Cognitive style

The cognitive style of the user may effect the approach taken to information seeking, Brindle (1981), and the relevance assessments made, Davidson (1977), but it is believed that these effects are superseded by task characteristics and the level of the users knowledge.

2.7 Human-computer interface design

Human computer interface design aims to produce systems which satisfy users requirements and minimise the effort associated with learning how to use a system proficiently. To this end attempts are made to design for users and the tasks they perform through user centred design (ISO 13407). Neilson (1993) states that user characteristics and the variability in tasks are the factors which have the most important impact on usability. The requirement for user centred design implies a need to understand the types of users and tasks performed during information retrieval. This section discusses literature concerning user classification and user types within design.

This is followed by a review of task categorisations, and modelling user behaviour and its use in interface design.

2.7.1 User classification and user types

The importance of the user, and the types of users of a system, was emphasised in Hansen's (1971) list of user engineering principles and this was added to by Neilson in (1993), who declared that in order to set appropriate limits for interface complexity the characteristics of the user population must be known. This theme of designing for a type or category of user was re-iterated by Faulkner (1998) who noted that designers categorise users based on their characteristics and then design to support these characteristics. This promotes the idea that different classes of user have different system requirements. Novices may expend more cognitive effort in the use of systems because interaction involves knowledge level reasoning whereas experts have compiled procedures for specific situations (Rassmussen 1983). Novices require confirmation of success and more support. Experts, on the other hand, find long supportive user dialogues disturb patterns of activity. This is because experts organise interaction with the systems into procedural chunks reducing the low level detail. Faulkner synthesised the following general guidelines when designing systems for novices and experts. The following guidelines describe aspects of systems which assist novice behaviour whereas experts require the opposite.

1. All initiative should come from the computer
2. Each required input should be brief
3. Input procedures should be consistent with user expectations. The user should not be faced with anything that seems inconsistent or unlikely
4. Special training should not be necessary
5. All systems messages should be clear and unequivocal
6. User decisions should be made from a small set of options
7. Users should control the pace of interaction
8. User decision making should be in response to a specific request for action
9. There should be sufficient feedback: novice users like to know if they are on the right track

The definitions of novice and expert for information retrieval are a little unclear as knowledge of the task domain and appropriate strategy effect the interaction patterns and the facilities required as much as device knowledge.

2.7.2 General models of interaction and the user

Models of information retrieval (Bates 1989; Markey and Atherton 1978; Belkin 1993; Marchionini 1995; Ingwersen 1982, 1996) were discussed earlier in section 2.3. Here, literature concerned with modelling in HCI is addressed to indicate solutions and methods for modelling interaction, cognitive behaviour, the effects of user knowledge and the influence of the task. The relevant types of models used in HCI are general models of cognition, models representing user tasks and goals, and models of the interaction and dialogue.

The general models of cognition highlight the components necessary to address the causes and types of cognitive activity underlying user's behaviour e.g. Model of Human Processor (Card et al 1983) and Interacting Cognitive Subsystems (Barnard 1988). The Model of Human Processor provides a simple view of user-system interaction. The model is divided into: perceptual, motor and cognitive systems each consisting of memories, processors and principles of operation dictating behaviour dependent on context. The perceptual system processes sensory information from external sources to the user, while the motor system controls the user action on the external world. These two systems are connected by the cognitive system which models problem solving based on the sensory and memory information available, and outputs decisions causing the motor system to manipulate objects in the environment. The architecture highlights the need for a strong view of the environment, task and system, and cognitive processes in understanding behaviour. Interacting Cognitive Sub-systems (Barnard 1988) describes a resources architecture consisting of the following nine components acoustic, body state, visual, articulatory, limb, morphonolexical, implicational, propositional and object subsystem. The sub-systems all have the same structure comprising of inputs, outputs and memory stores. Cognitive activity is described as the processes required, knowledge and process attributes.

HCI offers techniques and languages to enable the description of the hierarchical nature of users tasks and goals e.g. Goals Operators Methods and Selection (Card et al 1983) and Cognitive Complexity Theory (Keiras and Polson 1985). The GOMS model provides descriptions of expert behaviour for routine tasks using a four component model. Goals describe the user's goals, operators describe actions to achieve goals, methods are composed of the operators and selection rules dictate the methods to be used. The concentration on expert and error free behaviour mean GOMS is unsuitable for modelling the IR task as retrieval is very diverse and dynamic. Cognitive Complexity Theory (Keiras and Polson 1985) enriches GOMS like models to provide more predictive power. The model uses descriptions of the users' goals and of the computer system. The model is able to model novice behaviour using rules which may fire in a given context. The use of production rules enables many plans to be represented. The model of the computer system takes the form of generalised transition networks. CCT is used to measure the complexity of the interface through the number of production rules in the CCT description.

Both of these goal hierarchy descriptions produce well defined plans based on structured goals. However the majority of information retrieval tasks tend to be reactive, dynamic and adaptive based on information encountered in the retrieval process. The goal hierarchy descriptions do not concentrate enough on the effects of systems output on behaviour. The inclusion of the effect of systems output is essential to modelling of the IR task which is fundamentally a reactive process, rather than a pre-ordained, complete and structured interaction sequence.

Finally to conclude this section the methods available to describe action sequences and interactive dialogue are discussed. The approaches to modelling discussed are Task Action Grammar (Payne 1984, Payne and Green 1986), situated action (Suchman 1987) and Normans model of interaction (Norman 1986). Task Action Grammar is used to model the user's interaction as a language in an attempt to assess the 'cognitive difficulty' of the interface. The language allows the generalisation of operations which have the same interaction structure. TAG includes a representation for information already known by the user which can be used to define the effect of user knowledge on the commands used. Harrison (1994) stated that the language was inadequate for the design of systems based around dialogue and thus its use in information retrieval is limited. Situated actions (Suchman 1987) provides a more reactive richer view of user interaction which is particularly suited to the IR task as it states that context is fundamental to the interactions observed. The use of the system, its information content and setting define the course of action providing a useful metaphor from which to view the models of observed IR behaviour discussed in section 2.3. The Model of action (Norman 1986) proposes seven stages of human computer action: forming the goal, forming the intention, specifying the action, executing the action, perceiving the systems state, interpreting the systems state, evaluating the outcome. The stages are incorporated into cycles of action and evaluation. This iterative, cyclic and dynamic view of interaction is particularly relevant to information retrieval.

2.8 Summary and conclusions

This chapter has highlighted the changing nature of the information seeking task and the reasons why performance is likely to progressively get worse unless it is possible to make retrieval activity more effective and efficient. The distributed nature of information resources (e.g. WWW), increases in the number of information access points, the diversification in the tasks for which electronic information systems are used and the removal of the librarian as the intermediary has caused a requirement for user centred systems which assist the retrieval process. Current models of the search process address the problems in IR at a high level of abstraction for a specific context but lack the necessary precision, detail and predictive power to predict behaviour and users' strategy. The models describe observable action and are a good starting point for understanding retrieval behaviour but they need extending, synergising and improving as they fail to explain action at process, query development and cognitive levels. This more detailed view of search activity is required due to the complex interactions and iterations in the search process. The models generally have a restricted view of the retrieval task that assumes an expert intermediary is present, whereas this has been superseded by the move to end-user retrieval and a diversification in the variety of situations in which IR is performed. Research in IR has identified many context factors, tactics and strategy types which can influence search behaviour but the current models of the search process appear to neglect some of these strategies and tactics. Empirical studies have shown that IR activity is complex, dynamic and interactive with many variables (especially the system) effecting activity. This leads to a necessity for the computational implementation of models if behaviour is to be predicted and the system is to be optimised for specific users and tasks. Attempts at replacing expert intermediaries with computer intermediaries have shown improved retrieval performance in terms of recall but unfortunately this is often at the expense of increased search times or more user effort. The expert intermediaries must be based upon more comprehensive models of the retrieval process and a mediation between models of expertise, models of actual behaviour and the search context. This all leads to the need for a comprehensive cognitive model of the search process which can account for the effects of different types of users, tasks, search strategies and system designs on retrieval behaviour.

Chapter 3

Quantitative studies of end-user information seeking

This chapter describes the results of studies on end user retrieval behaviour for a typical commercial retrieval system

Chapter three: Quantitative studies of end-user information seeking

3.1 Introduction

An empirical investigation of information retrieval was carried out to study variations in user behaviour, performance and to investigate reasons for sub-optimal searches. Focusing on performance and behaviour patterns enables the success of information retrieval systems in enhancing and improving the efficiency of user retrieval to be tested. The effects of user knowledge and task characteristics on activity, performance, need articulation, system facility usage and the strategies used are assessed by analysing users behaviour within the retrieval process for different categories of users on different types of IR tasks.

This chapter describes experimental studies of end-user performance and search activities for information retrieval sessions using the WinSPIRS interface for MEDLINE[1]. The chapter is organised as follows: section two describes the experimental design and the analytic method. This is followed in section three by performance results which covers recall, precision and time spent. This section concludes with analysis of query construction, strategies and the system's facilities used. User perceptions of the usability of the interface and the systems functionality are discussed in section four. The implications of the findings for producing an improved understanding of the IR process and the need for a sound methodology for their design are discussed in section five.

3.2 Experimental Method

3.2.1 Experimental design

Seventeen medical students (thirteen males and four females, aged between 24 and 26) who were three months away from their clinical final examinations took part in the experiments. The subjects' experience in the domain and information retrieval system was assessed by a pre-test questionnaire (Appendix 3a). The answers given were used to categorise subjects' profile of domain and device knowledge and to allocate individuals into either a group who had some experience in the use of

MEDLINE (experts) or novices. Three pilot subjects undertook the experiment before the trials were started, to refine the medical scenarios for the information required and to test the experimental procedure.

All subjects were given the following search tasks which were developed by a independent medical expert.

1. Please use the MEDLINE database to investigate the socio-economic reasons for increased failure rates on the oral contraceptive pill *(Experimental task code OG1),*
2. Please utilise the MEDLINE database to compare caesarean sections being performed in planned and emergency situations from the standpoints of: maternal and foetal safety, infection control, statistical and economic data *(Experimental task code OG2),*
3. Using the MEDLINE database please assess the importance of blood sugar levels and lipid profile in the cause of myocardial infarction within the male population *(Experimental task code PH1),*
4. Utilise the database to determine some areas of increased and decreased efficiency within the NHS since the introduction of formal clinical auditing *(Experimental task code PH2),*

The version of the MEDLINE system used for the experiments runs under MS Windows 3.11 running on a personal computer using the WinSPIRS interface, see figure 3.1 for the query interface.

[1] MEDLINE is the copyright of the U.S. National Library of Medicine and SilverPlatter

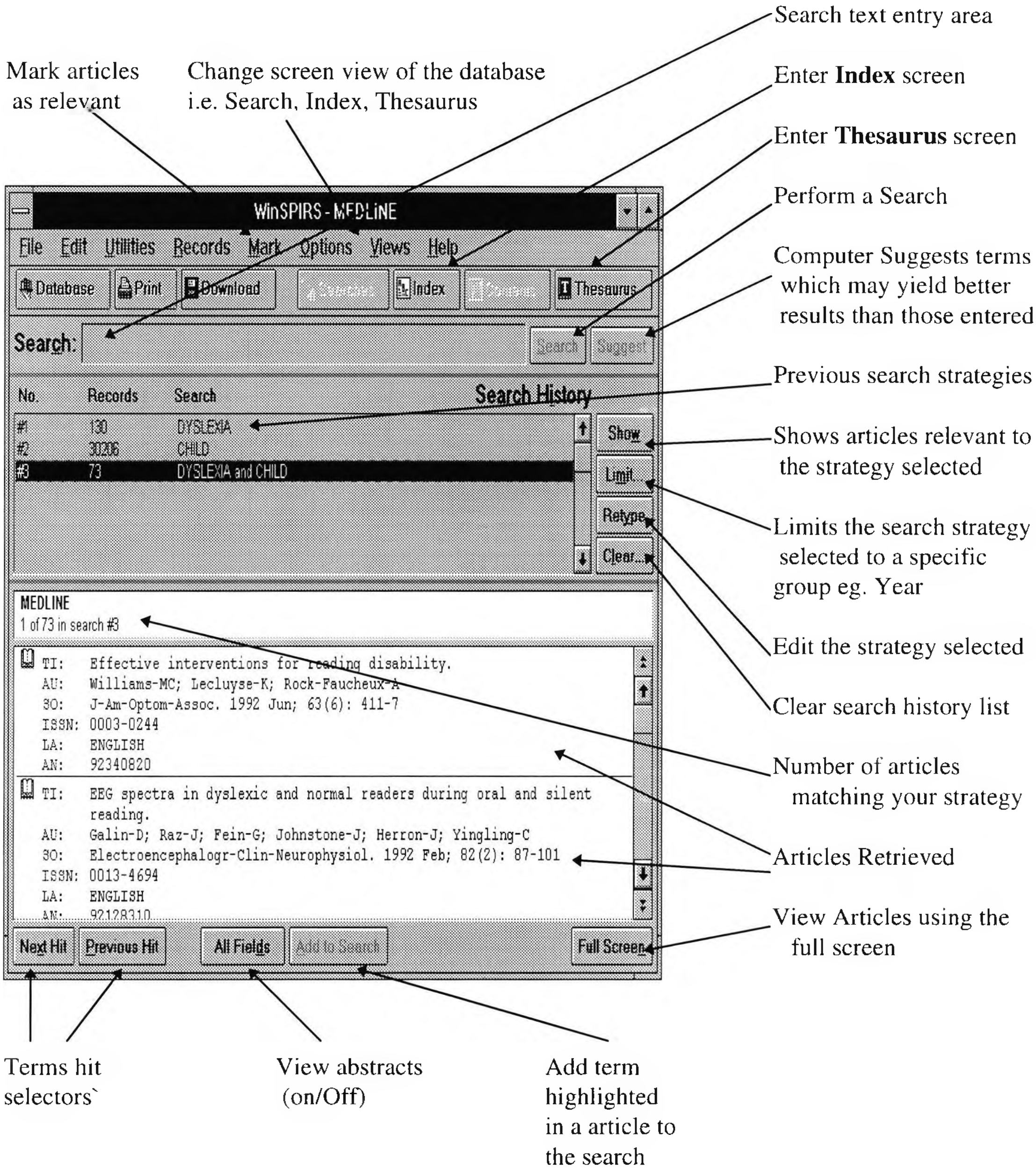

Figure 3.1. WinSPIRS search interface for MEDLINE showing a summary of retrieved results.

When questioned after the experiment, the subjects confirmed that the tasks were similar in style and complexity to ones they would normally use MEDLINE to solve. Subjects had access to the tasks throughout the trial and were instructed that they could record information as they searched. Each task was given to the subject after they had reported that the preceding task had been completed satisfactorily. The subjects' notes, physical actions, search strategies, views of the problem to be tackled and search history were recorded. The presentation of tasks was randomised to counteract any learning effects. At the end of the experiment the subjects were asked to rate the perceived difficulty of each of the tasks on a five point scale.

To ensure a high and uniform level of device knowledge for the expert group the ten subjects with experience using MEDLINE were given training in the use of the information retrieval system. The system facilities and brief guidance on the production of strategies were explained in a tutorial document (Appendix 3b) and a demonstration. The demonstration topic was learning disabilities and thus far removed from the domain of the experimental tasks. Novices were not trained. Subjects were requested to think aloud during the experiment and their verbal protocols, with associated physical actions, were recorded on video and audio tape. This chapter addresses the quantitative data from the users' behaviour, queries and retrieved results to ascertain the extent of any problems the users encountered while using retrieval interfaces.

3.3 Data analysis

Performance: Three performance measures were calculated for each task:

- Recall: the proportion of documents marked as relevant by the subject at any time during the session as a percentage of relevant records in the database as contained in a gold standard solution set.
- Precision: the proportion of the documents marked as relevant which were also judged to be relevant by an independent expert as a percentage of the records retrieved and marked as relevant by the subject. Note that precision was calculated from the documents the subject marked as relevant, not the overall number retrieved during querying.
- Search times: the total time spent searching and evaluating retrieved records. These times were extracted from the users' interaction with the system, the state of the system interface and user protocols.

Recall performance scores were calculated as a percentage of an ideal 'gold-standard' solution synthesised from all the subjects' searches and queries created as a collaboration between independent domain experts and a information retrieval specialist. Two independent medical experts were asked to assess the articles for relevance to the task scenarios. Differences between experts' document classification were discussed and an agreed cluster of relevant documents was produced. The reassessment of article relevance for the articles selected by the subjects, despite their expertise in the medical domain, was performed so as to eliminate irrelevant articles from the recall data as some subjects selected all the results of a particular query without assessing the articles for relevance.

Query construction: The subjects' queries were analysed to test the following hypothesis:

1. Better performers will construct better queries. This is measured by the number of different terms used, the number of terms shared with the gold standard and the number of incorrect terms used i.e. not in gold standard.
2. Better performers will be more diligent in querying. This is measured as the number of query iterations per task and total search time.
3. Better performers will use richer queries to express the information required. This is measured by the range of query terms employed and use of complex Boolean query syntax.

Strategies used: It is hypothesised that better performers would use more effective search strategies, as reported in the literature (Harter et al 1985, Marchionini 1995). The subjects' search histories and timing data were categorised to ascertain the strategy types employed.

Systems facilities used: It is hypothesised that the use of certain systems facilities may assist users in obtaining better performance. The analysis investigates which were the most popular facilities and which seemed to aid retrieval success.

3.4 Results.

3.4.1 Search performance.

Table 3.1 shows recall and precision results of the subjects for the four tasks. The subjects failed to retrieve much relevant information as their average recall was 13.94%, when compared to the gold standard. The gold standard queries represented a single search without any evaluation and gave an average recall of 76.66% of all the relevant articles possible, demonstrating that a high percentage of the relevant articles could be accessed if the need was articulated correctly. Expert searchers had, on average, better recall than novices, but some novices (eg. L6 and L7) had good recall scores (see table 3.1).

Subject	**OG1**		**OG2**		**PH1**		**PH2**	
(Experts)	*Recall (%)*	*Precision (%)*	*Recall (%)*	*Precision (%)*	*Recall (%)*	*Precision (%)*	*Recall (%)*	*Precision (%)*
H1	27	*60*	7	*33*	13	*85*	5	*100*
H2	9	*100*	24	*30*	5	*80*	9	*71*
H3	45	*71*	17	*63*	5	*57*	12	*47*
H4	9	*33*	21	*75*	1	*100*	5	*75*
H5	27	*38*	14	*21*	6	*71*	5	*38*
H6	36	*44*	21	*67*	4	*100*	5	*75*
H7	45	*83*	21	*67*	9	*53*	2	*14*
H8	27	*50*	3	*8*	11	*75*	5	*60*
H9	18	*67*	14	*67*	16	*88*	9	*56*
H10	18	*50*	31	*40*	1	*25*	5	*50*
Mean (H)	***26.1***	***59.6***	***17.3***	***47.1***	***7.1***	***73.4***	***6.2***	***58.6***
(Novices)								
L1	18	*100*	14	*44*	8	*100*	2	*20*
L2	9	*25*	17	*55*	22	*76*	9	*63*
L3	9	*50*	7	*12*	28	*77*	12	*50*
L4	9	*20*	10	*25*	6	*83*	2	*25*
L5	9	*100*	24	*58*	9	*80*	5	*100*
L6	64	*39*	7	*14*	12	*67*	5	*75*
L7	27	*100*	14	*14*	18	60	8	*71*
Mean (L)	***20.71***	**62**	***13.29***	***31.71***	***14.71***	***77.57***	***6.14***	***57.71***
Gold standard query	**72.72**	***72.72***	**83**	**39**	**63**	**75**	**87.93**	**76.11**
Relevant records available	**11 records**		**29 records**		**84 records**		**58 records**	

Table 3.1 Analysis of recall and precision performance by task and subject groups

The only significant difference in recall between the two groups was on task PH1 ($p < 0.05$, Mann Whitney U) when novices out-performed experts. There is no immediately apparent reason for the performance difference on this task; but on this task alone novices spent longer on a retrieval session than experts and the proportion of time spent evaluating articles was much higher in novices (74%) than experts (57%). In PH1 recall was positively correlated (Spearmans rank) with the users perception of task difficulty ($p < 0.005$), total time spent ($p < 0.05$) and time spent evaluating ($p < 0.05$). In PH1 it appears that searchers who were more diligent in the amount of time spent in the retrieval process, and evaluating the results reaped rewards in terms of improved recall. Correlations for the other tasks were not significant.

Friedman's and L Page's (non parametric ANOVA) tests were used to establish if similarities occurred in the ranking of searchers' performance across tasks. Friedman's test establishes if a significant inter-subject rank order of performance exists for the tasks while L Page's test determines the order of that ranking. The experts had significantly similar ranking of the tasks for recall ($p < 0.01$) in the order of PH2, PH1, OG2, OG1 ($p < 0.001$). It appears that this group found the public health (PH) tasks more difficult than the obstetrics and gynaecology (OG) tasks. The recall ranking for the tasks agreed with our independent medical experts view of the tasks' difficulty (PH2 harder than PH1 and OG2 harder than OG1). The novice subjects exhibited no significant ranking of recall performance by task, so it appears that they found all questions to be equally difficult.

There were no inter-group differences in any of the tasks for precision (Mann Whitney U). This is not surprising given both groups had expert knowledge of the medical domain and thus they had the same ability to select articles relevant to the information need from the results sets. Precision scores for both groups and the gold standard query were lower for task OG2, possibly due to the large number of qualifying conditions which complicated the basic query on caesarean sections.

As the performance results were inconclusive the time taken by the subjects for a retrieval session, querying and evaluating results were investigated (see table 3.2). Time querying and evaluating results were extracted from user protocols and their actions on the interface. It appears the majority of subjects spent similar amounts of time searching and evaluating irrespective of the task. There were no significant differences for retrieval time overall or evaluation time. However, novices spent longer evaluating results than experts (experts 9.32; novices 10.17 minutes), but experts spent longer on the whole retrieval session (average time per session: expert 16.48; novices 13.58 minutes). Significant differences occurred between the groups for the time spent querying on the two more difficult tasks (Mann-Whitney U OG2 $p < 0.005$; PH2 $p < 0.01$) and experts spent substantially longer querying than novices (average time querying per session: expert 7.16 minutes; novices 3.41 minutes). Search times by task did not reflect expected difficulty as OG2 had the longest average retrieval time (20.49 minutes) followed by PH1 (16.05 minutes), OG1 (13.19 minutes) and PH2 (11.35 minutes).

Subject	OG1			OG2			PH1			PH2			Totals		
	% Search /eval	*Total time mins*	*Recall %*	*% Search /eval*	*Total time mins*	*Recall %*	*% Search /eval*	*Total time mins*	*Recall %*	*% Search / eval*	*Total time mins*	*Recall %*	*Mean % Search / eval*	*Mean Total time mins*	*Mean Recall %*
H1	62: 38	15.17	27	72: 28	28.39	7	49: 51	13.9	13	55: 45	9.89	5	60:40	16.84	13.00
H2	52: 48	35.31	9	40: 60	19.72	24	52: 48	4.55	5	78: 22	7.66	9	56:44	16.81	11.75
H3	18: 82	17.04	45	33: 67	31.46	17	44: 54	36.68	5	31: 69	25.01	12	32:68	27.55	19.75
H4	25: 75	8.87	9	35: 65	13.41	21	35: 65	10.57	1	20: 80	8.62	5	29:71	10.37	9.00
H5	28: 72	6.94	27	35: 65	21.94	14	54: 46	7.24	6	24: 76	9.98	5	35:65	11.53	13.00
H6	34: 66	6.4	36	39: 61	22.27	21	17: 83	12.16	4	25: 75	5.45	5	29:71	11.57	16.50
H7	58: 42	29.31	45	50: 50	26.65	21	40: 60	11.33	9	65: 35	23.52	2	53:47	22.70	19.25
H8	51: 49	12.68	27	49: 51	19.26	3	53: 47	20.47	11	44: 56	29.87	5	49:51	20.57	11.50
H9	35: 65	5.28	18	45: 55	22.83	14	15: 85	21.08	16	14: 86	8.83	9	27:73	14.51	14.25
H10	43: 57	10.88	18	34: 66	19.59	31	66: 34	12.89	1	32: 68	5.14	5	44:56	12.13	13.75
Mean (H)	**41:59**	**14.79**	**26.1**	**43:57**	**22.55**	**17.3**	**43:57**	**15.09**	**7.1**	**39:61**	**13.4**	**6.2**	**42:58**	**16.46**	**14.18**
L1	20: 80	6.48	18	36: 64	13.93	14	25: 75	9.25	8	13: 87	11.92	2	24:76	10.4	10.50
L2	8: 92	7.04	9	36: 64	15.33	17	28: 72	21.56	22	10: 90	17.73	9	21:79	15.42	14.25
L3	50: 50	9.08	9	34:66	19.93	7	10: 90	18.13	28	33: 67	1.19	12	32:68	12.08	14.00
L4	40: 60	8.39	9	52:48	10.93	10	41: 59	6.59	6	20: 80	4.94	2	38:62	7.71	6.75
L5	29: 71	15.46	9	19: 81	23.98	24	31: 69	28.65	9	12: 88	9.15	5	23:77	19.31	11.75
L6	18: 82	16.12	64	31:69	16.14	7	25: 75	20.55	12	18: 82	7.81	5	23:77	15.16	22.00
L7	19: 81	13.81	27	16: 84	22.56	14	24: 76	17.29	18	21: 79	6.29	8	20:80	14.99	16.75
Mean (L)	**26:74**	**10.91**	**20.71**	**32:68**	**17.54**	**13.29**	**26:74**	**17.43**	**14.71**	**18:82**	**8.43**	**6.14**	**26:74**	**13.58**	**13.71**
All Subject mean	**34:66**	**12.85**	**23.41**	**38:62**	**22.05**	**15.29**	**35:65**	**16.26**	**10.91**	**29:71**	**10.91**	**6.17**	**34:66**	**15.02**	**13.94**

Table 3.2: Time spent searching and evaluating, and total session times for the four tasks. The subjects spending greater than 60% of the retrieval session evaluating articles are shaded.

On the whole the time spent searching were similar for subjects across tasks. There are occasional anomalies but these are to be expected due to subjects possibly losing their way during a search or encountering task specific difficulties. The percentage of the total retrieval time spent searching to evaluating showed significant inter-group differences on the public health tasks (Mann-Whitney U PH1 $p < 0.05$, PH2 $p < 0.05$ Experts spent a greater proportion of a retrieval session searching than novices). Novices, generally seemed to concentrate more on evaluation at the expense of query formulation as shown by the proportion of time spent searching to evaluating. All novices spent 60% or more of their total retrieval time evaluating records on all retrieval tasks, apart from L3 on task OG1. The novices L5, L6, L7 spent 65% or more of their total time on evaluation for all tasks as did L1 and L2 except on task OG2 (64%). Experts were less consistent; H3, H4, H5, H6, and H9 did favour evaluation (65% +) but not on all tasks. It appears that novices may be able to compensate for their lack of search strategy or device knowledge by concentrating on article evaluation.

Further investigations showed a correlation between time measures and task characteristics. The ratio of time spent querying to time spent evaluating was positively correlated with the complexity of users' queries on all search tasks (Spearmans rank: OG1 $p < 0.005$ OG2 $p < 0.005$ PH1 $p < 0.05$ PH2 $p < 0.05$). As the proportion of the total time spent on querying increases, the queries used become more complex.

3.4.2 Query construction

As an initial attempt to establish if the needs articulated by the two groups of subjects differed, the number of different terms used were analysed. Table 3.3 gives an overview of the use of query terms. There were no significant correlations between recall performance and either of the query construction measures which were number of different terms used, terms shared with the gold-standard and terms used which were not shared with the gold-standard. On average expert searchers used more terms in the queries submitted (9.1 terms) than novice users (6.6 terms) even though they had equivalent domain knowledge. There was a significant difference between experts and novices for the number of different terms used in queries for the two more complex tasks (experts used a > number of terms than novices for OG2 $p < 0.05$; PH2 $p < 0.05$ Mann-Whitney U). This result supports the argument that for complex tasks novice and expert users require differing task support; if only to accommodate and compensate for differences in the number of terms used during need articulation. However, the total number of terms used by each group was not significantly different for all tasks, although individuals did show a common ranking in the number of terms used per task (L Page's test PH2, PH1, OG1, OG2, $P < 0.001$ for both groups). This indicates individual searcher's reactions to each of the tasks characteristics were similar and they adjusted the number of terms needed to articulate the information need based on these characteristics.

Subjects	Percentage of terms used during retrieval occurring in the gold standard query		Percentage of the gold standard query terms used		Number of different terms used		Recall	
	Mean (%)	Range (%)	Mean (%)	Range(%)	Mean	Range	Mean	Range
(Experts)								
H1	71.19	66-75	33.27	26-45	15	10-26	13.00	5-27
H2	77.03	50-100	20.37	14-33	8.5	4-13	11.75	5-24
H3	80.16	71-100	17.89	17-19	7	5-9	19.75	5-45
H4	86.46	62-100	18.45	12-29	6.5	4-8	9.00	1-21
H5	84.72	66-100	19.41	18-22	7.25	5-9	13.00	5-27
H6	89.58	75-100	18.73	11-24	6.75	4-12	16.50	4-36
H7	67.82	61-81	27.28	21-32	12.25	11-14	19.25	2-45
H8	86.43	80-100	26.72	19-37	9.25	7-10	11.50	3-27
H9	88.84	75-100	23.38	11-26	8.25	4-14	14.25	9-18
H10	67.47	33-100	22.65	7-30	10.5	6-16	13.75	1-31
Mean (H)	*77.50*	*75-79*	*22.81*	*21-24*	*9.125*	*6-14*	*14.18*	*6-26*
(novices)								
L1	77.08	50-100	15.48	14-19	6.5	4-8	10.50	2-18
L2	85.63	62-100	14.48	4-21	5.75	1-9	14.25	9-22
L3	87.50	66-100	17.00	7-24	6.75	2-12	14.00	7-28
L4	63.69	50-71	11.24	7-15	5.5	3-7	6.75	2-10
L5	89.58	75-100	17.92	14-22	6.25	4-8	11.75	5-24
L6	60.49	50-71	17.00	14-19	9	6-13	22.00	5-64
L7	83.33	66-100	17.82	15-25	6.5	6-7	16.75	8-27
Mean (L)	*74.66*	*67-80*	*15.85*	*14-17*	*6.60*	*5-9*	*13.71*	*6-21*

Table 3.3 Query term usage for both groups summed across tasks. The first two columns give the terms used by the subject that were shared with the gold-standard as a % of all the terms they used; columns 3-4 give the terms used and shared with the gold-standard as a % of the total gold-standard terms.

The proportion of the terms used from the gold standard query was low (expert average 22.81%, novices 15.85%) and significant inter-group differences were found on the public health tasks (experts have a greater coverage of the gold standard query than novices PH1 $p < 0.05$, PH2 $p < 0.05$, Mann Whitney U). However, the percentage of terms used by most subjects that were shared with the gold standard was high (> 70%) apart from H7, H10 (67%) and L6 (60%). Experts expanded more of the components of the original need statement in queries than novices by adding terms from the system's thesaurus or from their knowledge of the domain (experts 60% of task elements explored, novices 37% of task elements expanded with synonyms or refined with sub-terms). On average experts produced 43% more queries (average 12.73 per task) than novices (average of 8.93 per task). The number of query iterations and the terms used that were shared with the gold standard were positively correlated (OG1 $p < 0.05$, OG2 $p < 0.01$, PH1 $p < 0.01$, PH2 $p < 0.01$ Spearman rank order correlation). Significant correlations were found between the number of terms searchers added to queries and the number of query iterations (OG1 $p < 0.001$, OG2 $p < 0.05$, PH1 $p < 0.001$, PH2 $p < 0.05$) so prolonged querying also appears to lead to richer and more complete queries.

The Boolean operators ('AND', 'OR', 'NEAR', 'IN', 'NOT' etc.) used in queries to express problem relationships are markedly different for the two subject groups. Novices consistently only used Boolean 'AND' relationships, whereas experts use more complex Boolean syntax. This confirms the findings of Sewell et al (1986) and Marchionini (1989) that for novice searchers the 'AND' relationship is used most commonly and users submitted multiple parallel searches instead of using 'OR' relationships. This may provide some of the reasons for the consistent use of successive term substitution strategies by novice subjects (table 3.13).

3.4.3 Query pattern analysis

Query logs were analysed by counting the number of searchers in a retrieval session who produced few (0-10 iterations), average (11-20 iterations) or many (>21 iterations) query iterations, and assessing query complexity as low (0-3 terms), average (4-7 terms) or high (>8 terms) based on the maximum number of terms used per query. This analysis demonstrates the different approaches between the novices and experts, see table 3.4. There were no significant correlations between recall performance and the number of query iterations. However, this analysis did demonstrate differences between novice and expert search strategies.

	Query iterations			Maximum terms/query		
	0-10	11-20	21+	0-3	4-7	8+
H1		✪ ✪	✪ ✪			✪ ✪ ✪ ✪
H2	✪ ✪ ✪		✪	✪ ✪	✪ ✪	
H3	✪ ✪	✪ ✪		✪ ✪ ✪	✪	
H4	✪ ✪ ✪ ✪			✪ ✪ ✪	✪	
H5	✪ ✪	✪ ✪		✪ ✪	✪ ✪	
H6	✪ ✪ ✪	✪			✪ ✪ ✪ ✪	
H7		✪	✪ ✪ ✪		✪ ✪ ✪	✪
H8		✪ ✪	✪ ✪		✪ ✪ ✪	✪
H9	✪	✪	✪ ✪	✪	✪ ✪	✪
H10	✪	✪ ✪ ✪		✪	✪ ✪	✪
Total	16	14	10	12	20	8

	Query iterations			Maximum terms/query		
	0-10	11-20	21+	0-3	4-7	8+
L1	✪ ✪ ✪	✪		✪ ✪ ✪ ✪		
L2	✪ ✪	✪ ✪		✪ ✪ ✪ ✪		
L3	✪ ✪	✪ ✪		✪ ✪ ✪ ✪		
L4	✪ ✪ ✪	✪		✪ ✪ ✪ ✪		
L5	✪ ✪ ✪ ✪			✪ ✪ ✪ ✪		
L6	✪	✪ ✪ ✪		✪ ✪ ✪ ✪		
L7	✪	✪ ✪ ✪		✪ ✪ ✪ ✪		
Weighted Total	23	17	0	40	0	0

Table 3.4: Query iterations and their complexity for novice and expert subjects analysed by task. The spatial distribution reflects the task order OG1, left hand side, OG2, PH1 middle locations, PH2, right hand side. The totals for novices are weighted to account for differences in the number of subjects in the novice group

The distribution of query iterations for the experts was fairly even; however the novices performed fewer iterations overall and never used more than twenty. The pattern is more striking for query complexity and inter-group differences were apparent (Mann Whitney OG1 $p < 0.05$, OG2 $p < 0.01$, PH1 $p < 0.01$, PH2 $p < 0.05$). Experts tended to use moderate to complex queries, whereas the novices all kept their queries simple. When individual differences are examined, four groups are apparent. Among the experts four subjects tended to use a low number of iterations with simple queries (H2-5). They shared this pattern with the novice subjects. Four experts may be characterised as using a high number of iterations with complex queries (H7-10), although H10 was less consistent on the high number of iterations. Finally one expert, H1, showed a high number of iterations with ultra complex queries.

3.4.3.1 Information searching strategies

Query logs were analysed to highlight patterns in the need articulation process. Analysis concentrated on cycles of narrowing, broadening and successive term substitution as illustrated in table 3.5. A subject was determined to be using successive term substitution if the queries used showed a pattern whereby a term relating to a particular concept is successively amended as subsequent queries are submitted. A subject was determined to be using narrowing cycles if a pattern of progressive search narrowing occurred. Observations of progressive narrowing were made if more specific terms were used, constraints were tightened, new concepts were added using Boolean 'AND' clauses or Boolean relationships between query elements were altered indicating narrowing (e.g. substituting 'AND' for 'NEAR'). A subject was determined to be using broadening cycles if a pattern of progressive query expansion occurred. Observations of progressive query expansion were made if more general terms were used, constraints were relaxed, new concepts were added using Boolean 'OR' clauses, or Boolean relationships between query elements were altered indicating broadening (e.g. substituting 'NEAR' for 'AND'). At a higher level of analysis assessments were made as to whether a searchers retrieval followed patterns of trial and error or could be characterised as favouring evaluation. A subject was determined to be favouring evaluation strategies if they spent 25% or greater of their total retrieval time on the uninterrupted evaluation of articles in between querying activity. A subject was determined to be using trial and error if the queries posed are characterised as having apparent logic jumps as the query is articulated. For example, if a subject was to submit myocardial infarction and males as a query and in the next query submit blood glucose and lipids then they would be determined to be using trial and error (or sub-goaling) as there is no direct link between the queries.

	Strategy types				
	Narrowing cycle	Broadening cycle	Successive term substitutions	Evaluation strategy	Trial and error
H1	✪ ✪ ✪ ✪	✪ ✪ ✪ ✪	✪ ✪ ✪	✪	✪ ✪ ✪ ✪
H2	✪ ✪ ✪ ✪		✪ ✪ ✪	✪	✪ ✪ ✪
H3	✪ ✪ ✪		✪ ✪	✪ ✪ ✪	✪ ✪ ✪ ✪
H4	✪ ✪		✪ ✪ ✪	✪ ✪ ✪	✪ ✪ ✪
H5	✪ ✪		✪ ✪ ✪ ✪	✪ ✪ ✪	✪ ✪
H6	✪ ✪ ✪	✪ ✪		✪ ✪ ✪	
H7	✪ ✪ ✪ ✪	✪ ✪ ✪ ✪	✪ ✪ ✪		✪
H8	✪ ✪ ✪ ✪	✪ ✪ ✪ ✪	✪ ✪ ✪	✪	✪
H9	✪ ✪ ✪ ✪	✪ ✪	✪	✪ ✪ ✪	
H10	✪ ✪	✪ ✪	✪	✪	✪ ✪
Total	32	18	23	19	20

	Strategy types				
	Narrowing cycle	Broadening cycle	Successive term substitutions	Evaluation strategy	Trial and error
L1	✪		✪ ✪ ✪	✪ ✪ ✪	✪ ✪ ✪
L2			✪ ✪ ✪	✪ ✪ ✪	✪ ✪ ✪
L3	✪		✪ ✪ ✪	✪ ✪	✪ ✪ ✪
L4			✪ ✪ ✪ ✪	✪ ✪ ✪ ✪	✪ ✪
L5	✪		✪ ✪ ✪ ✪	✪ ✪ ✪ ✪	✪ ✪ ✪
L6	✪ ✪ ✪	✪	✪ ✪ ✪	✪ ✪ ✪	✪ ✪ ✪
L7	✪ ✪	✪ ✪	✪ ✪ ✪	✪ ✪ ✪	✪ ✪ ✪ ✪
Weighted Total	11.42	4.3	32.4	31.4	30

Table 3.5: Strategy types used by searchers. The spatial distribution reflects the task order OG1, left hand side, OG2, PH1 middle locations, PH2, right hand side. The totals for novices are weighted to account for differences in the number of subjects in the novice group

Overall, about half of the expert subjects adopted a consistent search strategy across tasks, the more consistent subjects being H1, 7 and 8. Inspite of adopting this apparently effective strategy these subjects did not achieve better recall, so it appears that although there may be an 'expert' behaviour pattern, unfortunately, it is not necessarily always successful. Another point to note is that only a minority of subjects consulted the system thesaurus and that doing so did not lead to improved performance (see table 3.11). Most subjects followed more than one strategy. Expert searchers concentrated on cycles of narrowing and broadening; whereas novices favoured a trial and error approach by substituting the individual query terms until reasonable search results are produced. However, there were individual differences, for example, experts H2-5 did not use query broadening. This might reflect individual styles and search strategies. Novices made use of evaluation strategies more than experts. It is noticeable all the novices used evaluation strategies on the public health tasks (PH1, PH2) and all but one of the subjects used this approach for task OG1. However this was not observed in task OG2, possibly because this task promoted a richer query representation. In task OG1, complex queries are noticeably absent from novice subjects behaviour although iterative querying was adopted by three subjects, two of whom followed a narrowing strategy and scored well (L6, 7). Both groups substituted terms in queries, although novices favoured this strategy slightly more than experts. Sub-goaling to break a complex need into sub-parts was practised by both groups; however, it was less common for PH1, an easier task, but common for PH2, a difficult task.

To further investigate user strategies the querying and evaluation transitions are summarised in table 3.6. The table shows the number of bouts in subjects' activity and the average duration of bouts of searching and evaluating activity.

Subjects	*Number of search/ evaluation bouts*				*Average duration of search bouts (minutes)*				*Average duration of evaluation bouts (minutes)*			
(Experts)	**OG1**	**OG2**	**PH1**	**PH2**	**OG1**	**OG2**	**PH1**	**PH2**	**OG1**	**OG2**	**PH1**	**PH2**
H1	22	40	12	14	0.86	1.02	1.14	0.78	0.52	0.40	1.18	0.64
H2	54	24	6	15	0.68	0.66	0.79	0.75	0.63	0.99	0.73	0.24
H3	12	26	32	22	0.51	0.80	1.01	0.70	2.33	1.62	1.24	1.57
H4	16	20	16	13	0.28	0.47	0.46	0.25	0.83	0.87	0.86	1.15
H5	11	27	17	14	0.32	0.55	0.43	0.34	1.00	1.10	0.48	1.08
H6	8	26	6	14	0.54	0.67	0.69	0.19	1.06	1.04	3.36	0.58
H7	74	52	21	31	0.46	0.51	0.41	0.96	0.33	0.51	0.68	0.55
H8	18	28	32	33	0.72	0.67	0.68	0.77	0.69	0.70	0.60	1.05
H9	10	24	10	4	0.37	0.86	0.63	0.62	0.69	1.05	3.58	3.80
H10	22	24	12	6	0.43	0.56	1.42	0.55	0.56	1.08	0.73	1.17
Mean (H)	24.7	29.1	16.4	16.6	0.52	0.68	0.77	0.59	0.86	0.94	1.34	1.18
(Novices)												
L1	8	20	8	16	0.32	0.50	0.58	0.19	1.30	0.89	1.73	1.30
L2	2	16	14	8	0.56	0.69	0.86	0.44	6.48	1.23	2.22	3.99
L3	18	27	6	4	0.50	0.48	0.60	0.20	0.50	1.01	5.44	0.40
L4	12	16	10	4	0.56	0.71	0.54	0.49	0.84	0.66	0.78	1.98
L5	14	18	18	10	0.64	0.51	0.99	0.22	1.57	2.16	2.20	1.61
L6	18	24	26	12	0.32	0.42	0.40	0.23	1.47	0.93	1.19	1.07
L7	16	18	16	8	0.33	0.40	0.52	0.33	1.40	2.11	1.64	1.24
Mean (L)	12.57	19.86	14	8.86	0.46	0.53	0.64	0.3	1.94	1.28	2.17	1.66

Table 3.6: Activity transition data for the four experimental tasks.

Table 3.6 shows that experts appear to change between querying and evaluating activity many more times than novices irrespective of the task. These differences are less marked on task PH1. The bouts of searching behaviour produced by experts appear to be longer than those of novices; whereas the bouts of evaluation observed in the behaviour of novices tend to be much longer than experts. Differences in the number of transitions between searching and evaluating are significant for task OG2

(P< 0.05 Mann Whitney U experts produce a greater number of transitions than novices), and on PH2 differences in the average duration of bouts of search activity are significant (P<0.05 experts have longer bouts of searching than novices).

The sequence and structure of bouts of searching and evaluating activity of individual searchers are summarised as follows (see appendix 3c for actual transition data). Novices tend to follow evaluation based strategies irrespective of the task i.e. a small number of searching bouts interspersed with long bouts of evaluation . The pattern for experts is more complicated. Subjects H1, H7 and H8 follow a progressive query-evaluate strategy on tasks OG1, OG2 and PH2; but H1 and H7 followed an evaluation based strategy on task PH1. In general subjects H3, H5, H6, H9 and H10 follow evaluation based strategies on all tasks, the exceptions to this general trend are H3 and H10 on PH1 and H9 on OG2 where mixed strategies are followed.

3.4.3.2 Analysis of query patterns highlighting strategies

The patterns observed in the query histories were examined in an attempt to categorise subjects according to the characteristics of the queries used in a retrieval session. The analysis grouped the activity of each of the subjects on each task into one of four patterns of query activity.

The pattern that represents the first (simple queries, few iterations) and second groups (simple queries, many iterations) is shown in figure 3.2. This illustrates the pattern for a novice group 1 subject, L1 who achieved reasonable results (18% recall for task OG1) with few iterations and simple queries. There is little evidence of a coherent strategy such as narrowing the query by adding more terms or substituting synonyms; instead, this group seem to tackle the problem by trial and error. The second group showed the same query pattern but with more iterations and an example is L2 in task PH1 (see Appendix 3d of all searchers query histories)

Request No.	Number of records retrieved	Query
1	13	ORAL CONTRACEPTIVES AND FAILURE
2	0	ORAL CONTRACEPTIVE AND SOCIO-ECONOMIC
3	1	CONTRACEPTIVE PILL AND SOCIAL
4	7	ORAL CONTRACEPTIVES AND ECONOMICS

Figure 3.2: Query log for a novice searcher (L1) who performed few queries which were simple on task OG1.

The third group were expert subjects who submitted moderately complex queries but with few iterations. A typical example is H2 for task OG2, who was relatively successful in achieving 24% recall (figure 3.3). There is some evidence of query narrowing, use of alternative terms and consulting the system thesaurus.

Request No.	Number of records retrieved	Request
1	0	ELECTIVE and EMERGENCY and CAESARIAN SECTION
2	4	ELECTIVE and EMERGENCY and CAESAREAN SECTION
3	19	CAESAREAN SECTION AND INFECTION
4	13	#3 and (ENGLISH in LA)
5	108	explode "CESAREAN-SECTION"/ adverse-effects , economics , mortality , statistics-and-numerical-data
6	64	#5 and (ENGLISH in LA)
7	5	#6 and ELECTIVE
8	4	#6 and EMERGENCY
9	1	#6 and COST
10	0	#6 and EFFICIENCY

Figure 3.3: Query log for a novice searcher (H2) who performed submitted moderately complete queries but with few iterations on task OG2. The # denotes the re-use of a previous query.

Finally, the pattern for group 4 experts who submitted complex queries with many iterations is illustrated in figure 3.4, by subject H8 in task PH1. This subject tried three cycles of narrowing, use of alternative terms and more complex Boolean operators, although this only achieved a moderate performance of 11% recall.

Request No.	Number of records retrieved	Request
1	16	CLINICAL AUDIT
2	16	CLINICAL AUDIT
3	18	CLINICAL AUDIT*
4	0	#3 and NHS
5	93	NATIONAL HEALTH SERVICE
6	0	#3 and #5
7	3225	EFFICIENCY
8	4	#7 and #5
9	0	#8 and #3
10	3225	EFFICIENCY*
11	94	NATIONAL HEALTH SERVICE*
12	273	NHS*
13	24	#11 and #12
14	343	#11 or #12
15	9	#14 and #10
16	0	#5 and #3
17	334	MEDICAL AUDIT
18	0	#17 and #15
19	11	#12 and AUDIT
20	5407	DEFICIENCY
21	6	#20 and MEDICAL
22	7	#20 and AUDIT

Figure 3.4: Query log for a novice searcher (H8) who performed complex queries with many iterations on task PH1.

Analysis of experts' strategies

The patterns produced by experts were investigated for intra-subject variations across the tasks. H1 was a 'super expert' in terms of a consistent strategy of constructing complex queries with many iterations, but achieved a poor performance overall. The approach proved reasonably successful for OG1 (27% recall) and PH1 (13% recall), but achieved poor performance on the other tasks (OG2 7% and PH2 5% recall). Possible causes of these performance differences, given similar search characteristics in terms of query iterations and their complexity, can be found in the time measures. In OG2 this subject spent a greater proportion of the retrieval session querying (72%) than on the other three tasks (this subject's across task average 60%). It is possible the large number of terms for OG2 may have caused this subject to neglect the careful evaluation of results. On task PH2 the subject spent a very short time on the retrieval session compared to the other tasks (9.89 min against this subject's average of 16.84 min) and so he may have just given up. H2 used complex queries and many iterations on the first task then reverted to a 'novice' group 1 style of fewer iterations and simpler queries for the last three tasks with reasonable results, relative to the other subjects, on OG2 (24% recall) and PH2 (9% recall). Good choice of search terms seemed to account for this subject's success on OG2 and PH2. On OG1 it appears that poor choice of terms seems to have caused the poor performance for this subject as her query history showed that 65% of the queries submitted produced less than two articles (see appendix 3b for query history). H3 constructed simple queries with few iterations but spent more time evaluating the results on tasks OG1, OG2 and PH2, with reasonable success (recall OG1 45%, OG2 17%, PH2 12%). On task PH1 the same strategy was followed with more iterations and less evaluation (54% of total time spent evaluating compared to this subjects average of 68% of total time spent evaluating) and this achieved a poor result. Subject H4 was similar to group 1 novices on all tasks, and only achieved a good performance on OG2 (21% recall) with more iterations and a longer evaluation time. In contrast H5, used a novice pattern on task OG1 and PH2, but achieved good results only on the former (27-5% recall respectively), even though more time was spent on evaluation and the overall search duration was longer for PH2. Longer evaluation time and complex queries with many iterations may account for a better score on task OG2 (14% recall), but the same query pattern without the longer evaluation time in PH1 met with little success. H6 followed a novice strategy for tasks OG1 (36% recall), PH1 and PH2 with little success on the latter two. Complex queries, more iterations and a longer evaluation time brought reasonable success on OG2 (21% recall). H7 was one of our true, group 4, experts who followed a pattern of complex queries, with several cycles of narrowing and broadening with many iterations. This worked well for OG1 (45% recall) and OG2 (21% recall) but not for the PH tasks on which he spent less time evaluating the results. H8 was similar but was only successful on OG1 (27% recall). H9 was a hybrid, group 3, expert who used a complex strategy for OG2 with only moderate success (14% recall) and simpler strategies on the other three tasks, with reasonable results for OG1 (18% recall) and PH1 (16% recall). Finally H10 showed a consistent expert pattern; however, once again the strategy worked for task OG1 (18% recall) and OG2 (31% recall) but not for the PH tasks in which he spent a shorter time evaluating.

Analysis of novices' strategies

Novices were more consistent in their strategies across tasks. L1 illustrates the typical pattern of simple queries, few to moderate iterations and longer evaluation times than experts, with reasonable success on task OG1 (18% recall) and OG2 (14% recall), but poor results on PH1 and PH2. It was noticeable that although proportionately more time was spent evaluating results in the PH tasks, the overall retrieval time was short. L2 used many iterations with some evidence of narrowing cycles on tasks OG2 and PH1 achieving good scores (17% recall and 22% recall). On tasks OG1 and PH2 the proportion of the retrieval time spent querying (8-10% of the total time) was much lower than on the successful tasks (28-36% of the total time). L3 followed a similar pattern for task OG1 and OG2, but with poor results; then presented an anomaly by using simple queries with 2-3 iterations to achieve relatively good results on PH1 (28% recall) and PH2 (12% recall). Evaluation time may be the answer on PH1, but in PH2 it seems that choice of correct keywords resulted in a reasonably good score. L4 behaved consistently with a novice searching pattern and achieved indifferent results throughout, but also spent less time evaluating results than other novices. L5 evaluated for a good proportion of the time throughout his retrieval sessions but only scored well on OG2 (24% recall), when choice of correct terms appeared to have made the difference. Subject L6 came close to the hybrid group 2 pattern and spent longer evaluating results but was only successful in OG1 (64% recall) in which good choice of terms and cycles of narrowing appear to have brought rewards. Finally L7, showed a similar pattern with long evaluation times and narrowing/broadening strategies which brought reasonable rewards on the first three tasks (OG1 27% recall, OG2 14% recall, PH1 18% recall). Shorter retrieval and evaluation times on PH2 resulted in poor performance. Generally novices may have given up on the harder PH tasks.

At the task level some differences in performance were associated with query patterns and strategies, although these could not be tested statistically.

Task OG1, Recall means 26.1 (experts), 20.7 (novices)
For experts there seems to have been two routes to success on this task. Out of the six good expert performers, two (H1 and H7) used long retrieval times, many query iterations, a large number of terms in query expressions and a high coverage of the gold standard solution. The other group of good performing experts (H3, H5 and H6) showed poor query quality and term articulation but spent a large proportion of the retrieval session evaluating results and so careful evaluation may be another route to success. The final expert who performed well (H8) had retrieval behaviour which did not excel in any of the measures, so this subject's performance can only be accounted for by good choice of terms. The poor performing experts show different reasons for failure. H2 followed a reasonable pattern of retrieval on most measures but failed to spend an adequate proportion of the retrieval time evaluating results (48% of total time) and term choice was poor. H4 on the other hand, had poor query quality and term articulation scores. Distinguishing reasons for the relative performance of novices is a little harder due to the similarities between their retrieval behaviour. The two novices with relatively good performance (L6 and L7) did perform more query iterations than other novices but apart from that there are no other differences as all novice subjects exhibited poor query quality and term articulation. It appears that for this task longer evaluation time and iterative querying were the determinants of success. Unfortunately these are not reliable indicators as H2 achieved poor results with many query iterations but poor term choice.

Task OG2, Recall means 17.3 (Experts), 13.29 (novices)
The two good, expert performers constructed complex queries, but their query iterations and evaluation time were not exceptional. Subject L5 achieved a high score because he spent more time evaluating articles making careful relevance judgements; however, that was not so for subject H4. One successful novice subject (L5- 17% recall) had a high number of iterations and a high evaluation time, although L2 (21%) did not.

The majority of poor performers all suffered from defective queries as indicated by either inappropriate terms, or incorrect terms and query syntax (H8, L3 and L6), however, the poor query quality was not always apparent from the quantitative indicators. The other subject performing poorly (H1) submitted many complex queries using a wide variety of terms, but seems to have been tripped up by using several terms which were not in the gold standard and failing to spend a large enough proportion of the retrieval session evaluating the articles retrieved (28% compared to the task average for experts of 57% or this subject's average across the experimental tasks 40%).

Six experts' performance was relatively good. Half of the good performers (H2, H3 and H4) exhibited poor query construction and articulation measures but appear to have been able to compensate for this by spending a high proportion of their retrieval session evaluating results. The remainder of the good performers (H10, H6 and H7) exhibited characteristics associated with more effective retrieval (high evaluation times and high numbers of different terms in queries). Out of the two poorly performing experts, H8 showed poor coverage of the gold standard solution (<20% of gold standard terms used) and used few terms in query articulation (<4 different terms) and H1 seemed to fixate too much on querying (72% of total retrieval session time) at the expense of careful evaluation. Again for novices causes of performance differences on this task are less easy to spot but the success of the top two novices are explained as follows: L2 seemed to be successful due to the high number of terms used in query articulation and L5's success is attributed to a longer evaluation time.

Task PH1, Recall means 7.1 (experts), 14.71 (novices)

Performance for this task was poor and novices out performed experts, so only the top two novice subjects can really be considered to have been partially successful. Two of the top four novice performers (L3 and L7) spent more time evaluating results so this appears to be an alternative strategy for effective retrieval. The better performing experts used a high iteration strategy (H1), long evaluation and complex queries (H9), while the novices used high evaluation time and iteration (L2) or high evaluation time (L3). However, the best overall performance by L3 was achieved with few iterations and relatively poor queries. In this case it seems that to achieve modest levels of performance the subject had only to choose a few, correct keywords and spend 90% of the retrieval session evaluating articles.

Recall success was correlated to the time spent evaluating articles for task PH1. Two of the three experts who performed relatively well (H8 and H9) had long retrieval times, high query iterations and good scores on term articulation measures. H1 on the other hand used complex and well articulated queries to achieve recall success. Two of the three poorly performing experts (H4 and H6) used a low number of query iterations and poor query articulations. H10 neglected article evaluation causing his poor recall performance. The four novices who attained relatively good performance scores all had longer retrieval times and spent a high percentage of that time evaluating articles.

Task PH2, Recall means 6.2 (experts), 6.1 (novices)

The effect of strategies may have been masked on this task as performance was so poor that none of the subjects can be considered to have been successful. Two of the better performers (H3, L2) had high evaluation times, but the converse was true of H2 and L3. The query construction indicators showed no pattern, indeed, L3 achieved the top equal score in this task with a poor query, few iterations and little evaluation. In contrast a poor performer, H7, did all the right things and submitted many complex queries with a variety of terms for a miserable 2% recall. This subject may have fallen foul of using several inappropriate terms. The other poor performers all submitted few iterations and used few query terms.

In conclusion some strategies point to success but only for the OG1, OG2 and PH1 tasks. For the experts more complex queries and narrowing cycles seem to produce reasonable performance; whereas in novices, more query iterations and longer evaluation times produced good results. However, for tasks PH2, these explanations do not always hold, and choice of appropriate search terms may account for success. When query patterns were examined by task it appears that PH2 gave most problems as it received fewer iterations and simpler queries from both groups. This task may have been too difficult and subjects just gave up. The distribution for the other three tasks was reasonably even.

3.4.4 Term use analysis

The terms in users' queries were matched against gold standard queries to identify which terms may be linked to recall success. This analysis identifies the information components in the original need which searchers are and are not using in their queries.

3.4.4.2 Term analysis of task OG1

In task OG1 three terms "oral contraceptive, failure and socio-economic" were used by at least ten out of the seventeen subjects. Thereafter variations on oral contraceptive pill were the next most popular terms.

A full analysis of the queries used by individual subjects can be found in table 3.7. Queries from the top six subjects showed little overlap in the terms used. In this task iteration and evaluation of results appear to have been the major determinants of success. Of the top performers, H7 used complex queries with 11 gold standard terms whereas H3, L6, and L7 all submitted simpler, 4 term queries. Moreover, H3 performed few iterations, so choice of the terms "oral contraceptive pill, failure rates, failure, and accidental pregnancy" and long evaluation times may have been the key factor in this subject's good performance (recall 45%). In the category of terms relating to the socio-economic reasons for the failure of the contraceptive pill all the novices only used one term (socio-economic) whereas in general experts tended to use alternative expressions as well as 'socio-economic'. The top expert performer (H3) didn't use any terms relating to socio-economic as he instead focused on alternative expressions for the failure of the pill. The other experts who performed better (H7, H6, H1, H5, H8) all used multiple term alternatives for oral contraceptive pill. Out of these five searchers all bar one (H6) used multiple term alternatives for either the pill's failure or socio-economic reasons. In general it appears that in order to achieve a good performance the term oral contraceptive must be used along with multiple terms representing either socio-economic factors or failure of the pill.

Terms	H3	H7	H6	H1	H5	H8	H9	H10	H2	H4		L6	L7	L1	L2	L3	L4	L5
Oral contraceptive pill	✓		✓			✓	✓			✓					✓			✓
OCP		✓																
Oral contraceptive		✓	✓	✓	✓	✓	✓					✓	✓	✓		✓		
oral contraception		✓		✓				✓	✓			✓				✓		✓
contraceptive pill					✓	✓				✓				✓		✓	✓	
Combined oral contraceptive																		
Combined oral contraception																		
failure rates	✓	✓														✓		✓
fail							✓											
failed		✓					✓											
failure	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓		✓	✓	✓
accidental pregnancy	✓								✓				✓				✓	
accidental pregnancies					✓													
unplanned pregnancy																		
unplanned pregnancies																		
socio-economic reasons		✓																
educational level																		
educational status																		
explanation				✓														
explanations				✓														
socio-economic		✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	
epidemiological factors									✓									
social class		✓		✓		✓			✓	✓								
social status																		
marital status																		
reasons		✓		✓														
income		✓							✓									
Recall %	**45**	**45**	**36**	**27**	**27**	**27**	**18**	**18**	**9**	**9**		**64**	**27**	**18**	**9**	**9**	**9**	**9**

Table 3.7: Term usage of subjects for task OG1. Subjects are ranked by recall within the expert/ novice groups. The divisions between the different concepts are indicated by the double line.

The gold standard query (see figure 3.5) had three main concepts - oral contraceptive pill, increased failure rate and socio economic factors. Socio economic factors in particular was a difficult concept to articulate because there are many causes of socio economic differences and a large number of synonyms are required to achieve an effective query.

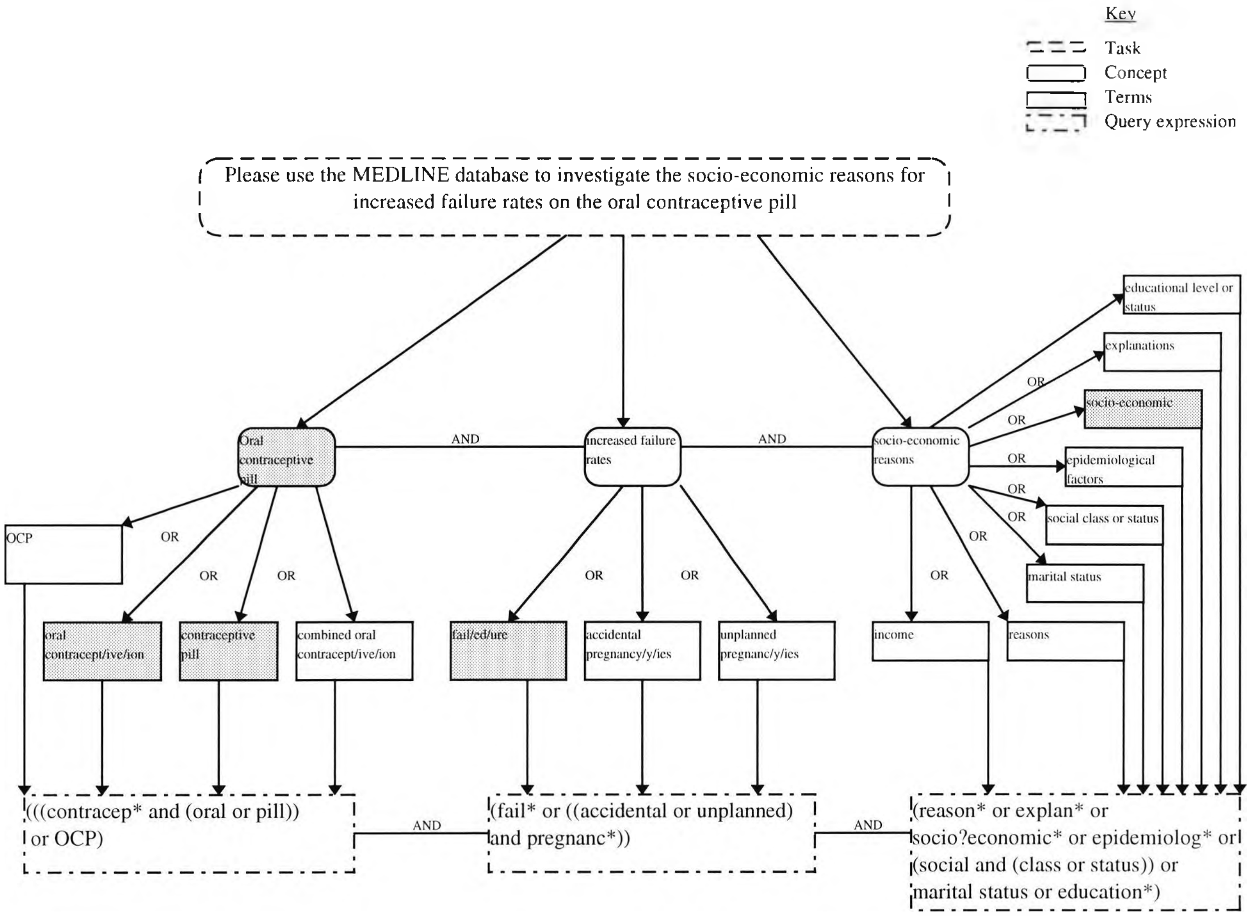

Figure 3.5: Choice of search terms for task OG1. The shaded boxes represent the 5 most popular choice of search terms.

The task is given at the top and decomposed into concepts at the next level: oral contraceptive pill, increased failure rates and socio-economic reasons. Concepts are refined into search terms in level two. The search terms are translated into the query syntax in level three showing how the gold standard query is constructed in terms of Boolean and system operators.

3.4.4.2 Term analysis of task OG2

In OG2, the four more common terms shared by 14 or more subjects were "caesarean section, emergency, safety and infection". This task had a widely spread distribution of terms (Figure 3.6).

While the top six subjects shared the four common terms (see table 3.8), so did many other subjects so success in this task was mainly accounted for by good query construction and iterate/evaluate cycle. The exception to this pattern, H4, used a simple, 4 term query "caesarean section, planned, emergency, economics" that achieved reasonable recall of 20.7%. Expert searchers all use either 'planned' or 'elective' to define the type of caesarean procedure used whereas only the better performing novices referenced this aspect of the problem at all.

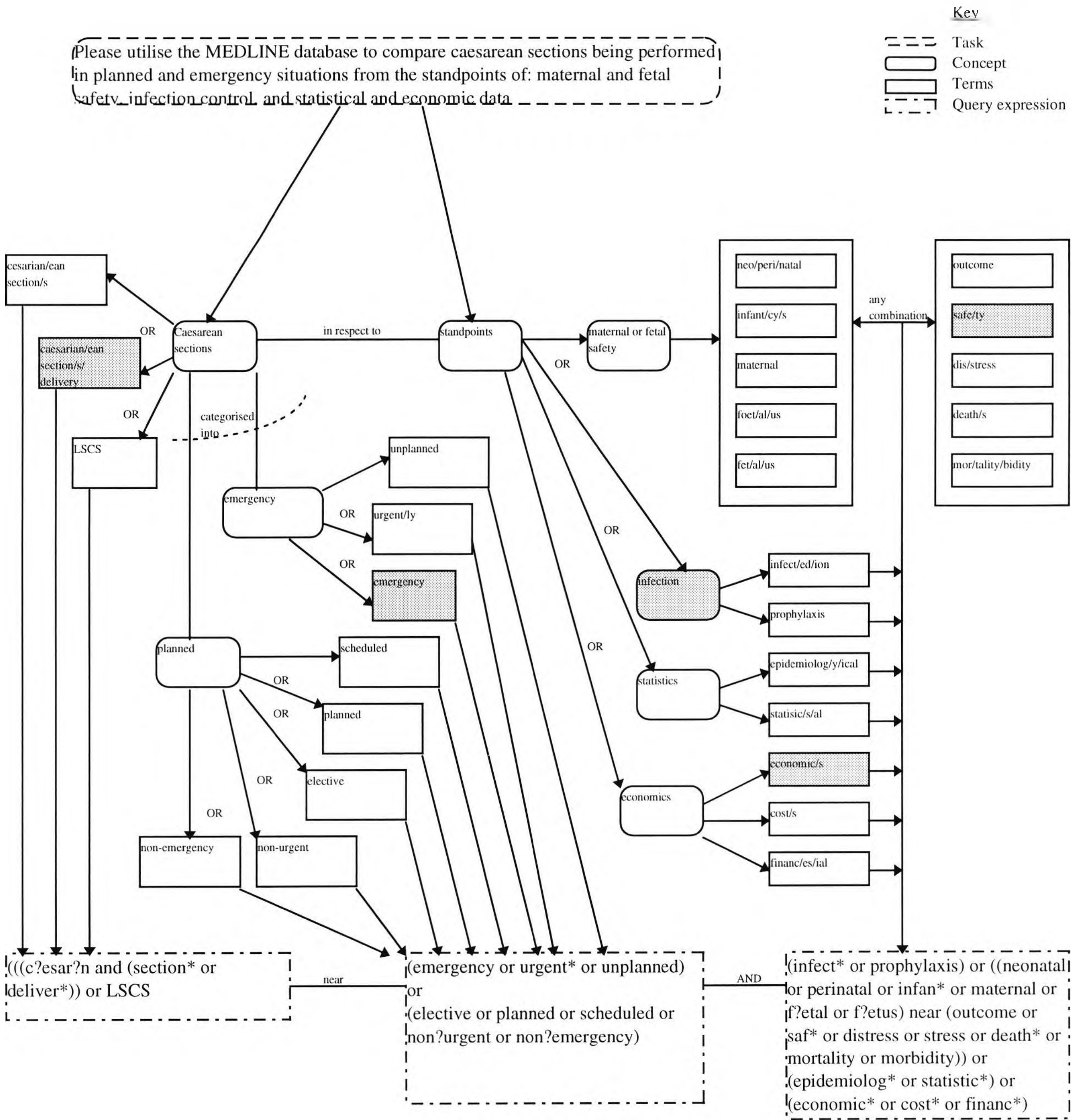

Figure 3.6: Choice of search terms for task OG2. The shaded boxes represent the 5 most popular choice of search terms.

The gold standard query has four concepts with the standpoints requiring a large number of synonyms for effective searches. The figure (3.6) breaks the gold-standard into its four component layers.

Terms	H10	H2	H4	H6	H7	H3	H5	H9	H1	H8		L5	L2	L1	L7	L4	L3	L6
cesarian section				✓				✓	✓			✓					✓	
cesarean section				✓													✓	
caesarian section			✓	✓		✓	✓	✓								✓	✓	
caesarean section	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
LSCS									✓									
scheduled																		
planned	✓		✓	✓	✓			✓	✓	✓		✓						
elective	✓	✓				✓	✓	✓	✓				✓	✓				
non-urgent																		
non-emergency																		
emergency	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓		✓
urgent									✓									
unplanned								✓										
neonatal	✓																	
perinatal																		
infants																		
infancy																		
maternal	✓				✓		✓			✓		✓	✓	✓	✓			
foetal	✓				✓		✓						✓		✓			
foetus										✓								
fetal	✓				✓							✓	✓					
fetus		✓								✓								
outcome				✓	✓									✓			✓	✓
safe				✓														
safety	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
dis-stress	✓																	
stress													✓					
death																		
mortality	✓					✓												✓
morbidity	✓																	
infected																		
infection	✓	✓		✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
prophylaxis																		
epidemiology									✓									
epidemiological									✓									
statistics						✓			✓					✓	✓		✓	
statistical					✓				✓				✓					
economic		✓			✓		✓	✓	✓	✓			✓		✓			✓
economics			✓	✓	✓	✓	✓	✓	✓	✓							✓	
costs		✓				✓		✓	✓								✓	
finances									✓									
financial									✓									
Recall %	**31**	**24**	**21**	**21**	**21**	**17**	**14**	**14**	**7**	**3**		**24**	**17**	**14**	**14**	**10**	**7**	**7**

Table 3.8: Term usage of subjects for task OG2. Subjects are ranked by recall within the expert/ novice groups. The divisions between the different concepts are indicated by the double line.

3.4.4.3 Term analysis of task PH1

The three more popular terms for task PH1 were "myocardial infarction, lipids, and blood glucose" with "male and blood sugar" chosen by over half of the subjects. Analysis of the queries used by individual subjects are summarised in table 3.9. The best performer, L3, used a four term query that added "lipid profile" to the top three terms, and achieved success without many iterations. Interestingly the next two top performers (both novices- L2 & 7) also added the same term. This shows how choice of a single term can help performance. The majority of differences in the terms used by expert searchers occur in relation to lipid profile and blood glucose as myocardial infarction and male are used by most subjects. The better expert searchers (H9, H1, H8) all used the following pairs of terms to articulate lipid profile and blood sugar respectively: lipid and lipids; blood sugar and blood glucose. Unfortunately this observation does not hold for novice searchers. In general novices appear to exclude the male component of the task in their queries.

Terms	H9	H1	H8	H7	H5	H2	H3	H6	H4	H10		L3	L2	L7	L6	L5	L1	L4
myocardial infarct										✓								✓
myocardial infarction	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
MI		✓		✓														
M.I				✓														
infarction																		
male	✓	✓	✓	✓	✓	✓		✓		✓				✓				
males		✓	✓		✓		✓								✓			
men	✓									✓							✓	
man																		
sex																		
lipid profile			✓	✓			✓					✓	✓	✓				
lipid	✓	✓	✓	✓	✓			✓		✓								
lipids	✓	✓	✓			✓	✓	✓	✓			✓			✓	✓	✓	✓
lipoprotein																		
lipoproteins															✓			
lipofuscin																		
fatty acids																		
sterols																		
HDL																		
LDL																		
cholesterol levels													✓					
blood sugar levels			✓										✓			✓	✓	
hyperglycemia										✓						✓		
hypoglycemia										✓						✓		
glucose tolerance																		
blood glucose	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓		✓			✓
blood sugar	✓	✓	✓	✓	✓				✓	✓				✓	✓			
Recall %	**16**	**13**	**11**	**9**	**6**	**5**	**5**	**4**	**1**	**1**		**28**	**22**	**18**	**12**	**9**	**8**	**6**

Table 3.9: Subjects term usage on task PH1. Subjects are ranked by recall within the expert novice groups. The divisions between the different concepts are indicated by the double line.

The gold standard query has four concepts with 'blood sugar levels' and 'lipid profile' requiring a large number of synonyms for effective searches. The figure (3.7) breaks the gold-standard into its four component layers.

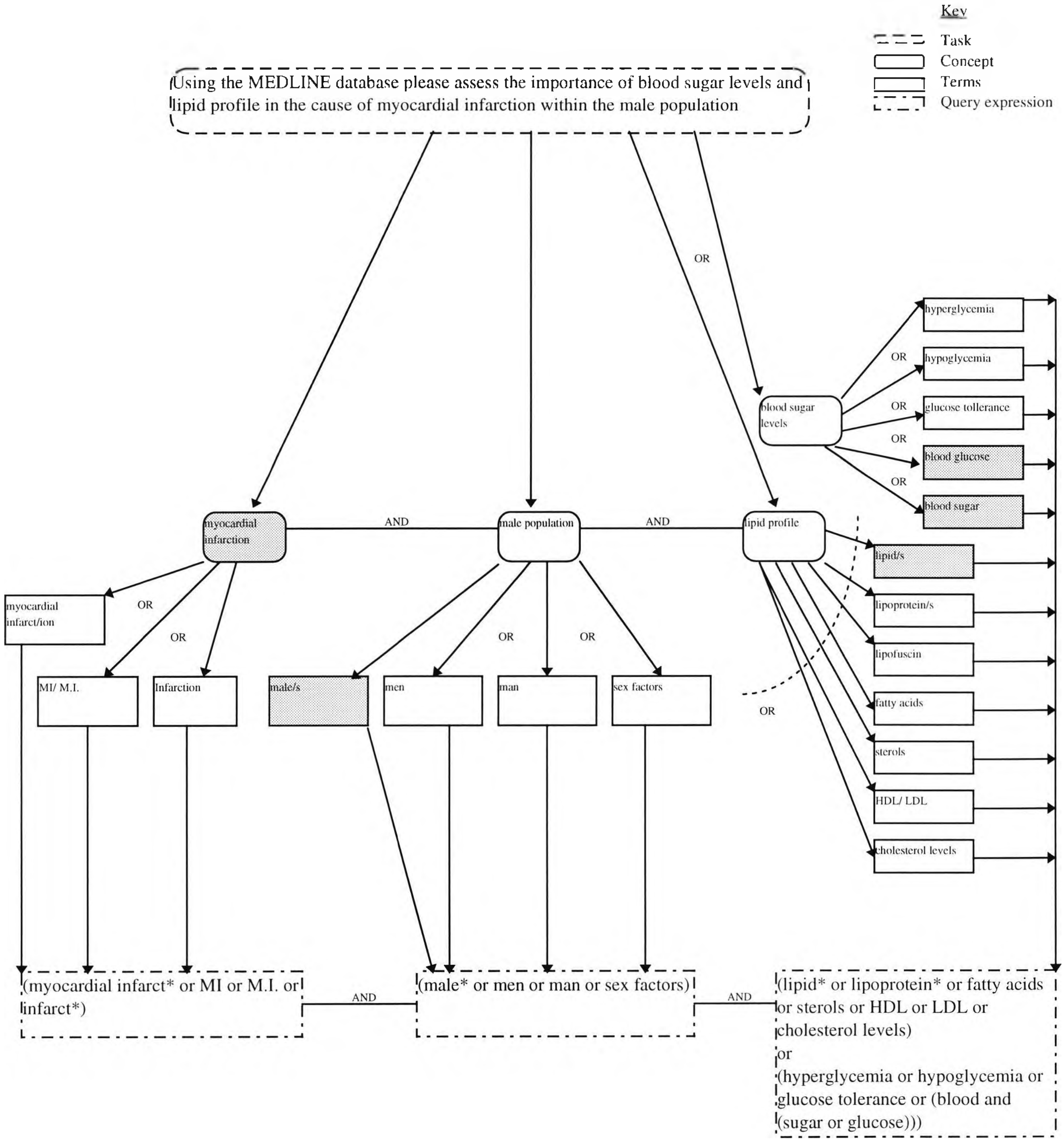

Figure 3.7: Choice of search terms for task PH1. The shaded boxes represent the 5 most popular choice of search terms.

3.4.4.4 Term analysis of task PH2

The more popular terms for task PH2 were "NHS, audit and efficiency" with variations in audit also being common; however, performance was poor for most subjects (see table 3.12). One of the two top performers used a two term query "audit and efficiency" and the other used five variations on audit with efficiency. The two key differences were in the terms used to articulate clinical audit and efficiency. Expert searchers considered substantially more terms in relation to these concepts than novices. Novice searchers did not use any alternative terms to express efficiency (except L4). L4 did not include any reference to efficiency in her query.

The ideal query in this case, illustrated in figure 3.8, possessed three concepts two of which required synonym elaboration. It is interesting that the query structure was possibly simpler than for the other tasks, yet performance for PH2 was poor. The figure breaks the gold-standard into its four component layers.

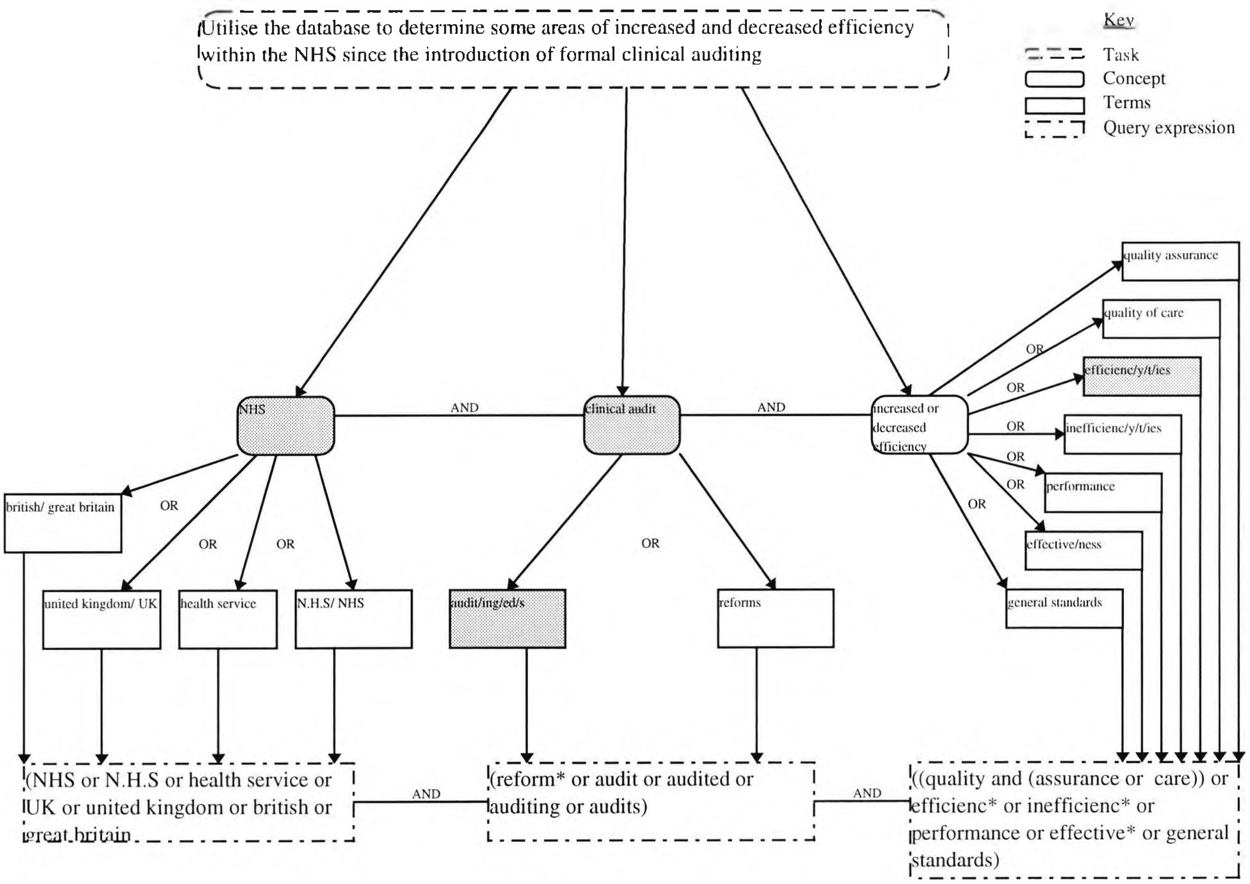

Figure 3.8: Choice of search terms for task PH2. The shaded boxes represent the 5 most popular choice of search terms.

Terms	H3	H2	H9	H1	H4	H5	H6	H8	H10	H7		L3	L2	L7	L5	L6	L1	L4
British																		
great Britain																		
united kingdom																		
UK																		
health service							✓											
N.H.S									✓	✓								
NHS		✓	✓	✓	✓	✓		✓		✓			✓	✓		✓		✓
clinical audit	✓					✓		✓	✓	✓			✓			✓	✓	✓
audit	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓		✓	✓		✓	
auditing	✓			✓	✓		✓	✓	✓	✓			✓					✓
audited	✓			✓	✓		✓	✓	✓	✓								
audits	✓			✓	✓		✓	✓	✓	✓								
reform																		
reforms																		
increased efficiency		✓		✓		✓				✓								
decreased efficiency				✓	✓					✓								
quality assurance																		
care quality																		
efficient	✓			✓	✓													
efficiency		✓	✓	✓			✓	✓		✓		✓	✓	✓	✓	✓	✓	
efficiencies				✓						✓								
inefficiency					✓					✓								
inefficient																		
inefficiencies																		
performance																		
effective																		
effectiveness																		
general standards																		
Recall %	12	9	9	5	5	5	5	5	5	2		12	9	8	5	5	2	2

Table 3.10: Subjects term usage on task PH2. Subjects are ranked by recall within the expert/ novice groups. The divisions between the different concepts are indicated by the double line.

3.4.3.3 Systems facility usage

Table 3.11 shows the systems facilities used during retrieval sessions. Term suggestion facilities provide alternative terms which could be incorporated into the search (see figure 3.9), while the thesaurus provided a controlled vocabulary of synonyms, related terms, preferred terms and definitions (see figure 3.10). The hierarchical nature of the thesaurus aided the selection of more general (broader) and more specific (narrower) terms. Each query is assigned a unique identifier by MEDLINE so previously submitted queries may be re-used. Term exploration facilities enable the extension of queries by term truncation and wild cards so partial matches can be found.

	Task support facilities																			
	Thesaurus				Term Suggestions				Query Re-use				Term exploration using operators *?				Order of execution			
H1				✪				✪	✪	✪	✪	✪	✪	✪	✪	✪				
H2	✪								✪	✪	✪	✪	✪				✪			
H3		✪	✪	✪		✪	✪	✪	✪	✪	✪	✪				✪	✪	✪	✪	✪
H4		✪	✪						✪	✪	✪	✪				✪				
H5									✪	✪	✪	✪		✪	✪					
H6														✪	✪	✪				
H7					✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪
H8	✪	✪	✪	✪					✪	✪	✪	✪		✪	✪	✪				
H9		✪				✪			✪	✪	✪	✪	✪	✪	✪		✪	✪	✪	✪
H10			✪				✪		✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	✪	
Total	13				10				36				26				16			

	Task support facility																			
	Thesaurus				Term Suggestions				Query Re-use				Term exploration using operators *?				Order of execution nesting operators ()			
L1																				
L2										✪	✪	✪								
L3										✪										
L4									✪	✪	✪	✪		✪						
L5		✪	✪																	
L6									✪	✪	✪	✪								
L7									✪	✪	✪	✪								
Weighted Total	2.9				0				22.9				1.42				0			

Table 3.11: Analysis of systems facilities used by searchers. The spatial distribution reflects the task order OG1, left hand side, OG2, PH1 middle locations, PH2, right hand side. The totals for novices are weighted to account for differences in the number of subjects in the novice group

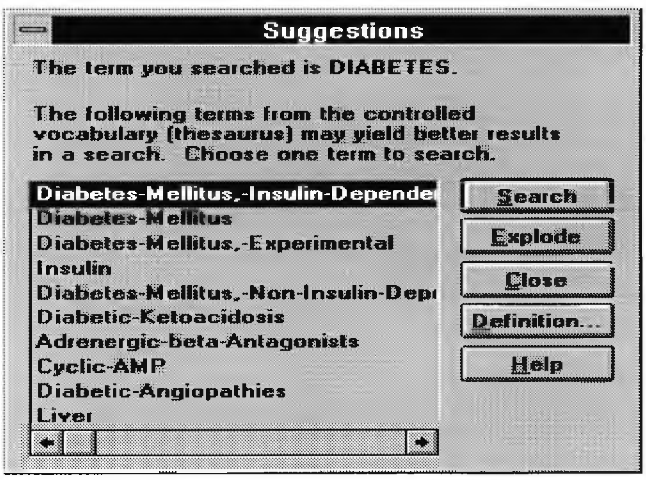

Figure 3.9: Term suggestion facilities

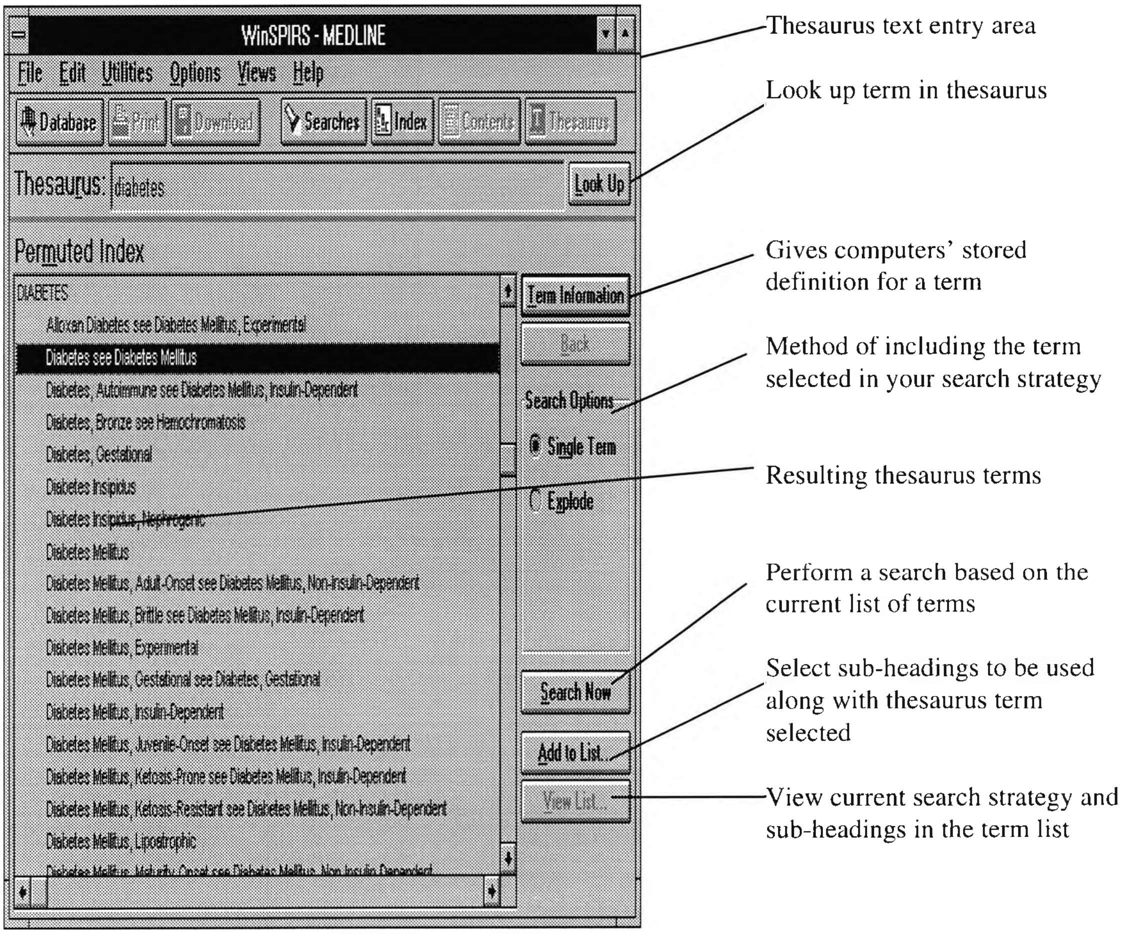

Figure 3.10: Thesaurus facilities in MEDLINE

Experts used the term exploration facilities sparsely, whereas only one novice used the facilities. Eighty percent of experts used the thesaurus or term suggestion facilities to explore concepts at some point during the experiment, but use of these facilities was inconsistent. None of the subjects used the index facilities and only one subject consulted the help system. Both groups of subjects (14 out of the 17 subjects) re-used queries at some point in the experimental session (using system attributed query identifiers as short cuts); however, expert searchers were more consistent in this (9 out of 10 subjects on all tasks). Experts used term extensions to increase the scope of queries, but novices did not. The expert searchers who used term extension facilities did so consistently irrespective of the task.

Experts used Boolean and query structuring operators to sub-divide complex queries into functional components, whereas the novices constructed only simple queries; and this pattern was consistent across tasks. Experts used ordering mechanisms (use of () operators to group query components) consistently across tasks (see table 3.11) and generally adopted more complex strategies than novices. Those strategies should have resulted in better performance, but the performance data demonstrated otherwise. Generally, usage of system facilities was idiosyncratic, possibly reflecting hit and miss discovery of system functions and inadequate knowledge of MEDLINE.

At the end of the experiment searchers were encouraged to give their views of the interface to MEDLINE and its facilities. The comments given by the users were categorised as illustrated in table 3.12.

Comment	**Number of subjects**
Frustration due to poor feedback	1
Poor thesaurus (hindrance not help)	3
Articulation difficulties (terms and relationships)	7
Systems representation of search histories poor	4
Aesthetic configuration changes wanted	2
Desire for automatic associated keyword generation	6
Desire for spell checker	1

Table 3.12: Usability problems highlighted by subjects

The searchers' comments seem to back up the problems and errors found in their searching activity. This level of awareness of the faults of the system gives reason for concern. It indicates that the user population are having problems with current information retrieval systems and aren't satisfied with the support facilities.

3.6 Summary of key reasons for search success or failure

The different reasons for success and failure in search behaviour are summarised in table 3.13. The possible characteristics of a successful search are: the use of a large number of terms to articulate the information need (1); the elaboration of a high number of the problem concepts in queries (2); longer retrieval sessions (3); a high number of query iterations (4); the use of complex queries (5); the use of systems facilities to improve database coverage (6); a high number of transitions between querying and evaluation (7); longer bouts of article evaluation (8) and cycles of narrowing or broadening(9). The characteristics indicating poor retrieval practise are: the use of a small number of terms to articulate the information need (10); the failure to elaborate the problems concepts in queries (11); short retrieval sessions (12); a low number of query iterations (13); the use of simple queries (14); failure to use of systems facilities to improve database coverage (15) .

Subject	OG1		OG2		PH1		PH2		Total		
	Factors	Recall	Factors	Recall	Factors	Recall	Factors	Recall	Good	Bad	Average recall
H1	**1, 2, 3, 4, 5, 7, 9**	27	**1, 2, 3, 4, 5, 7, 9**	7	**1, 2, 5, 8, 9**	13	**1, 2, 5, 9**	5	23	0	13.00
H2	**1, 2, 3, 4, 7, 9**	9	**7, 9**, 11, 13	24	**9**, 11, 12, 13	5	**9**, 10, 13	9	10	7	11.75
H3	**3, 8, 9**, 10, 13	45	**3, 7, 8, 9**, 14	17	**7, 8, 9**	5	**1, 3, 7, 8**, 13	12	14	4	19.75
H4	**9**, 10, 13	9	**9**, 10, 13, 14	21	**9**, 11, 13, 14	1	**1, 2, 8**, 13	5	6	9	9.0
H5	**2, 8**, 12, 13	27	**7, 8**, 14	14	**1, 9**, 12	6	**8**, 10, 11, 13	5	7	7	13.00
H6	**8, 9**, 10, 12, 13	36	**7, 8, 9**	21	**8, 9**, 13	4	**1, 9**, 12, 13	5	9	6	16.50
H7	**1, 2, 3, 4, 7, 9**	45	**1, 3, 4, 7, 9**	21	**1, 2, 7, 9**	9	**1, 2, 3, 4, 5, 7, 9**	2	22	0	19.25
H8	**2, 9**	27	**7, 9**	3	**1, 2, 3, 4, 7, 9**	11	**1, 3, 7, 8, 9**	5	16	0	11.50
H9	**2, 9**, 12, 13	18	**1, 2, 7, 8, 9**	14	**1, 2, 3, 4, 8, 9**	16	**4, 8, 9**, 10	9	16	3	14.25
H10	**7**, 10	18	**1, 7, 8**, 14	31	**1, 2**	1	**1, 5, 8**, 12	5	9	3	13.75
L1	**8**, 10, 12, 13	18	11, 14	14	**8**, 10, 11, 13, 14	8	**3, 8, 9**, 10, 13, 14	2	5	12	10.50
L2	**8**, 10, 11, 13	9	**8**, 11	17	**3, 8**, 14	22	**3, 8**, 10, 13, 14	9	6	8	14.25
L3	**2**	9	**7, 8**, 14	7	**3, 8**, 10, 13, 14	28	**9**, 10, 11, 12, 13, 14	12	6	9	14.00
L4	10, 13	9	10, 12, 14	10	10, 11, 12, 13, 14	6	**8**, 10, 12, 13, 14	2	1	14	6.75
L5	**3, 8**, 10, 13	9	**3, 8**, 10, 13, 14	24	**3, 8, 9**, 13, 14	9	**8**, 10, 11, 13, 14	5	8	11	11.75
L6	**3, 8, 9**, 10	64	**7, 9**, 10, 11, 14	7	**3, 7, 8, 9**, 14	12	**8**, 10, 11, 13, 14	5	10	9	22.00
L7	**8**, 10	27	**8**, 10, 11, 14	14	**3, 8, 9**, 10, 11, 14	18	**8, 9**, 10, 11, 13, 14	8	7	11	16.75

Table 3.13: Summary of the characteristics of good and bad strategies present in the behaviour of searchers. The recall of the subjects on the four tasks are also given in the summary table. The numbers in the table refer to the characteristics or factors described in the previous paragraph. The success characteristic identifiers are in bold and those in a standard font are characteristics of poor retrieval.

The table highlights that novices exhibit few characteristics of retrieval behaviour which should lead to success. The commonest characteristics of success produced by novices are longer bouts of evaluation and longer evaluation sessions and this did sometimes lead to higher recall. Experts exhibit more characteristics of successful retrieval and fewer characteristics of poor retrieval behaviour than novices.

3.7 Implications for user modelling of information retrieval

- Overall, searcher performance was poor. This may be attributed to the system's user interface and the task support facilities provided rather than the underlying retrieval mechanism given the performance which could be achieved by the gold standard.
- Experts exhibited significantly similar ranking of recall on the four tasks whereas novices did not, so it appears that novice searcher performance was more due to chance. This and the idiosyncratic use of system facilities indicates that IR systems should provide specific assistance, or targeted task support, for users.
- Multiple strategies exist in any context and many of these, or a combination of these, can be successful. Searchers are not consistent in their strategy across task. The strategies used by the searcher may change during a retrieval session.
- As the proportion of the retrieval session spent querying increases the complexity of the queries used increases. Novices favour evaluation (on average spending 74.5% of the retrieval session evaluating articles) and experts favour spending a greater proportion of their retrieval session querying (query to evaluation 41.5:58.5% average). These expert-novice differences have implications for modelling retrieval behaviour and the tools required to assist these different activities.
- Novices use simple queries, and spend more time evaluating the results, whereas experts submit more complex queries, use more search iterations and strategies such as cycles of narrowing and broadening. Significant differences occur in query complexity between experts and novices.
- Experts generally spend longer completing a retrieval session and consider more alternative term expressions than novices but this is still only a small proportion of the possible search concepts (expert average 22.75%; novices 15.97% of the gold standard queries). The majority of search terms used to articulate the need were found in the gold-standard query, but coverage of the gold-standard query was low.
- Experts expanded more components in the original need statement than novices by adding terms from the system thesaurus or from their knowledge of the domain (experts 60% explored, novices 37% terms expanded with synonyms or refined

with sub-terms). Experts considered more term alternatives than novices but this was only significant on the more complex tasks. However, individuals did show a common ranking in the number of terms used per task. The number of terms used is related to the characteristics of the task for both groups, so task characteristics may be useful in predicting searchers' strategies and behaviour. An important issue for modelling and design is that the different categories of user require different advice and synonym generation assistance. This combined with the performance results of subjects indicate that both groups of searchers require some task support to assist in need articulation.

- Experts re-used queries more, and used substantially more queries to articulate their need to the system than novices (experts average 12.73 queries; novices average 8.93 queries). This difference is attributed to the alternative strategies employed. Experts concentrate on cycles of narrowing and broadening whereas novices favour trial and error by substitution. Novices favour evaluation while experts use systems facilities to explore alternative terms more than novices. Novices only use Boolean 'ANDs' to express relationships between keywords whereas the Boolean relationships used by experts are more diverse. The number of query iterations performed were positively correlated to the coverage of the gold standard.
- If the searchers' retrieval included characteristics of good query construction and need articulation coupled with careful evaluation then in general recall was improved but the choice of specific terms can have dramatic effects.

In conclusion, the overall pattern is inconsistent. In the tasks with better overall performance many query iterations, longer evaluation times and sound query construction are reasonable indicators of success. Iterations and long evaluation times were also reasonably reliable indicators for novice subjects. However, these predictions broke down for the poor performance tasks (PH1 and PH2); furthermore, any one indicator was unreliable for predicting recall performance. It appears that there are several ways of achieving good performance, and that individuals can be successful by following a single good strategy even though their approach may be flawed in some respects (e.g. poor queries but many iterations). The converse seems to hold for poor performance. In some cases when individuals were following apparently successful strategies, they came unstuck with a single mistake, such as choosing inappropriate keywords; although many poor performers just submitted a few poor queries. Overall, experts are more likely to construct sound queries and follow more complex search strategies, whereas novices are more likely to spend more time evaluating results to achieve better performance.

3.7 Conclusions

Although evidence for behavioural differences were found between novice and expert searchers, no simple correlations between behaviour and performance were immediately apparent. Instead, our analysis revealed a complex picture. There are many contributing factors to effective performance, such as use of complex queries, many search cycles, narrowing strategies and careful evaluation of retrieved results. However, these factors were not evenly distributed across our subjects. Experts relied more on complex query formulation and iterative cycles of searching whereas novices relied on careful evaluation. However, success was neither guaranteed by any one strategy nor by a combination of all of them. Some of our subjects exhibited expert behaviour with poor results, and conversely some followed poor strategies with good results. Although these exceptions were a minority, they can only be explained in terms of choice of appropriate search terms. Hence content of queries appears to be at least as important as the search strategies employed.

In many tasks poor evaluation of search results may have caused poor performance among our experts, whereas novices who evaluated results more carefully performed quite well. It was noticeable that task PH2 seemed to cause a motivational problem for our subjects who exhibited short, inadequate searches. This raises important questions for the design of IR systems about how to motivate users with early rewards of search success. One possibility is to investigate the concept of information scent (Furnas 1997) whereby hints of search paths towards desired targets are provided; however, Furnas was considering a hypertext structure rather than free text retrieval. The lack of correlation between effective strategies and search success has been noted in ecological studies of on-line searches with intermediaries (Smithson 1994), who also found that poor evaluation of results early in the search cycle caused poor performance. The inter-task performance differences encountered demonstrate how task characteristics may determine search strategies.

The subjects were not only inconsistent in their search strategies but also in their choice of search terms. There was little overlap with the gold-standard solutions and the inter-subject consistency was also low. This agrees with Iivoenen's (1995) findings of individual differences and lack of searcher consistency, although consistency in detailed terms of lexical agreement has not been analysed.

Semantic distance between a searcher's articulation of information needs and the search system index or document description, is one of the perennial problems of information retrieval (Brook 1995, Ingwersen 1996). These results, on a pessimistic interpretation indicate that search strategies can not overcome the more fundamental problem of semantic distance; although more optimistically, advising subjects on strategy and particularly query development activities may lead to improved performance. Many of the subjects exhibited the expert strategies noted in the literature, e.g. query narrowing, use of alternative terms (Marchionini 1995); so can their lack of success be attributed to poor system design? This may be so, especially where system support bears upon choice of search terms. Term expansion facilities with relevance feedback in experimental IR systems have demonstrated good results (Hancock-Bealieu et al 1995). As our subjects experienced difficulty with choice of terms, improving thesaurus facilities may provide the answer but MEDLINE did have thesaurus facilities, so maybe active guidance is required or query dialogues which cue term expansion. Several explanations for the searchers' difficulties are possible. First, choice of appropriate search terms that matched the document contents in free text search. It is possible that searchers were unaware of alternative terms held in the system thesaurus, considering that only a minority of our subjects consulted it. Links to other concepts could have been found but few subjects consulted the suggestion and thesaurus facilities. It seems that educating users about these facilities only goes part of the way to solving the problem as shown by the differences in facility use by the two subject groups. Secondly, they may have had difficulty using these facilities as demonstrated by three of the searchers who explicitly complained about the usability of the thesaurus (out of the eight who used the thesaurus). The term suggestion and thesaurus facilities required several interaction steps so, for most searchers, the effort in using these facilities may not have been justified unless the search failed. Finally, the system thesaurus did not always contain appropriate keywords and spellings, hence leading to failed searches. MEDLINE's thesaurus has been set up with US spellings, hence articles with British spellings and keywords are not found by free text retrieval; also the subjects had some difficulty finding thesaurus entries.

Tasks OG1 and PH1 were both cause and effect questions of similar difficulty and yet expert recall on task OG1 (26.1%) was much better than on PH1 (7.1%). The proportion of time spent querying to evaluating results and the total search time were approximately the same for these two tasks. The only task differences seem to be that two document sets are required for PH1, i.e. lipid profile and blood sugar level, whereas only one document set, socio-economic reasons, was required for OG1. The variance in performance of novice searchers on these tasks were not as larger as experts (novice recall: OG1 20.71%, PH1 14.71 %). In general experts performed much better on the Obstetrics and Gynaecology tasks than the Public Health tasks which may indicate some domain familiarity effects. All subjects performed poorly on task PH2 and spent the least amount of time on this search task. The task may have been too difficult because the need was expressed ambiguously.

The feedback offered by MEDLINE was found to be ineffective by the searchers; for instance, the advice given on retrieval of a null set is basic. The system could indicate methods of diagnosing search failures and provide hints on more successful strategies, e.g. replacing terms, reducing query constraints, etc. Searchers require search histories which encourage and assist query re-use and diagnosis of inappropriate searches. Most of the information required to assist the searcher is already maintained by the system and this should be utilised more effectively to target support at users' specific problems.

It is clear from searchers' views of the system that they want assistance in query generation and need articulation (see table 3.12). If computerised retrieval systems aim to take over the role of the librarian as the negotiating mechanism (Ingwersen 1993) then their designers will have to pay more attention to the users' task needs when designing user-system dialogue and search support facilities. The study has shown that a searcher's device knowledge effects the strategies used in need articulation. The systems interface, task support facilities and user guidance are failing to assist the searcher in need articulation and their retrieval activity, as even domain experts are unable to retrieve a high proportion of the relevant information in the database. The performance figures, and especially the results achieved by the gold standard query, highlight the requirement for improved IR interfaces and intelligent system guidance to assist searchers. The differences between the query articulation activity and search strategies of the subject groups shows a need exists for targeted assistance based on the device knowledge held and the pattern of a user's query history. Differences in the query patterns and terms used further support this requirement for user-specific task support based on their knowledge. Experts generally considered more term alternatives than novice searchers even though they have equivalent domain knowledge. This may be due to differences in their mental models of system functionality, index structures and document representations. A searcher's device knowledge effects the perception of the degree to which term exploration should performed for need concepts. This has design implications for the articulation assistance provided by the system and the advice offered to searchers. If searchers are encouraged to develop more specific representations of the information required the gap between their current and the optimal performance may be reduced. The study has highlighted the requirement for an alternative approach to IR interface design and indicates some of the factors and strategies which must be included in a model of user behaviour in its description of retrieval sessions. These aspects of users retrieval are included in the cognitive model described in chapter 4.

Chapter 4

Explanations of information seeking through cognitive task models

This chapter describes a cognitive model of the information seeking task and discusses how this is tailored to specific search situations.

Chapter four: Explanations of information seeking through cognitive task models

4.1 Introduction

The need for improved interface designs and targeted task support was established in chapter three, based on performance data gathered in an empirical evaluation of a typical commercial database: MEDLINE[1]. Chapter 2 discussed the different models of IR behaviour and how these need to be adapted and augmented to offer a more complete view of user behaviour. This chapter describes theoretical work addressing the cognitive activity underlying search activity and the synthesis of this into a cognitive task model of information seeking behaviour. The motivation for this is to understand, explain and minimise the mismatch between information retrieval interfaces, their associated functionality and users' mental models of the search process. It is not intended to alter the underlying search algorithms, Robertson et al (1994), but the aim is to try to improve the matching between the searcher's need and the information in the database through interface design, dialogue design and by providing appropriate IR services for individual user needs. The chapter is organised as follows: section two introduces the architecture and nature of the mechanisms used by the cognitive model to simulate the varied behaviour, strategies and tactics occurring in information retrieval. Section three describes the cognitive activities and correspondence rules which give the model its adaptivity, before discussing aspects of the model which relate to the influence that information retrieval systems have on behaviour. This section discusses the narrowing of the taxonomies of all the possible systems facilities as the model is adapted to simulate user behaviour with a specific retrieval systems design. Section four discusses the representations used for search results within the model. In conclusion the implications of our modelling activities on IR systems design are discussed.

4.2 Architecture

The architecture posits many components (see figure 4.1) which allow a general model of the IR process to be 'context adjusted' to the situations faced during a

[1] MEDLINE is the copyright of the U.S. National Library of Medicine and SilverPlatter

retrieval session. The process model can be configured to describe the many possible routes and outcomes for retrieval behaviour using correspondence rules. Correspondence rules dictate the search activities (cognitive and physical) performed and selection decisions made given an information need type (general-specific, known-unknown, simple-complex etc.) and users context (knowledge, motivation, time etc.). Information retrieval is a process of dialogue between the IR system and the user (Raskutti et al 1997). For this reason the architecture contains a representation of the user's plan of action and a representation of the system's responses, as the later of these effects how plans of action may be modified based upon the results produced. To facilitate this dialogue a user must translate the plan of action into a device specific execution plan (interaction plan). To model this process the architecture contains a representation of the system facilities present, and known about, and the different situations in which these facilities may be required.

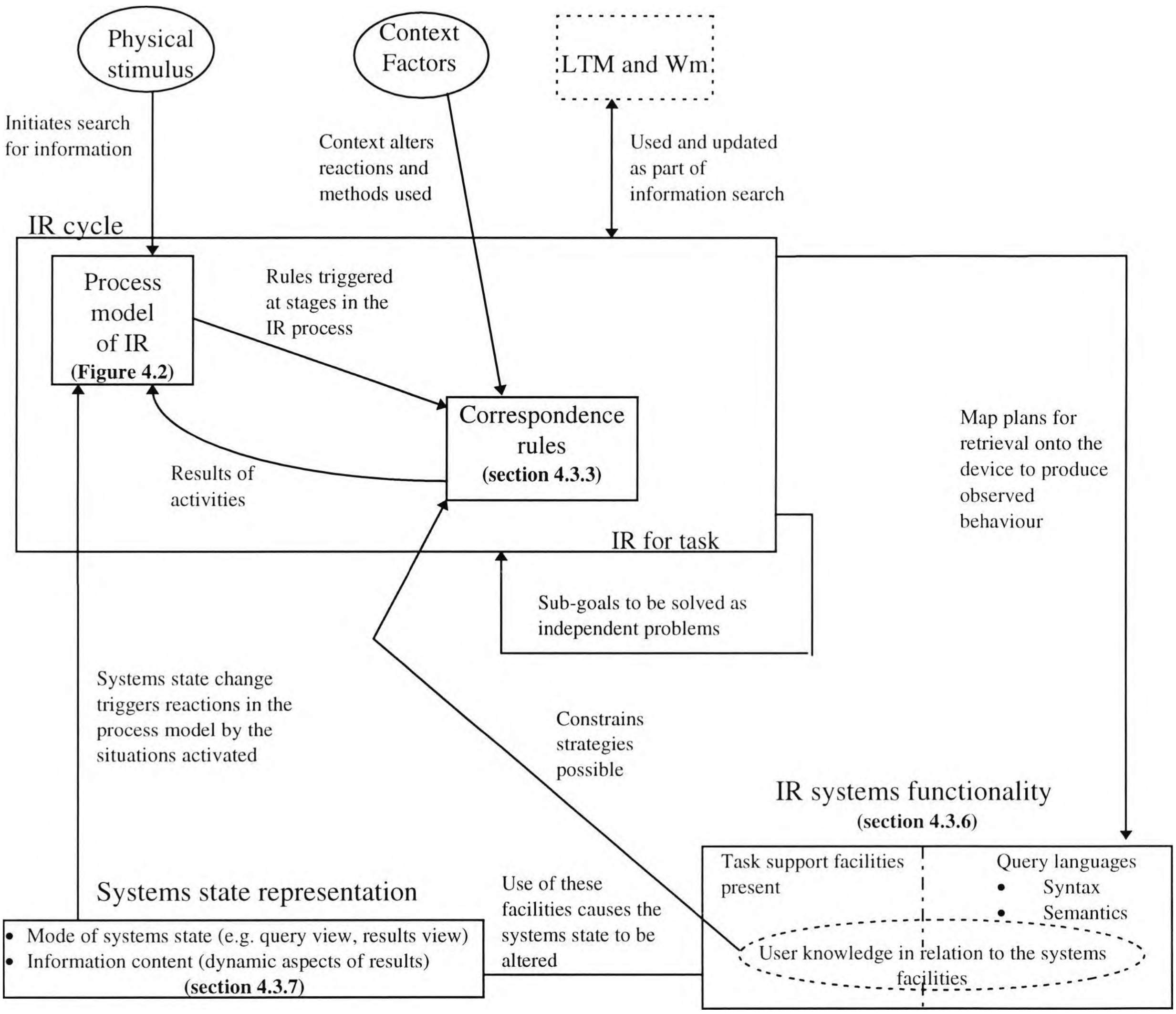

Figure 4.1 : Architectural view of the IR problem and the situation inputs to the behaviour produced.

The major components of the architecture are:

- IR cycle simulating the activity underlying search behaviour (process model and correspondence rules),
- the perceptual mechanisms, action operators and physical activity components (physical stimulus, generic problem development operators and context factors) ,
- memory resources (Long Term Memory and Working Memory),
- systems specific components (systems state and IR systems functionality representations).

IR cycle:
Information retrieval has a flexible and non-deterministic nature so a general model is posited through which there are many pathways. The pathway taken is dependent on the correspondence rules, causing the generic process model to be context specialised into an instance specific model. The process model is tailored, or configured, to the task, structure and level of the users knowledge and the context factors present; see figure 4.1. This flexibility and configurability permit simulation of evolving search behaviour given the different situations encountered in a retrieval session.

The need to build adaptivity and configurability into the model stems from the acknowledgement that user goal states, and the methods used to locate and extract the information required, may not be determined at the search outset (Kwasnik 1992). Any ambiguity in the external task must be resolved during the search process as the information need is structured and search directions are identified (Taylor 1968) or queries may under specify the need leading to a greater number of articles to be evaluated. For example if a concept (A) can have different semantics in two different domains then unless other concepts are included in the query to distinguish between the different contexts (for A) the system will match both groups of articles. It is evident from the dynamic nature of users information requirements the search process is not rigid, but is adaptive in the light of changes occurring in a searchers' environment (Ingwersen 1996).

Perceptual mechanisms, action operators and physical activity related components:
Perceptual mechanisms relate the cognitive activity performed to the external world. The mechanisms are required owing to the influence of the external task, system and results on the retrieval process. The model accounts for the effects of the external world on retrieval behaviour at a description level and so the effects of users having different perceptions of the same information are eliminated. The model requires the user beliefs about the task and results produced to be described and does not attempt to model user perception. The perceptual component is included in the architecture for completeness as the results of the mechanism effect the structure of the model and the strategies applicable.

Memory resources:
The structuring of problem information and the identification of search options are achieved by referencing long term memory and the systems information space (Michel 1994; Allen 1991, 1994). The effects of knowledge resources are included in the model through generic types of Long Term Memory (LTM), i.e. domain and device (expert and novice). For this reason the model can be applied in different search domains as it does not require a complex instance specific domain representation. Knowledge resources are segregated from other context factors in the architecture as knowledge plays a dominant role in directing and determining the cognitive activities performed. The cognitive activities performed in turn determine the strategies and tactics used, and thus behaviour observed. The other aspect of memory resources are the different working memory (Wm) representations of the information need as the search progresses.

Systems specific components:
The system, its facilities and the representation of the information retrieved have an important effect on the user strategies, tactics and activities assisted and supported. The users ability to diagnose the success of the search depends on their perception of the state of the system at any point in time, i.e. active facilities and retrieved information. The outcome of this diagnosis affects the subsequent user actions on the interface. The search results are explicitly represented within the architecture along with representations of the system facilities. The facilities and interface influence the retrieval activity, the users reaction to search results and directions followed in the search.

In conclusion the architecture posits that the search process is a multi-layered, complex and highly inter-related structure of IR cycles. Each IR cycles in the search process may involve decisions, strategies, tactics, cognitive and physical activity. The outcome of these activities and their sequence is dependent on the searcher's knowledge profile i.e. the domain and device knowledge held, and the characteristics of the information need. For example searchers with low device knowledge may evaluate larger results sets earlier in a search due to difficulties operating the device; whereas experts may try and specify the information need as accurately as possible before evaluating articles in an attempt to minimise the number of irrelevant articles evaluated. An optimal search involves discovering effective queries to locate the required information, reasoning and the selection of appropriate strategies, tactics and terms. If the search is not successful due to constraints (time, motivation etc.); poor knowledge of the domain or device then users may fall back on compensatory strategies such as exploration in device facilities, amendment of search goals, deletion of need components or the evaluation of large results sets. The course of action taken depends on the perceived importance of the obstacle to search success and the expected difficulty associated with finding its solution.

4.3 Process model of information retrieval

The generic process model (see figure 4.2) synthesises previous research (see chapter two) with the findings from empirical studies on factors effecting searching (see chapter three) into a unified structure. The process model can produce many different behaviour patterns by incorporating dynamic features in the search process. The four processes shown in figure 4.2 are driven by the situation, task context and searcher abilities to produce the output of decisions and activities.

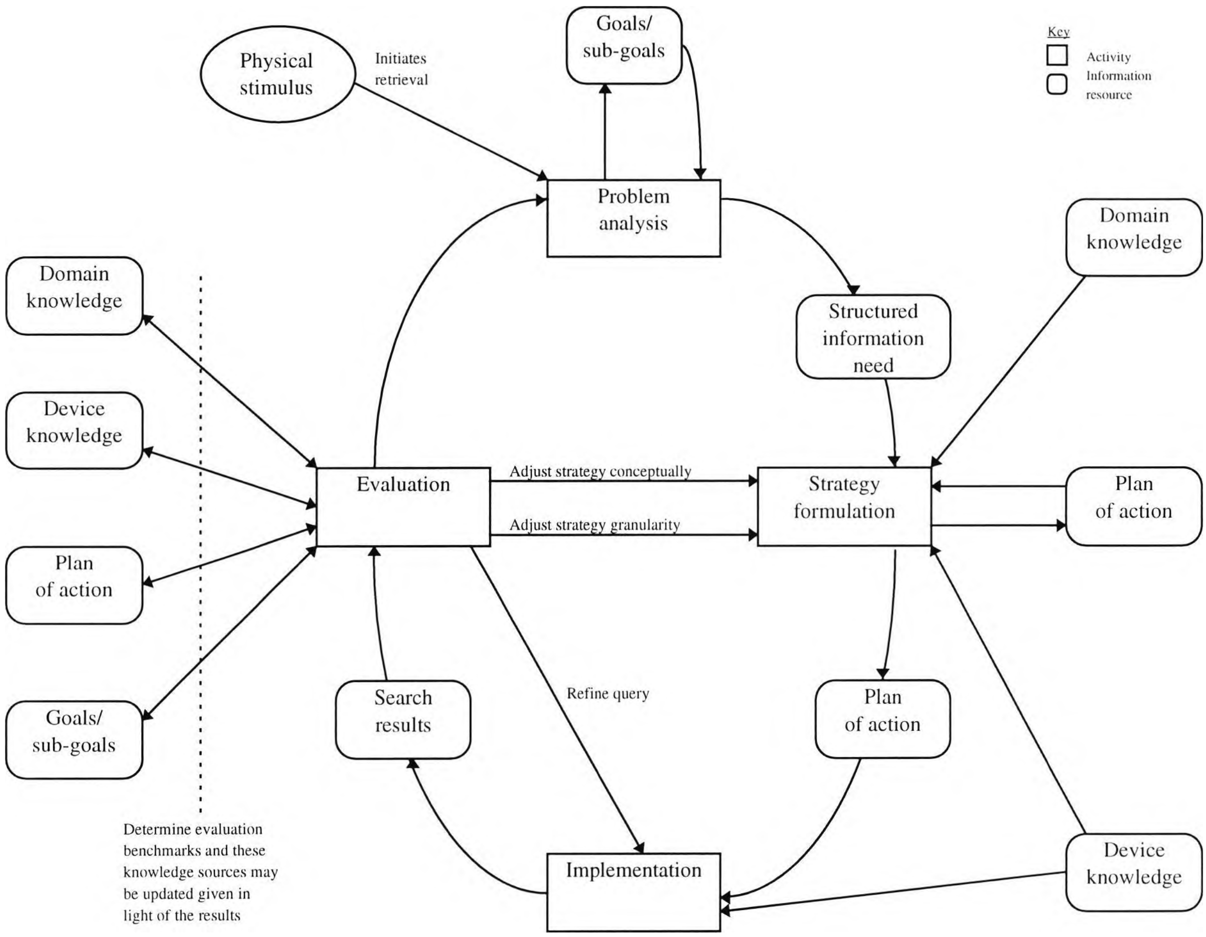

Figure 4.2: General process model of IR and its information flow. The route taken through the general model is determined by the users knowledge and context.

In the process sequence a information need is presented; a plan for search solution is formulated based on the applicable strategies (strategy formulation). Strategies are refined into tactics that lead to search implementation, predicting how the query should be executed in a retrieval system according to the available task support facilities (TSFs). Once search results are returned the evaluation activity advises on the query reformulation strategies to follow; given a description of the volume, relevance and precision of the results set. Further plans of action in terms of strategies and tactics are developed based on the outcome of the retrieval activity (strategy formulation).

The types of knowledge effecting the nature of the search process and its outcome have been reported by Michel (1994), Allen (1991, 1994) and Ingwersen (1982) and are incorporated into the framework of the model as categories. The decision to condense the knowledge categories known to affect retrieval behaviour is justified on the following grounds:

- problems occur in eliciting the level of a users knowledge for the 11 categories used by Michel (1994),
- the added utility of informing system design using detailed knowledge representations is minimal,
- the inherent difficulties which occur in the production of complete representations of knowledge at a low level of detail and the difficulties in updating such knowledge as a search progresses.

On the other hand, a higher level of abstraction of searchers' knowledge with broader categories (domain, device facilities and strategic IR knowledge) helps to predict the level and type of activities the user can perform when interacting with information retrieval system functions. The effects of differing levels of knowledge on activity serve to focus searching behaviour and retrieval activity in core areas of the general model of IR eg. strategy formulation or results evaluation.

4.3.1 Problem analysis

The problem analysis process is dealt with for completeness but it is not a functional part of the model. The complex and well researched areas of reasoning and problem solving are not described because of the complexity of acquiring and representing domain knowledge which also would limit the generality of the process model (Eysenck et al 1990, Newell and Simon 1972), but the activities which should be performed are acknowledged. Instead information needs are categorised into types, an approach used by Ingwersen (1996), and these types are used to direct subsequent user behaviour and strategy selection (see section 4.4.1). The use of categories of information need enable the complex issues related to variations in users interpretation of the information sought to be disregarded.

4.3.1.1 Activities performed

The recognition of the existence of an information need and its initial structuring into a series of goals is the fundamental function of the analysis phase. The information need representation directs future cognitive activity, strategy selection and the related search behaviour.

The possible processes that can be performed are:

1. Recognition of the existence of an information need and the production of beliefs relating to its solution.
2. Identification of concepts and their inter-relationships within the problem. Determine the information attributes which indicate search topic directions, search terms and sub-problems within the task.
3. Structure and prioritise goals and concepts in the information need. For example lipid profile is a member of the lipid class of material and thus is an acceptable value as a narrower expression for the lipid class but within the same goal.
4. Filter out irrelevant concepts as the view of the information required becomes focused. For example when dealing with multi-faceted problems the whole 'picture' of the information required only becomes apparent as context information is gathered. This has the effect of causing a shift in the focus of attention within the search and renders some concepts as no longer relevant.
5. Trigger the follow on process when the problem has been analysed. This can have two possible outcomes. The one chosen will be dependent upon the complexity of the task, cognitive load constraints and the granularity of the searchers knowledge:
    a) Trigger a new IR cycle with part of the problem acting as the goal (sub-goaling);
    b) Proceed to strategy formulation using the current problem description.

4.3.2 Strategy formulation and re-formulation

Strategy formulation is a very diverse and complicated process. This is due to the iterative, dynamic and evolving nature of the information seeking process (Rassmussen 1996). Different strategies and tactics may be used to produce a plan of action when a new problem is presented or adapting a current plan of action if there is a change in the users context. The user's context dictates the outcome of the strategy formulation activities performed during the production of the plan of action as it:

- limits the strategies and tactics which are known.
- limits the strategies and tactics which are applicable given the characteristics of the information need, the user and if reformulating a strategy the results retrieved.

The strategy formulation process can have many possible outcomes depending on variations in user context. Strategy formulation depends on the complexity and characteristics of the task. The possible general outcomes in terms of the architecture and process model are:

- Trigger a cycle of the general model of IR with the sub-goal acting as the information required.
- Implement a search based on the plan of action developed.
- Terminate the retrieval as the problem goal appears insoluble.

The user's interaction with the retrieval system depends upon the strategies and tactics embedded within these outcomes and the task support facilities which are available to assist their execution.

The outcome of a reformulation decision is to target activity with the purpose of either:

- selecting an alternative method for problem solution.
- altering the granularity at which the cognitive activities operate,
- update memory concerning domain exceptions and expectations, expand the domain knowledge held and produce a reformulated strategy based on the new domain knowledge and altered expectations.
- alter the strategies and tactics used as the current action plan does not allow a true representation of the information need.
- the results produced serve to trigger memory concerning similar problems or search approaches. The user adapts the current strategy, tactics and problem components given the results retrieved owing to the change in the users context.
- apply more systematic search methods due to the failure of heuristic approaches. For example a searcher may submit a list of terms in the hope the system returns a result without any pre-search planning or synonym exploration but if this fails then they may try a more methodological approach gradually iteratively adding concepts and exploring alternative synonyms for each of the concepts included in the query.

This rest of this section is structured as follows: first the activities performed in strategy formulation and reformulation are defined, next all the possible strategies and tactics are defined. This is followed in section 4.3 2 3 by a description of the strategies which are applicable if a new strategy is produced given variations in the characteristics of the information need and user knowledge. This is followed in section 4.3.2.4 with a description of the possible strategies and tactics for reformulation given variations in search diagnosis and the users knowledge.

4.3.2.1 Activities performed

The activities performed in strategy formulation enable the production of a plan of action to satisfy a need, within the constraints of the user's cognitive resources and context. First the different strategies and tactics for locating information are explored given the characteristics of the need and user's knowledge. From the possible strategies and tactics a plan is selected based on a trade-off between the effectiveness and effort associated with each approach. In this context the effectiveness of a strategic approach refers to the expected completeness of the information solution and the interaction costs required to enable its execution. The plan selected may include multiple strategies and tactics. The strategies and tactics selected alter the facilities used and thus the interaction with the device. The level of a searcher's knowledge influences the number and quality of search plans, strategies and tactics available.

The main activities performed in strategy formulation are:

1. Assess the appropriateness of the different strategies and tactics for locating information to satisfy a need. The options available to a specific searcher are a subset of all 'possible' alternatives and are a function of the information need characteristics, searchers' knowledge, reasoning abilities, motivation, resource requirements of pursuing alternative courses of action and the 'prize' value of the solution. The 'prize' value attached to a problem solution refers to the benefits or consequences of success or failure: Elkerton et al (1984); Irwin et al (1957) and Wendt(1969).
2. A plan of action for the search is selected from those possible given the users knowledge and need context. For example given low domain and high device knowledge a searcher is likely to choose a path of early implementation and

iterative searches using rapid prototyping rather than reasoning about the need and formulating queries carefully. A searchers' behaviour is directed by their perception of their own abilities, and judgements on the most effective search methods. The strategy selection is a trade-off between the simplicity of implementation, efficiency given the task characteristics and that which will produce the most complete solution. This trade-off minimises the cognitive resources spent whilst satisfying need expectations. In this way cognitively expensive paths are avoided. The plan of action may include multiple strategies and tactics; and strategies can comprise of sub-strategies. When all the possible plans of action have been exhausted the searcher is forced to give up the search.

3. Execute strategy development activity based on the strategies and tactics included in the plan of action.
4. Trigger the follow on situation which can be:
    - Split needs into sub-goal components and plan retrieval for each separately;
    - Proceed to search implementation for the plan of action developed;
    - Give up as the search requires too much effort to continue.

4.3.2.2 Possible strategies and search tactics definitions

4.3.2.2.1 Strategy types

Different strategy types identified by IR research (Harter et al 1984, Marchionini 1995), have highlighted the diversity of approaches to information seeking. The selection of the strategies applicable in any context is determined by correspondence rules which dictate the behaviour produced given the information need type and users' knowledge.

The strategy types possible are:

- rapid prototyping (iterative refinement based trial and error),
- direct query composition (need is well formed so problem becomes an articulation issue at the implementation level),
- direct query composition and cognitive exploration (explore synonyms in LTM, reason about problem ambiguities, refine need articulation and carefully construct queries),

- script or template based retrieval (rehearsed and familiar problems; domain or device templates),
- exploration of aspects of the problem using the device (e.g. use of thesaurus and guidance facilities to solve need problems before an effective query strategy can be formulated),
- sub-goaling (sub-divide the information need to minimise the complexity of the queries executed. These sub-goals may be aggregated later on in the search),
- goal directed browsing (berry-picking type behaviour and information component selection Bates 1989),
- exploratory browsing (knowledge, experience and skills acquisition),
- ignore goal and terminate retrieval for this IR cycle.

Rapid prototyping: This is manifest as rapid refinement and alteration of the queries submitted. A searcher with non-optimal knowledge, either in respect to the domain, device or IR strategies, is able to get a 'feel' for the number of records associated with a particular concept or query. The focus of activity, and the searches generated, is on utilising the system to sample the database. This strategy is reactive to feedback in iterations of query development rather than a systematically planned approach. The emphasis of activity is on evaluation rather than pre-search planning. This 'trial and error' approach is a method of reducing cognitive effort in the hope that one will be lucky and chance upon the information required. Large quantities of relevant information may be missed using this approach. The searcher terminates the iterative query-evaluation cycle when the results set meets their minimum information requirement. The iterative cycles of query development can: add, subtract or replace terms or concepts in the queries submitted. If a concept is replace then all terms used to articulate that concept are replaced with terms used to articulate the replacement concept.

Direct query composition without synonym exploration: This strategy aims to produce a complete representation of the information need, as the need and structure of the query are known. The information need components are complete and thus do not require elaboration.

Direct query composition and cognitive exploration: The information need is incomplete and may include ambiguities. The searcher elaborates their mental model of the need and extends the scope of the queries used by exploring possible synonyms and lexical variations, and reasoning about the problem. The ability to perform this depends on:

- domain knowledge of alternative terms for a concept
- knowledge of the retrieval system approaches to query term matching against document representations
- strategic knowledge that the exploration of synonyms and the use of richer need representations will increase search efficiency

Concept exploration may produce many term alternatives which must be judged for their usefulness in the present search context. Term selection focuses attention on the terms expected to produce coherent and complete results at the appropriate level of detail.

Script or template based retrieval: The information need is matched to a template plan of action, and possibly a re-useable query, as the information problem is known and rehearsed. This provides a fast track mechanism that minimises the effort required to satisfy a goal. This approach can have pit falls as a searcher may try and fit a specific problem to a more general template causing the search to be inefficient and non-optimal. Templates are based on experience and may relate to either the domain or specific retrieval device. Domain templates are pre-formed questions to be answered if the different aspects of the information need are to be covered. Device templates are pre-formed plans of interaction and TSF selections.

Exploration of aspects of the problem using the device: Motivations for the exploration of information in the device are:

- if the user does not have sufficient knowledge of the domain to articulate a information need accurately but has sufficient knowledge of either IR strategies or the device to realise that synonym elaboration may increase query effectiveness.

- to obtain a more accurate mapping between the users knowledge of the domain and index terms,

Exploration in the device may be performed to:

- Explore the structure of the information need by exploring inter-relationships.
- Explore the system's classification structure to obtain a general overview of the domain and context to the current information need. If the system does not support free text retrieval this activity may also allow the categories used to index documents to be discovered.
- Explore any synonym terms which can be used to express a concept occurring in the information need.
- Explore systems definitions for concepts to promote a greater understanding of the information need.
- Explore or browse the document space to extend ones knowledge of the types of articles available.
- Explore the device facilities available and how these should be used based on the help system.

Exploration activity is used to 'narrow' the gap between a users mental representation of the information required and systems indexing structures, but may cause frustration if, as in the empirical study described in chapter three, searchers find systems facilities, such as thesaurus, miss-interpret their 'meaning'. Exploration of the device may be used to determine a greater understanding of the domain, information need or the system facilities.

Sub-goaling: Sub-goaling divides the information need into smaller more manageable chunks. These sub-information needs are used to trigger new IR cycles (see section 4.2). The sub-information need becomes the focus of attention until an acceptable solution is found. When all of the sub-information needs have been satisfied they are aggregated to satisfy the initial information need. For example a high level cycle of the process model might produce a plan of action to address myocardial infarction and contra-indications to streptokinase as independent sub-goals before looking at the inter-relationships of these within the male population.

Goal directed, semi-directed and exploratory browsing (Herner 1970; Apted 1971): Browsing strategies vary based upon the support provided by the device. Browsing activity can use specialised system tools, e.g. concept or information space maps, or use standard query mechanisms as browsing tools e.g. applying general searches and skimming the articles retrieved for items of interest. The different types of browsing identified in the literature require similar cognitive activity but the differences occur in the facilities used and the focus of the items determined to be of interest e.g. goal directed browsing is likely to focus on information related to the goal whereas exploratory browsing may have no focus at the outset and thus items of interest are more serendipitous. These activities focus on an implementation-evaluation cycle to locate different information elements of interest in the device or resulting document space. This browsing activity can be used to explore and satisfy goals in the task or to acquire knowledge. Browsing can utilise the other strategies and tactics discussed but a searchers reaction to the results and the evaluation judgements are not directly goal driven e.g. selection of information snipits as opposed to a need solution.

Ignore and terminate: The search relating to the sub-goal/ goal is terminated as the solution is seen to fall into one of the following categories:

- the effort associated with searching is not justified by the perceived value of the information,
- the problem is seen as insoluble,
- the sub-goal is seen as a diversion from the original information need,

The user determines if a ‘give up’ category has been reached based on:

- the value of the information sought (Elkerton et al 1984, Irwin et al 1957, Wendt 1969),
- motivation, number of queries submitted and the success of those queries,
- completeness of the solution set of articles selected as an answer to the information need
- the expected success associated with further querying based on the results produced so far and the number of tactical, strategic, term or concept options still to be explored.

4.3.2.2.2 Tactic types

Tactics are lower level actions than strategies allowing a plan of action to be formulated as a reaction to the systems results. Tactics can operate at the query level or concept level. The result of query level tactics are observed by changes in the concepts used in the query. The result of concept level tactics are observed by changes in the terms used to articulate concepts already in the query. The tactics used in the model are as follows:

- broaden query by adding a new problem concept e.g. if the concepts myocardial infarction and diabetes are current in the query and the concept male is added. Note many terms may be used to articulate each of these concepts e.g. myocardial infarction could be articulated as the Boolean expression (myocardial infarction or heart attack or heart disease) but this is still only one query concept represented by multiple query terms.
- broaden query by substituting a problem concept
- broaden query by deleting a problem concept associated to the problem with a 'AND', 'WITH' or 'NEAR' Boolean relationship
- broaden query by changing relationships between concepts if this doesn't alter the semantics of the information need e.g. change 'WITH' to 'AND'. The different manifestations of this are dependent on the relationships supported by the device.
- broaden concept articulations by adding term synonyms at the same level of granularity with 'OR' relationships
- broaden concepts by substituting the terms used with more general synonyms
- broaden concepts by relaxing the constraints used in query terms
- broaden by exploring the problem in the device
- broaden by exploring definitions and alternative articulations for concepts in the device
- narrow query by adding a new problem concept
- narrow query by substituting a problem concept
- narrow query by deleting a problem concept associated to the problem with a 'OR' Boolean relationship

- narrow query by changing relationships if this doesn't alter the semantics of the information need e.g. change 'AND' to 'WITH'. The different manifestations of this are dependent on the relationships supported by the device.
- narrow concept articulations by adding term synonyms at the same level of granularity with AND relationships
- narrow concepts by substituting the terms used with more specific synonyms
- narrow concepts by tightening the constraints used in terms
- narrow by exploring problem in the device
- narrow by exploring definitions and alternative articulations concepts in the device
- aggregation of sub-goals or previous queries

The results of device exploration are then incorporated in the query used or if no information is available in the device then an alternative tactic is used. The tactics used depend on the diagnosis judgements made in results evaluation (see section 4.3.4.2) and those which the user is able to perform given their knowledge.

4.3.2.3 Possible strategies given need type present

The strategies and tactics discussed (see section 4.2.2.2) are the list of all possible strategies and tactics used within our model of users retrieval behaviour. Knowledge of these strategies and the ability to execute them on a retrieval system is dependent on the user's knowledge of the domain and the device. The strategies and tactics applicable depend on the information need characteristics e.g. if the information need is well known then exploration based tactics and strategies are unnecessary. In the following section the effects of user knowledge and the need type on the strategies and tactics applicable are discussed.

4.3.2.3.1 Domain and device experts

The possible strategies are dependent solely on the characteristics of the information need as the users' knowledge should not restrict search options. Table 4.1 shows the possible strategies given the different categories of information need. Well known information needs concentrate user activity on query formulation. The choice over which strategy to use depends upon the fit between the need and pre-formed plans of action, and the precision of the need. Unknown needs focus activity on using the device to overcome need anomalies. The complexity of the information need effects the 'cognitive load' placed on a user by the different strategies possible at any point in time. The 'load' a user will accept as reasonable will depend on their motivation and the time available to interact with the retrieval system. The sub-division of a information need is likely to increase system interaction but should reduce the errors and effort associated with query articulation. The precision required of the results effects the level and type of synonym elaboration required.

Need category	Possible strategies	Dependencies and notes
known	1. fit to domain and device templates (pre-formed plans of action); 2. direct query composition with synonym elaboration 3. direct query composition without synonym elaboration 4 additive rapid prototyping	complex, time, motivation, general, specific, critical nature of the solution
unknown	1 explore using the device (at problem level or concept articulation level); 2. substitution based rapid prototyping (too determine the most acceptable articulation for a concept); 3. exploratory browsing (learning and update knowledge of the problem domain) 4. direct query composition with synonym elaboration and using systems functions 5. direct query composition without synonym elaboration	Favour strategy 1 complexity, time, motivation(completeness and importance of need success)
complex	1. sub-goal and form strategies for each information component: 2. direct query composition utilising complex query syntax 3. iterative additive rapid prototyping	Favour strategy 1. time; motivation, known-unknown, general-precise as these characteristics of the information need may cause changes in complexity and additional attentional requirements if the need must be specified in a greater detail
simple	1. direct query composition with synonym elaboration 2. direct query composition without synonym elaboration	motivation, critical nature of the solution and time dependent
general need but for general information and therefor the need is at the correct specificity	1. direct query composition with synonym elaboration at the same granularity as need concepts 2. direct query composition without synonym elaboration	motivation, time and complexity
general but for specific information as the need is currently under specified	1. direct query composition with synonym elaboration using more specific synonyms for need concepts 2. explore using the device (locating more specific synonyms) 3. direct query composition using the query language to extend the terms used through truncation and wildcards	
precise	1. direct query composition without synonym elaboration	complexity

Table 4.1 Initial strategy determination table for domain and device experts

4.3.2.3.2 Domain expert but device novice

The possible strategies are constrained by low device knowledge and thus the searcher uses strategies which minimise interaction with the device. The only anomaly to this is in the form of substitution based rapid prototyping in which the interaction cost is minimal as it just involves changing terms in an existing query structure. Complicated inter-relationships may exist between need concepts but users generally tend only to use 'AND' relationships between need concepts due to difficulties constructing complex query syntax because of their low device knowledge. The evaluation of results dominates user activity as it is easier to produce general queries and assess article relevance by evaluating larger results sets than producing complex and precise queries.

Need category	Possible courses of action	Dependencies and notes
known	1. Fit to domain template 2. direct query composition with synonym elaboration 3. direct query composition without synonym elaboration	motivation, time, complex and quality of activity template
unknown	1. substitution based rapid prototyping; 2. direct query composition with synonym elaboration 3. exploratory browsing in the records retrieved by queries	Favour strategy 2 over strategy 1 complex, precision
complex	1. sub-goal 2. substitution based rapid prototyping; 3. additive/ deductive iterative prototyping	Favour strategy 1 due to the complexity of the interaction involved with trial and error given a complex problem situation, motivation, time. Therefore multiple parallel searches as opposed to complex Boolean statements
simple	1. direct query composition with synonym elaboration 2. substitution based rapid prototyping	precise, general, need can be articulated effectively but this can be obscured by difficulties using the device
general but for general information	1. direct query composition with synonym elaboration at the same granularity as need concepts 2. direct query composition without synonym elaboration	
general but for specific information	1. direct query composition with synonym elaboration using more specific synonyms for need concepts 2. direct query composition without synonym elaboration	complexity
precise	1. direct query composition without synonym elaboration	complexity

Table 4.2 Initial strategy determination table for domain experts but device novices

4.3.2.3.3 Domain novice but device expert

The searcher favours interaction with the device as they do not have the domain knowledge necessary to supply query terms from memory but they are able to leverage the device to explore alternative query articulations. An example of this type of user could be a librarian, in search situations they may not fully understand the information need but they are able to use the system effectively to compensate for their lack of understanding. This reliance on device facilities and more sophisticated tactics may cause longer query times and the acceptance of all the records produced when a information need has been articulated as precisely as possible given the device operators and query syntax of the system. The searcher knows the type of search activities and problem development processes which should be performed if the need is to be satisfied, eg. term elaboration, and try to achieve this by using systems facilities to compensate for their own lack of domain knowledge.

Need category	Possible courses of action	Dependencies and notes
known	1. Fit to device template 2. explore information in the device to assist specific concept articulation (using systems functionality to extend the scope of terms in the query)	complex, time and motivation. Favour proceduralised activities based on device knowledge
unknown	1. explore in device facilities to gain context information (either problem or concept level) 2. substitution based rapid prototyping (using systems functionality to extend the scope of query terms)	favour strategy 1 over strategy 2. Depends on the interaction steps required to explore device facilities, complex, time, motivation
complex	1. direct query composition without synonym elaboration using complex device facilities to specify the need using short-cuts 2. iterative substitution rapid prototyping due to domain term unfamiliarity	includes using one complete query or multiple progressive queries
simple	1. explore in device facilities for improved term articulations; 2. direct query composition without synonym elaboration	time, motivation and critical nature of the solution
general but for general information	1. explore in device facilities for improved term articulations at same granularity: 2. direct query composition without synonym elaboration	
general but for specific information	1. explore term articulations in the device and execute (select sub-ordinate terms to articulate concepts) 2. direct query composition without synonym elaboration	complex, time, motivation and critical nature of the solution
precise	1. direct query composition without synonym elaboration 2. rapid prototyping using cycles of narrowing by addition so as not to over shoot results	complex, time, motivation and critical nature of the solution

Table 4.3 Initial possible strategy determination table domain novices but device experts

4.3.2.3.4 Domain and device novices

The searchers options are minimal so most of the queries include only the terms given in the external task and users accept all/ most of the results returned. Retrieval sessions are short and inadequate as the searcher doesn't understand the domain or how to use the device effectively.

Need category	Possible courses of action	Dependencies and notes
known	1. Fit to general problem solving template 2. Direct query composition without synonym elaboration 3. additive iterative rapid prototyping	complex, even though known a reluctance to use the device exists and they do not have the knowledge to produce synonyms
unknown	1. substitution based rapid prototyping 2. explore in the device by inspecting the results sets and matching to domain knowledge (learning)	Favour strategy 2 complex so their is a need to acquire knowledge inorder to specify the need
complex	1. additive iterative refinement and substitution based trial and error and early termination using manual article evaluation to by-pass articulation difficulties 2. substitution based rapid prototyping evaluating large number of results to compensate	cut off points are determined by results set size rather than completeness of the need articulated
simple	1. substitution based trial rapid prototyping at the same granularity as terms in the need and early termination	alternative terms based on knowledge acquired from article evaluation
general but for general information	1. substitution based rapid prototyping with early termination	alternative terms based on knowledge acquired from article evaluation
general but for specific information	1. substitution based rapid prototyping evaluating large number of results to compensate	
precise	1 direct query composition without synonym elaboration	

Table 4.4 Initial possible strategy determination table domain and device novices

4.3.2.4 Possible strategies given re-formulation, diagnosis and user context

4.3.2.4.1 Domain and device experts

The searchers' knowledge allows them to produce tactics to by-pass search failures for any situation that may arise. The tactics available are the optimal and thus should be the basis for advice given as guidance. The searcher is able to diagnose the cause of the results not matching expectations and has the appropriate knowledge to overcome any obstacles encountered. The tactics operate at either the term or query level depending on the evaluation of the results.

Diagnosis category	Possible courses of action or combinations of actions	Dependencies and notes
terminate current IR cycle	1. articles selected acceptable to solve information need 2. current problem element OK but sub-goals exist (aggregate previous goals or move to next goal to be solved) 3. Motivation or time available low so return to previous queries relaxing judgements on article relevance	time, motivation, solution critical, outcome dependent on the evaluation judgements made
broaden query current terms at the correct level of granularity or no assessment made	1. broaden query by adding new problem concepts (OR) 2. broaden query by substituting query concepts or terms 3. broaden query by deleting query concept (if related to the rest of the query with 'AND') 4. broaden query by changing relationships if it doesn't alter the semantics of the information need 5. explore structure of problem, document space, index structures and concept definitions	
broaden query broaden terms	1. broaden query by adding new problem concepts (OR) 2. broaden query by substituting query concepts or terms 3. broaden query by deleting query concept (if related to the rest of the query with 'AND') 4. broaden query by changing relationships if it doesn't alter the semantics of the information need 5. explore structure of problem, document space, index structures and concept definitions 6. broaden concept articulation by adding term synonyms with OR relationships 7. broaden term by substitution with more general synonyms 8. broaden term by relaxing constraint clause 9. cycles of queries with successive term substitution (altering term granularity)	
broaden terms	1. explore structure of document space and index structures 2. broaden concept articulation by adding term synonyms with OR relationships 3. broaden term by substitution with more general synonyms 4. broaden term by relaxing constraint clause 5. cycles of queries with successive term substitution (altering term granularity)	
narrow query current terms at the correct level of granularity or no assessment made	1. narrow query by adding new problem concept (AND, WITH, NEAR, NOT etc.) 2. narrow query by substituting query concept or term 3. narrow query by deleting query concept (if related to the rest of the query with 'OR') 4. narrow query by changing relationships if it doesn't alter the semantics of the information need 5. explore structure of problem, document space, index structures and concept definitions	
narrow query narrow terms	1. narrow query by adding new problem concept (AND, WITH, NEAR, NOT etc.) 2. narrow query by substituting query concept or term 3. narrow query by deleting query concept (if related to the rest of the query with 'OR') 4. narrow query by changing relationships if it doesn't alter the semantics of the information need 5. explore structure of problem, document space, index structures and concept definitions 6. narrow concept articulation by adding term synonyms with AND relationships 7. narrow term by substitution with more specific synonyms 8. narrow term by tightening constraint clause 9. cycles of queries with successive term substitution (progressively altering term granularity)	
narrow terms	1. explore structure of document space and index structures 2. narrow concept articulation by adding term synonyms with AND relationships 3. narrow term by substitution with more specific synonyms 4. narrow term by tightening constraint clause 5. cycles of queries with successive term substitution (progressively altering term granularity)	

Table 4.5 Strategy reformulation determination table for domain and device experts

4.3.2.4.2 Domain expert but device novice

The tactics applicable are constrained by the lack of knowledge of the facilities provided by the device, and so any alteration to the queries must be based on domain knowledge or terms extracted from the articles retrieved by the previous query. This limits and restricts query development as the searcher aims to minimise device interaction. The facilities used are usually limited to the initial query interface.

Diagnosis category	Possible courses of action or combinations of actions	Dependencies and notes
terminate current IR cycle	1. articles selected acceptable to solve information need 2. current problem element OK but sub-goals exist so either formulate a query for the sub-goal or aggregate previous sub-goal solutions 3. Motivation or time available low so return to previous queries relaxing judgements on article relevance	time, motivation, solution critical
broaden query current terms at the correct level of granularity or no assessment made	1. broaden query by adding new problem concepts (OR) 2. broaden query by substituting query concepts or terms 3. broaden query by deleting query concept (if related to the rest of the query with 'AND')	Favour 2 or 3 as unlikely to use OR relationship
broaden query broaden terms	1. broaden query by adding new problem concepts (OR) 2. broaden query by substituting query concepts or terms 3. broaden query by deleting query concept (if related to the rest of the query with 'AND') 4. broaden term by substitution with more general synonyms 5. broaden term by relaxing constraint clause 6. cycles of queries with successive term substitution (gradual adjustment of term granularity)	unable to broaden by synonym addition with OR relationship due to difficulties with query formulation
broaden terms	1. broaden term by substitution with more general synonyms 2. broaden term by relaxing constraint clause 3. cycles of queries with successive term substitution (gradual adjustment of term granularity)	
narrow query current terms at the correct level of granularity or no assessment made	1. narrow query by adding new problem concept (AND) 2. narrow query by substituting query concept or term 3. narrow query by deleting query concept or term (if related to the rest of the query with 'OR')	Favour 1,2, 4, 5 as unlikely to have used OR relationship in previous query
narrow query narrow terms	1. narrow query by adding new problem concept 2. narrow query by substituting query concept or term 3. narrow query by deleting query concept or term (OR) 4. narrow term by substitution with more specific synonyms 5. narrow term by tightening constraint clause 6. cycles of queries with successive term substitution (gradual altering of term granularity)	Favour 1,2, 4, 5 as unlikely to have used OR relationship in previous query
narrow terms	1. narrow term by substitution with more specific synonyms 2. narrow term by tightening constraint clause 3. cycles of queries with successive term substitution (gradual altering of term granularity) 4.	

Table 4.6 Strategy reformulation determination table for domain experts but device novices

4.3.2.4.3 Domain novice but device expert

In this case the users tactics aim to promote the use of the device to compensate for short comings in domain knowledge. The query terms used are based on the information in the external task or those extracted from task support facilities, or the documents retrieved. Tactics focus on finding solutions to small parts of the task by leveraging the device facilities.

Diagnosis category	Possible courses of action or combinations of actions	Dependencies and notes
terminate current IR cycle	1. articles selected acceptable to solve information need 2. current problem element OK but sub-goals exist so either formulate a query for the sub-goal or aggregate previous sub-goal solutions 3. Motivation or time available low so return to previous queries relaxing judgements on article relevance	time, motivation, solution critical
broaden query current terms at the correct level of granularity or no assessment made	1. broaden query by adding new problem concept (OR) 2. broaden query by deleting query concept (if related to the rest of the query with 'AND') 3. broaden query by changing relationships if it doesn't alter the semantics of the information need 4. explore structure of problem, document space, index structures and concept definitions	
broaden query broaden terms	1. broaden query by adding new problem concept (OR) 2. broaden query by deleting query concept (if related to the rest of the query with 'AND') 3. broaden query by changing relationships if it doesn't alter the semantics of the information need 4. explore structure of problem, document space, index structures and concept definitions 5. explore term articulations in the device and using as term alternatives 6. broaden by relaxing constraint clauses	
broaden terms	1. explore structure of document space and index structures 2. explore term articulations in the device and using as term alternatives 3. broaden term by relaxing constraint clause	
narrow query current terms at the correct level of granularity or no assessment made	1. narrow query by adding new problem concept (AND) 2. narrow query by deleting query concept (if related to the rest of the query with 'OR') 3. narrow query by changing relationships if it doesn't alter the semantics of the information need 4. explore structure of problem, document space, index structures and concept definitions	
narrow query narrow terms	1. narrow query by adding new problem concept (AND) 2. narrow query by deleting query concept (if related to the rest of the query with 'OR') 3. narrow query by changing relationships if it doesn't alter the semantics of the information need 4. explore structure of problem, document space, index structures and concept definitions 5. explore term articulations in the device using term alternatives to narrow 6. narrow term by tightening constraint clause	
narrow terms	1. explore structure of document space and index structures 2. narrow term by substitution with more general synonyms 3. explore term articulations in the device facilities using term alternatives to narrow 4. narrow term by tightening constraint clause	

Table 4.7 Strategy reformulation determination table for domain novices but device experts

4.3.2.4.4 Domain and device novices

The tactics a searcher can employ are very limited as low domain knowledge means that term alternatives can not be recruited from memory and low device knowledge means the searcher is unable to use the device facilities to find alternative terms. The tactics concentrate on the addition or deletion of concepts or terms based on the content of the information need. The searcher is able to relax numerical constraints to widen the scope of the query but term substitution are limited to alternatives expressed in the original information need.

category	Possible courses of action or combinations of actions	Dependencies and notes
terminate current IR cycle	1. articles selected acceptable to solve information need 2. current problem element OK but sub-goals exist so either formulate a query for the sub-goal or aggregate previous sub-goal solutions 3. Motivation or time available low so return to previous queries relaxing judgements on article relevance	time, motivation, solution critical
broaden query current terms at the correct level of granularity or no assessment made	1. broaden query by adding new problem concept (OR) 2. broaden query by deleting query concept (AND)	Favour 2 as unlikely to use OR relationships
broaden query broaden terms	1 broaden query by adding new problem concept (OR) 2. broaden query by deleting query concept (AND) 3. broaden term by relaxing constraint clause	Favour 2 as unlikely to use OR relationships
broaden terms	1. broaden term by relaxing constraint clause	
narrow query current terms at the correct level of granularity or no assessment made	1. narrow query by adding new problem concept (AND) 2. narrow query by deleting query concept (OR)	Favour 1 as unlikely to have used OR relationship in previous query
narrow query narrow terms	1. narrow query by adding new problem concept (AND) 2. narrow query by deleting query concept (OR) 3. narrow by relaxing constraint clause	Favour 1 as unlikely to have used OR relationship in previous query
narrow terms	1. narrow term by tightening constraint clause	

Table 4.8 Strategy reformulation determination table for domain and device novices

4.3.3 Implementation

The implementation process initially allows the facilities to be used to execute plan of action for the information need. This process also enables the 'mental image' of the required information to be translated into the appropriate query syntax and semantics for its execution in the information retrieval system. The process maps the search tactics, strategies, concepts to the system to be searched. A users implementation success depends on the granularity of their knowledge and the interaction metaphors of the system, as these influence the users ability to translate the information need into a device dependent representation.

Search implementation can operate at the following levels:

- instance based experience (rehearsed and familiar problem situation);
- general rule based experience (domain specific rules);
- template matching or pre-formed query structures (interaction plans based on past experience and previous searches success)
- thesaurus and systems assistance based operation

- hit and miss query development (implement queries and amend activities by correcting error and action slips)
- systematic implementation and execution of the query

The implementation method is dependent on the knowledge held, motivation and the time available for the retrieval. For example if a searcher has high domain and device knowledge but is pressured by time or has low motivation then they are likely to try and produce a query that is sufficient rather than complete. If they are not pressured by time or are highly motivated then they may try a more iterative search making checks on success (especially relationship operators, term representations and order of execution).

The trigger situations for the implementation process are:

- *Strategy formulation:* plan of action consisting of strategies, tactics, concepts and terms,
- *Evaluation:* The possible reformulation outcomes are:
    - change device facilities used to implement query;
    - change term representations used to express concepts in the information need;
    - rectify action slip by changing relationships between query components or correct typing mistakes;

These different triggers for implementation enable activity to be targeted at the core refinement or execution problems.

The output of this process is observable as physical actions performed on the interface as the query is articulated to the computer system. The execution consists of a sequence of action steps performed on the retrieval system.

4.3.3.1 Activities performed

The activities performed in the implementation of a search are :

1. Refinement: This permits the searcher's information requirement to be constrained and directed towards the information contained in the retrieval system. The process

translates the plan of action into user interface commands or manipulations. A searcher's experience with the retrieval system frames refinement and execution activity. This complies with the results of Large et al (1994) that search strategies, and specifically access paths used to get to the information required, are a function of user's device related experience and not necessarily the most appropriate given the current information requirement. Refinement can be based on the adaptation of the current query in the light of search results rather than the production of a new query. The phases of the refinement process are defined as:

- Selection of the device facilities e.g. query interface, hypertext, concept map etc. A device facility may have mental scripts associated with it which specify the necessary interaction steps based on past search experience (Belkin 1993). If experience with system facilities is not available then interaction will be determined by trial and error.
- Selection of terms to represent concepts: The searcher selects the most efficient method of articulating the concepts required to the system, within the bounds of device knowledge and available facilities. The terms chosen to articulate a concept are based on the lexical variations and synonyms generated in strategy formulation. For example audit, audited, audits or auditing can be condensed to audit* in MEDLINE. The refinement tries to find the best match between search concepts and the database contents or index. Users can either guess a term, know it or explore external resources to find it (thesaurus, concept maps, navigation aids, experts etc.). The choice of which method to use varies by motivation, ease of execution and expertise.
- Select relationship operators: Boolean operators and order of execution are used to express the relationships identified in the information need. The use of combination and execution control operations allow a variety of different set relations to be incorporated into the query, but these are bounded by the operators supported and users understanding of their meaning.

2. Execution of the search: The execution of physical actions on the interface conveys the search to the retrieval device. This process manages the users interaction with the system and monitors the state of the interface for change. When an appropriate change occurs, as defined by expectation, the evaluation of a search is triggered.

4.3.3.2 Possible facilities given the current plan of action and user selection criteria

A list of the facilities which could be used by the user generated from the task support facilities (TSF) supported by the device (see section 4.5). The list is based on the task resource requirements (TSR) which are the operations that a specific task support facility can be used to carry out eg. provide a definition for a concept, explore the domain structure etc. Multiple TSFs may be applicable to a plan of action. The choice of which to use is based on the perceived difficulty of interaction and the perceived utility of the facilities based on the user's experience.

4.3.4 Evaluation

The evaluation permits a comparison of the results against the target information need, or goal, and the production of inferences about the cause of any differences between the two. Possible causes of differences are:

(i) the original information need included errors and inconsistencies,

(ii) retrieval system is perceived to be flawed in that it does not contain the information required or makes this information inaccessible to the user e.g. use of American spellings in thesaurus for article written in English,

(iii) the concepts used to represent the information need are incorrect or insufficient,

(iv) the terms used to articulate concepts are incorrect or insufficient,

(v) search granularity is incorrect.

(vi) Slips have been made e.g. typing errors or use of incorrect query syntax.

The searcher looks for clues in the results set for causes of failure. For example given a null set for a particular term a searcher may assume that either they have made an action slip (iv), they need to use a more complete term (iii), the item is indexed differently within the IR system than expressed in the query (v) or the item wasn't contained in the system (ii). Clues to which of these is the most likely cause are gained from the state of the interface; query used, systems facilities and results retrieved. The strategy, plan of action or query, can then be refined given knowledge acquired during the search. Inferences made during the evaluation (figure 4.6) dictate the re-entry point in the process model and these are incorporated into correspondence rules.

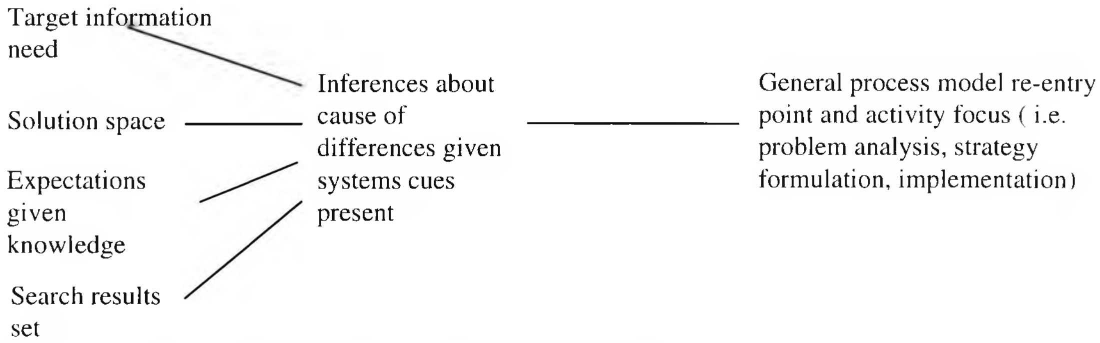

Figure 4.6: Cause and effect of the evaluation process

The evaluation process selects the records perceived to be relevant to the information need. Further search activity can then be specified according to how well the articles selected satisfies the information need.

4.3.4.1 Activities performed

The activities performed and the granularity of the evaluation depends on the search task, knowledge resources available and the ability to match the results against the search goal. The different evaluation tactics based on the amount of information used (e.g. quantity, abstract etc.) and the evaluation method for the results (e.g. skimming, reading etc.) are shown in figure 4.7. The evaluation process breaks down into four main stages:

1. Evaluate the number of records retrieved and determine the evaluation tactics (compare number produced with number expected),
2. Evaluate the records based on their information content (compare records retrieved to information need e.g. scan headers, sample first based on systems article ranking (first 5%), assess abstracts by sampling hits at random, sample hits systematically (1 in 5), skim abstracts or read abstract),
3. Evaluate the search successes and failures (compare strategy and need to results and articles selected as the solution set),
4. Select future plan of action, given the search results and the successes and failures identified.

The evaluation process differs from the other processes in the general model of IR in that the activities performed do not follow a linear progression. The outcomes of decisions made in the evaluation process are shown in figure 4.7.

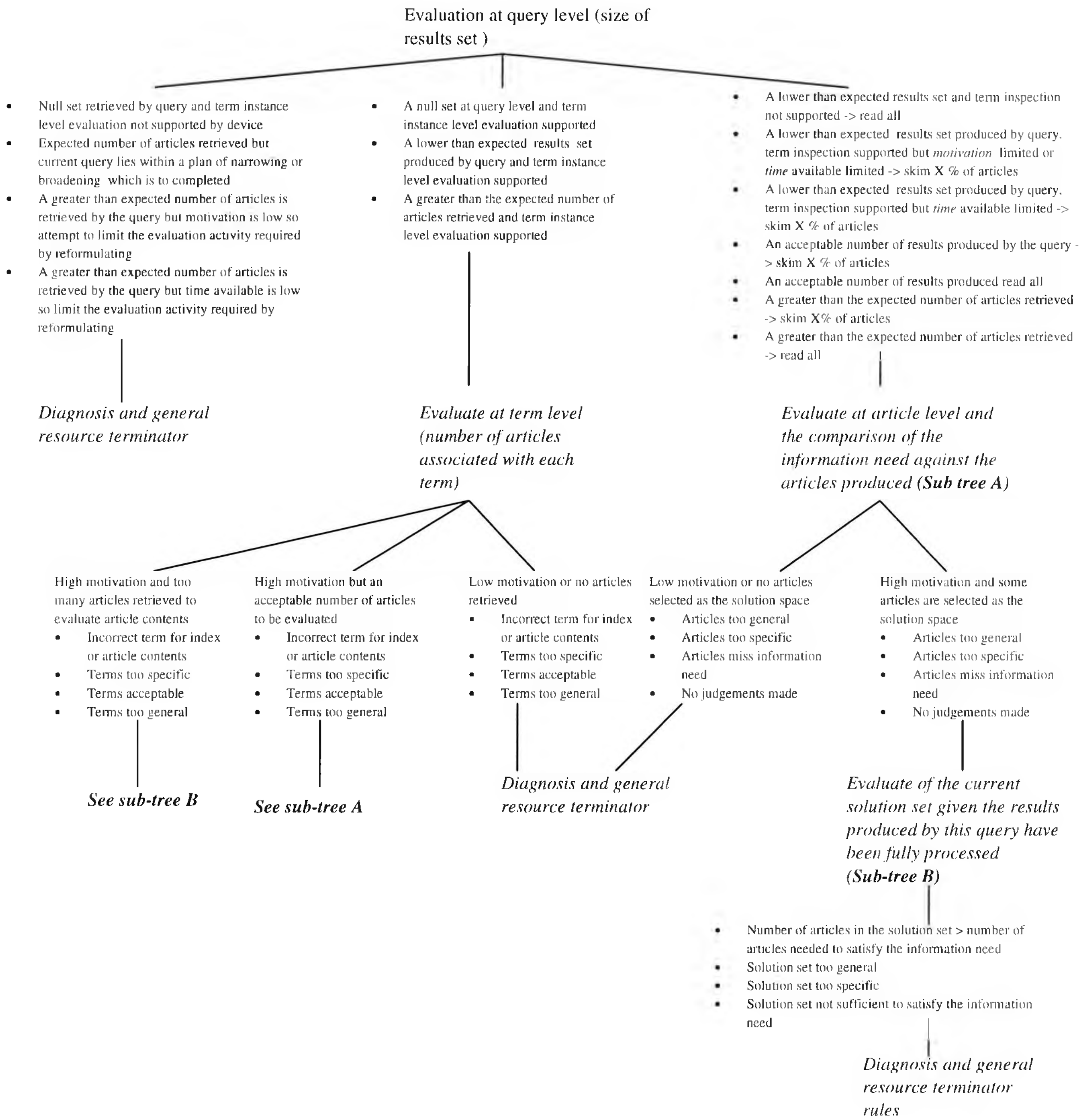

Figure 4.7: Decision tree controlling the activity performed in the evaluation process. The bullet pointed lists indicate the conditions under which this fork of the decision tree is followed. The tactic and strategy outcomes of these evaluation decisions are shown in tables 4.9-4.13.

The outcome of search diagnosis is dependent on the levels at which the evaluation is performed and the outcomes of these evaluations.

The decision tree uses an approximate categorisation of the result such as acceptable, high, low etc. and determination of category boundaries is dependent on the search context, user knowledge and user characteristics. For example searchers could react in the following ways to the retrieval of a large set of information:

- *Low device knowledge*: the production of a large amount of information improves their confidence in retrieving relevant information from the system. The searcher may thus proceed with detailed evaluation of the results early in the retrieval session,
- *High device knowledge*: the production of a large set of information is seen as a failure to specify the information required tightly enough. Obviously this depends on how general their information need is but in general the searcher is likely to reformulate the search.

4.3.4.2 Possible search diagnosis outcomes

Query level evaluation judgements (see table 4.9) are made based on the number of articles retrieved by the current query. The reformulation outcomes are very general due to the minimal amount of information available but they are used to direct high level reformulation and determine other evaluation tactics to use e.g. article skimming, reading etc. The effect of the user's expectation and knowledge (domain and device) on the diagnosis will depend on the information need type and a users perception of the results. No attempt is made to specify what constitutes a low number of results for a query as this depends on a multitude of factors not least of these being motivation and length of time already spent on a retrieval session.

category	Evaluation level	Possible courses of action	Dependencies and notes
Null result	Query level	1. broaden query 2. broaden query, broaden terms 3. broaden terms 4. terminate retrieval for this goal 5. shift problem focus	
Low number of results produced	Query level	1. evaluate articles for further diagnosis (read all / skim X%) 2. broaden query 3. broaden query, broaden terms 4. broaden terms	low knowledge categories may read all as compensatory mechanisms whereas high knowledge will only skim a percentage
Expected number of results retrieved	Query level	1. evaluate articles for further diagnosis (read all skim X %) 2. incorporate query into other queries as progression of retrieval	low knowledge categories may read all as compensatory mechanisms whereas high knowledge will only skim a percentage
excessively high number of results produced	Query level	1. evaluate articles for further diagnosis (read all/ skim X%) 2. narrow query 3. narrow query, narrow terms 4. narrow terms	low knowledge categories may read all as compensatory mechanisms whereas high knowledge will only skim a percentage

Table 4.9 Search diagnosis determination at the query level

Evaluation judgements made at the query component level (see table 4.10) fall into the same categories as those at the query level, but the outcomes operate on the terms used to articulate concepts instead of the whole query. Evaluation of query components depend on the ability to inspect the results at this granularity, which is a function of the presentation facilities provided by the device.

category	Evaluation level	Possible courses of action	Dependencies and notes
Null result	Query component level	1. broaden terms by substitution 2. delete terms as information is determined to not be in the system 3. explore information to develop articulation used for term	
Low number of results produced	Query component level	1. broaden terms by addition (OR) or substitution for super-ordinate 2. explore information to develop articulation used for term (in device or results set)	
Expected number of results retrieved	Query component level	1. maintain current granularity of the components used in queries	No alterations made at the query component level
excessively high number of results produced	Query component level	1. narrow terms by addition (OR) or substitution 2. explore information to develop articulation used for term (in device or results set)	

Table 4.10 Search diagnosis determination at the query component level

Evaluation judgements are made on the relevance and quality of articles to the information. The categories are broad as judgements of relevance are complex and user specific (Su 1994, Barry 1994). The important aspect of this type of evaluation is not how relevance decisions are made but what effects these decisions have on the strategies and tactics employed.

category	Evaluation level	Possible courses of action	Dependencies and notes
Articles too general	Articles produced by the current query level	1. narrow query (substitution or more specific articulation) 2. narrow query, narrow terms 3. narrow terms	
Articles too specific	Articles produced by the current query level	1. broaden query (substitution or more general articulation) 2. broaden query, broaden terms 3. broaden terms	
Articles at the correct level of granularity	Articles produced by the current query level	1. Maintain current granularity of the components used in queries	No alternations made as content matches need granularity
Articles are none relevant information to the need	Articles produced by the current query level	1. change the focus of search activity	

Table 4.11 Search diagnosis determination at the articles produced by the current query

The searcher may make evaluation judgements not only on the success of the particular query just performed but also on the success of the retrieval session up to that point in time. These evaluation judgements use the same diagnosis categories as those used to assess a specific query, but the outcomes in terms of strategies and tactics are different.

category	Evaluation level	Possible courses of action	Dependencies and notes
Selected articles are too general	Articles selected as part of the set which constitute the answer to the information need	1. narrow query (substitution or more specific articulation) 2. narrow query, narrow terms 3. narrow terms	
Articles too specific	Articles selected as part of the set which constitute the answer to the information need	1. broaden query (substitution or more general articulation) 2. broaden query, broaden terms 3. broaden terms	
Articles at the correct level of granularity but not enough information to satisfy need	Articles selected as part of the set which constitute the answer to the information need	1. broaden query	
Articles at the correct level of granularity but not enough information to satisfy need	Articles selected as part of the set which constitute the answer to the information need	1. terminate IR cycle	

Table 4.12 Search diagnosis determination at the articles selected as the solution to the information need during the retrieval session

Finally search termination judgements made based on the resources available to pursue further retrievals. These judgements over-ride the diagnosis decisions already discussed and have two possible outcomes if triggered: search termination or the re-evaluation of results of previous queries with relaxed matching constraints followed by search termination e.g. if a user reaches a 'give up' decision they may return to previous queries extracting articles.

category	Evaluation level	Possible courses of action	Dependencies and notes
Time	Resource related diagnosis	1. terminate IR cycle and make best of retrieval already performed	
motivation	Resource related diagnosis	1. terminate IR cycle and make best of retrieval already performed	

Table 4.13 Search diagnosis determination based on user termination boundaries

4.4 Correspondence rules

Correspondence rules dictate the decisions and activities performed in the model based on the need type, current context, user knowledge and changes in these as the search progresses. It would be unrealistic to expect all searches and behaviour patterns to be represented by a single configuration of the process model due to the diversity of tasks in information retrieval, differing levels of detail in users' knowledge and the history of a search. Thus a sub-set of the correspondence rules available are used to predict the activity, strategies and tactics for specific search and user characteristics. The rules adapt, configure and target the generic model based on the situation faced. The correspondence rules operate using a *if preconditions A, B, C are satisfied then perform activities F, G.* Activities F, G can be specific strategies, tactics or actions on the problem. The correspondence rules have been condensed into the tables of outcomes already addressed in sections 4.3.2.3, 4.3.2.4 and 4.3.4.2.

The correspondence rules are based around the assumption that they utilise a general cognitive architecture. Thus issues relating to how: long term memory (LTM) is represented, perceptual mechanisms effect information encoding and working memory (Wm) is represented are, at present, out of the scope of the model. The model is an information processing model which tries to explain, and predict, the cognitive activities and reasoning actions which underlie a searchers' behaviour. The correspondence rules act as the decision mechanisms in determining the manipulation and transformations of information applicable as well as the strategies and tactics to be used

4.4.1 Information need types

As search strategies are partially determined by the need (Taylor 1968, Ingwersen 1996), reflected in the user's goal, there is considerable motivation to taxonomise need types. Need types, may change as searching progresses as the user discovers new needs after evaluating retrieved results. The user's need is characterised by the following dimensions:

Generality- either high level information (general) or low level detail may be required (precise). This describes the user's expectation of the level of detail required at the beginning of the search. A precise target occurs when the user knows either the identity or a specific description of the search target e.g. drug interactions between streptokinase and paracetemol; this is contrasted with a search when users require only general information e.g outline changes of government policy on the NHS since its conception. Note that precise knowledge of the target is a function of known-ness of the need, but it is possible to have an unknown, yet specific need (e.g. when you have forgotten a specific reference but you know it is in the database somewhere).

Known-ness- how well the user can describe the required information is characterised along the dimension of poorly known to well known. This dimension may interact with generality and will change during the search (cf. ASKs Belkin et al 1982). The familiarity with a task domain also effects the level of device support required if a search is to be successful.

Complexity- an approximate measure of the search target in terms of the number of components and links in the information need and the number of separate items in the information need. Complexity is measured by the connectivity in a concept level representation of the need extending Saracevic and Kantors' (1988) view which was based on the number of concepts in the task. Aspect queries, for instance, tend to be complex. The complexity of a task alters the granularity and amount of cognitive processing required if satisfactory results are to be obtained Large et al (1994), Borgman (1985).

Information types- the types of information required are currently only described as values or text. With multimedia databases this distinction is becoming blurred; however, value/text queries have different connotations in terms of the strategies and tactics applicable, e.g. use of ranges in value domains to narrow queries.

Solution critical/ none critical solution: consequences of search success or failure for the users task. The effects of this characteristic are seen through the limits set for motivation, evaluation cut-offs and time available for the retrieval. This need type is different from the others as it is based on the users perception of the value of the information required.

The information need determines strategy formulation, according to the user's knowledge of the domain and the search device within the constraints imposed by time and user motivation. The latter is a compound of importance of the need and general motivation.

4.4.2 Context factors effecting search activity outcomes

The other factors affecting search behaviour are:

Motivation (Jacobson et al 1992), time, costs, value of information, payoffs (Irwin et al 1957 and Wendt 1969): These factors each effect the longevity and granularity of search activity. Their influence is dictated by correspondence rules which control cut-off points in search activity based on the search history and critical nature of a need.

Cognitive effort/ resistance (Eysenck and Keane, 1990). The cognitive effort/ resistance applies to the perceived usability of the system as well as the mental behaviours necessary. This is a variable the user/ model minimises by favouring strategies and tactics which play on knowledge strengths. For example device experts may favour strategies which utilise the systems facilities rather than reasoning intensive strategies. Compromises are made between completeness and associated effort in determining the most appropriate course of action for the search.

User expectations: Expectations of the search effort effect the duration of searches and decisions made in search evaluation. The ability to produce realistic expectations depends on domain and device knowledge. Flaws in user expectations can lead to search reformulation or premature search termination. Expectations are used to assess the effort associated with a particular course of action and the relative success of queries.

4.5 Information retrieval systems functionality

The IR system is represented as a number of task support facilities. The different facilities and their interaction metaphors support different aspects of users retrieval activities. The different types of task support facilities are categorised as follows:

- Exploration support (user knowledge expansion or compensation for),
- Query expression mechanisms (need articulation),
- Search evaluation support (search results presentations).

Guidance and help facilities are an exception to this categorisation as they operate at all these levels in an attempt to inform the searcher about the characteristics of the device and the search domain. The representation of task support facilities (TSFs) can either be a perfect system or a specific system being tested, i.e. a sub-set of the TSFs that models the functionality of the specific system being used. In this way the effects of altering the interface designs, or altering the TSFs supported by the system, on user search activity can be predicted and simulated. A TSF is represented as follows:

- Identifier (e.g. thesaurus)
- Task support requirement -TSR (e.g. concept expansion, concept definition)
- Interaction steps (e.g 6 interaction steps are required to execute this TSF for a single concept)
- Utility to the user based on past experience. This is a users belief about how successful a search will be using this facility, conforming with Large et al (1994) findings that users often use facilities linked to past success.

Table 4.15 shows the taxonomy of the TSFs which assist in the exploration of the different possible terms for search concepts. Most systems offer a selection of these facilities and use different interaction and results presentation metaphors. In our modelling the aim is to provide design advice specifying the facilities which assist users in their retrieval activities.

Problem space exploration support (concept expansion/ exploration)		
Facility	**Description**	**Examples**
Thesaurus	Active or passive term expansion or definition generators	OKAPI, MEDLINE: best alphabetical match
Concept maps	Graphical navigation metaphors which allow information space navigation, possible term alternatives or the document space	tables, maps, network diagrams, subject trees and hierarchies of meta-data or classification structures
Metadata dictionary	Descriptions of available databases and their contents	Microsoft's search pages showing abstract descriptions of information sources and allows brokering between different search engines
Indexing structures, Subject guides	Information classification structures and indexing terminology	MEDLINE's index search shows the structure of the different indexes on the information for the database
Hypertext navigation	Hot link based navigation mechanisms through the document space	Microsoft's help facilities allow the navigation of help information through hard coded hot links between related material, web based search engines or site maps
Term suggestion facilities	Active or passive generation of term alternatives	MEDLINE: term suggestion facilities

Table 4.15: Taxonomy of problem space exploration facilities

The facilities provide different methods of relating a searcher's view of their information need onto systems categorisations of the domain.

The taxonomy of problem expression facilities (table 4.16) shows the different methods which may be used in the articulation of a information need.

Need articulation/ problem expression mechanisms		
Facility	**Description**	**Examples**
Boolean query languages	Allows the specification of relationships between problem elements in respect to the documents in the system.	SQL; MEDLINE query language: AND, OR, NEAR, NOT, WITH, IN, etc.
Syntax directed editors	System provides assistance in the production of queries by targeting user retrieval through the interface provided. The system removes the complexities of constructing SQL queries and translates user selections into the relevant database language.	Access query view assists a searchers in query syntax construction.
Keyword queries (non Boolean- implicit ANDs, ORs)	Systems based around the use of a number of keywords which are bound by AND's. Enables narrowing or broadening by addition or deletion of terms.	Excite search engine on the WWW presents the results of queries AND keyword combination first followed by OR keyword combination.
Query by pointing , Query by objects	Hypertext concept maps or hotspots on realistic domain images bounded by the current problem space.	Access allows the selection of objects in the database to be searched.
Query by example or retrieval by similarity	Find similar related items given the retrieved results or specific items in the results set.	Lycos or webcrawler (WWW) allow the selection of related articles to a specific article.
Pre-formed queries, Template based form fills	Provide guidance towards profitable sub-areas and the structuring of a problem space through the query interface.	No example known.
Re-usable queries, Progressive query refinement, Iterative query development and the ability to refine searches	These enable problem development and its reuse as part of iterative query development. The mechanism provide shortcuts to these developments.	Embedded menus, combination of previous queries. MEDLINE allows search history re-use through labelling metaphor.
Query filters	Restrict retrieval set based on pre-defined value operators whose characteristics are controlled by user interaction with filtering mechanisms	Dynamic query sliders act as the interface to the document space: Shneiderman (94).
Negotiative query development	System takes on the role of the librarian intermediary to enhance need elicitation. This uses conversation metaphors for need elicitation in which systems advice and the query dialogue assist in the development of the query articulated.	CONNIT retrieval system in which user critiques the results produced
Constraint based querying	Restrict the solution space based on the physical characteristics of documents	MEDLINE allows a searcher to restrict the document set produced based on pre-defined document characteristics. e.g. date.
Query expansion by activation and user defined scope of conceptual distance	User controlled spreading activation, controlling scope, focus of attention and level of detail, guidance towards profitable sub-areas	OKAPI allows a user to incorporate systems suggestions based on the query elements.
Menu driven querying	Menu driven structures guide the searcher through options which allow the narrowing of the document space.	Niss, Bids, provide menu driven structures to enable the narrowing of the problem space by structuring interaction
Term weighting by statistical properties or user defined problem area dominance	Allow the determine of the order of results relevance to a specific query based on the importance (perceived or implied) of the elements in the query and their occurrence in the documents to be ordered.	OKAPI allows term weighting based on statistical properties. No known example of a system whereby the user can specify the ranking of results based on the importance of query concepts
Fuzzy matching	Truncation and term extension operators, beginning, end or middle truncation	MEDLINE: use of * and ? to truncate problem space elements
Multi- threaded search engines	Ability to search multiple databases and to target aspects of the information need at specific information sources.	Brokering search facilities in which one view of the information space is used to send queries to multiple databases. Microsoft's web search pages do this in a limited form.
Elimination by localisation in current possible document space	Articulation mechanisms allowing a searcher to alter the granularity of the possible document space by selecting areas of the problem domain for further investigation.	MEDLINE allows the re-use of a component part of a query in a subsequent investigation.
Query by sketching	Enables the production of image queries based on the user sketches of the information required.	querying by sketching.(Charles 1990)
Natural language querying	Full or restricted natural language to permit a user to express their information need as if talking to a human.	PLEXUS tries to automatically convert an informal user expression of the information required into a formal query (Vickery et al 1987)
Automatic query reformulation	Adjustment of the query used the searcher based on their current context	RABBIT query reformation based on users critiquing of the results retrieved (Williams and Tou 1982)

Table 4.16: Taxonomy of query articulation mechanisms

The query mechanisms provided by a specific system are a sub-set of this taxonomy. This aspect of the model enables the impact of different configurations of facilities on user's behaviour to be simulated. The main problem for IR designers is deciding which facilities to provide, how to be systematic about this functionality selection and how to choose the most effective interface organisation and presentation metaphors for this functionality.

The search evaluation facilities offered, their granularity and the operations they allow for problem development effect any subsequent retrieval activity.

Search evaluation support		
Facility	**Description**	**Examples**
Quantity view	Number of hits in a specific category	Scatter plots, tables, string, starfield displays, bar charts, Venn diagrams, density maps, graphic maps, web views, global maps and local maps, path, a map and a scope line, Fish eye views, zoom techniques
Clustering techniques	Allow the clustering of articles and the information space around a categorisation schema	LSI (Deerwester et al 1990), Scatter gather (Pirolli 1996)
Article maps of the resulting document space produced by a query	Dynamic maps or representations of the articles associated with elements of the queries posed. This goes further than simple quantity data as it gives a perspective on the semantic distance between the articles retrieved and the query elements	Ahlberg and Sheidermans (1986) starfield displays and hit density displays
Article summary	Displays document characteristics and high level views.	MEDLINE displays title, date, author and language in summary display mode allowing skimming and speeding up results evaluation.
View results contents	Abstract or whole document view of the results set	MEDLINE full record display enabling the evaluation of articles based on their semantic content.
View mark and re-use in query development, query element weighting based on results sets	Relevance feed back of document space produced, Book marking of results	OKAPI relevance feed back mechanism
Back tracking, history lists, time stamps and foot prints	Representations of previous problems which can be iteratively refined	MEDLINE allows parts of previous queries or whole queries to be re-used using systems designated labels.
Relevancy ranking of results produce	This depends on how relevancy is determined and conveyed to the searcher but it is an attempt to rank query results based on the components of the users need.	OKAPI ranks retrieved results based on statistical probabilities of relevance. No known example of relevancy ranking based on user specifying which of the query components are most important to their information need.
Adjacency indicators	Highlights and show articles, subject areas or terms loosely coupled to the results produced.	OKAPI related terms display

Table 4.17: Taxonomy of evaluation support facilities

The granularity at which users are able to inspect the retrieved results and the clarity of the presentation metaphor used effects how they react in terms of the problem development activities, strategies and tactics performed. As a consequence results presentation effects evaluation strategies and subsequent retrieval behaviour owing to the precision of search diagnosis (see section 4.3.4). The representations used for results need to support evaluation at these different levels, i.e. size of results set, number of articles associated with a concept, gist level overview of article and article content, and transitions between these different levels of information should be simple to implement as an attempt to prevent the interface from being overloaded. Attentional design techniques should be used to assist user evaluation at the different levels of evaluation e.g. highlighting term occurrences with articles to assist skimming techniques and relevancy judgements. The representation of the articles retrieved should allow a user to trace the articles they have evaluated very quickly and easily, e.g. visual highlighting of the level at which articles have been reviewed, as this will facilitate judgements on the quality of the results retrieved. The ability to determine the quality of the results set in relation to the query is essential if the query is to be updated and targeted to retrieved information relevant to information need satisfaction.

4.6 Systems representation

This is a representation of the state of the information retrieval system at any point in time. This is included in the architecture as although it is not a part of the cognitive model it influences the reaction of the searcher and diagnosis decisions made. The systems state has the following characteristics: mode and information content. The mode of the systems state is dependent on the active task support facilities chosen from the taxonomies of facilities supported (see section 4.3.7). The information contained in the active facilities are the content of the results e.g. number of articles, relevance, precision. The state of the system serves as an input to the process model of IR through evaluation reactions and reformulation decisions, strategies and tactics.

4.7 Model Walk through

In this section a walk through of part of the model is presented for a specific information searching scenario. The scenario is driven by a complex, known and specific information need. The user is described as having high domain and low device knowledge (experimental task OG2 in chapter three).

The user initially identifies the elements (concepts and relationships) within the information need. The complicated nature of the need causes the initial use of divide and conquer (sub-goal) strategies as the need is too complicated to be addressed as a single search and involves many different aspects in relation to the same core concepts (planned or emergency caesarean section). The subsequent strategy is to form a main results set for the core need concepts and narrow this in relation to the different search aspects (maternal and foetal safety, infection control, statistical and economic). The possible strategies for a user with this need and knowledge profile favour the use of domain knowledge and the use of a minimum of system functions. The user selects the strategy of refining queries in the device; as the interaction and effort necessary to substitute the different search aspects into a simple query is minimal; as the structure of the query and syntax is only attended to once. The initial query is for caesarean sections and this produces a high number of results, so the query is refined by adding an extra concept to the query (emergency sections and safety) which produces too few results. The subject thus broadens the term used to articulate safety using maternal in its place. The results retrieved are acceptable for this aspect of the information need. The subject substitutes the safety aspect of the query with infection. This cycle of substitution behaviour continues until all aspects of the information need have been covered.

4.8 Discussion

This chapter has described a complex yet adaptive model capable of simulating the retrieval behaviour of users with different types of need and knowledge characteristics. The model ties user retrieval activity to the context of the information system through the taxonomies of facilities present.

The user model predicts behaviour given a certain need type and user configuration and enables this to be compared with optimal 'gold standard' strategies, or process configurations. The four possible foci for using the model are:

- Predicting the task support facilities which enhance users searching activity given certain need characteristics and searcher contexts,

- Embedding the model within the system as rules to give guidance to searchers in their retrieval activities (intelligent IR guidance based on a user profile, query characteristics and search history).
- Improving interface design by supporting the cognitive activities performed in information searching. Part of this improvement will be achieved through attention to dialogue design or information presentations geared to support and encourage user process,
- Assessing the implications of alternative design decisions on the effectiveness of search activity, as prescribed by the model. The effects of different designs on users ability to retrieve the information required can be compared in an attempt to develop customised IR systems.

The development of the cognitive model has enabled some of the complexities found in searching behaviour to be explored. This chapter has given an overview of the IR process and highlights the strategies, tactics and activities which need to be supported by IR systems.

The model accounts for a variety of contexts in which the effects of need types and user characteristics on the search process can be predicted. Investigation of the search process can indicate areas in search behaviour when there are likely to be errors, user problems and divergences from optimal search plans. These are explored as good design properties for the development of systems interfaces which will improve the retrieval behaviour for the various user groups (Chapter six). The model highlights the activities searchers' attempt to perform in the retrieval process. This information can be used to provide support tools aimed at enhancing these activities to assist the searcher.

The model has put the different IR strategies, e.g. searching and browsing, in the context of their underlying processes and as such gives a greater insight into the strategies, tactics and problem development activities requiring support, and how this can be best achieved. The necessity to distinguish between the behaviour of experts and novices has also been established highlighting the different types of retrieval support required. The model enables the testing of the effectiveness of different TSFs in supporting the retrieval activities of its user population. The model is evaluated by comparing empirical observations of users information retrieval against the models predictions of behaviour for a specific situation in chapter five. Design implications attempt to link the model of human action to what users need to know and the task support provided by an IR system, in an attempt to indicate and justify design requirements. In this way attempting to define the facilities which support user process in chapter six.

Chapter 5

Evaluating the model's predictions

This chapter describes the results of a protocol study to enable model validation

Chapter 5: Evaluating the model's predictions

5.1 Introduction

This study examines the cognitive activities performed in the retrieval process. User information retrieval sessions were studied through examinations of the verbal reports of the activities performed. The empirical work presented in this chapter aims to investigate patterns of search behaviour for a user population and to define the task support facilities which aid the production of complete, efficient and effective searches. This chapter is an initial validation of the model of information retrieval and in doing so addresses the impact of the following on retrieval activity:

- the effects of variations in user task and thus need type on the users' reported activity;
- the effects of variations in user device knowledge on activity.

A further aim is to investigate if the interface and task support facilities (TSF) of a typical IR system are supporting the retrieval activity of all types of users. This in itself is a measure of the quality and effectiveness of systems as it evaluates the match between the activities users try to perform and the effectiveness of IR systems at meeting that demand. The study is motivated by the desire to validate the predictive power of the model of IR (presented in chapter 4) which indicates the task support facilities needed to support users' problem solving and information searching.

This chapter expands the analysis of the experiment data collected and reported in chapter 3. The experiment performed produced a very rich set of data and thus the analysis was split into a quantitative analysis (chapter 3) and verbal protocol data (chapter 5) for users retrieval sessions. The results found in chapter 3 and current literature enabled theory development. The verbal reports were not analysed until initial theory development was complete. The analysis and categorisation of verbal reports was undertaken to validate the theory. The subjects, tasks and experimental design reported in this chapter are the same as those found in chapter 3.

5.2 Experimental Method

5.2.1 Experimental design

Protocol analysis was used to investigate the information retrieval behaviour of seventeen medical students (thirteen males and four females, aged between 24-26 yrs) who were three months away from their clinical final examinations. These subjects were the same subjects who produced the data reported in chapter 3. Three pilot subjects undertook the experiment before the trials were started, to refine the medical scenarios for the information required and the experimental procedure. The protocol of one of the subjects had to be discarded due to her failure to verbalise. Data from the subjects provide the basis for the results discussed in this chapter.

Subjects were requested to 'think aloud' giving verbal protocols. The protocol, with associated physical actions, were recorded on video and audio-tape. The subjects were advised to take their time when verbalising and not be afraid of verbalising too much, in line with the practices of Ericsson and Simon (1984). To train the subjects to produce verbal reports they were required to 'think aloud' whilst performing a simple eight puzzle. The puzzle task was intended to familiarise users with verbalising whilst operating a computer. The experimenter encouraged the subjects to report the reasons for behaviour at interaction points, indicating why a specific action had been performed. The puzzle task interactions were designed to include time lags in order that the experimenter's prompts to verbalise caused minimal interruption to the subjects' behaviour.

Subjects were requested to complete a pre-test questionnaire to score their domain and device knowledge allowing them to be categorised as experts or novices in relation to the MEDLINE interface and system. Questions were chosen carefully so as not to relate to the task scenarios. To ensure a high and uniform level of device knowledge for the expert group and to accentuate the differences between the groups, the ten subjects with experience using MEDLINE were given training in the use of the information retrieval system. The system facilities and brief guidance on the production of strategies were explained in a tutorial document and a demonstration (see appendix 3b for training documentation). The subject domain of the demonstration was learning disabilities and thus far removed from the scenarios of the experiment. Novices were not trained.

All subjects were given 4 medical problem scenarios requiring them to find information to solve an information need:

1. Please use the MEDLINE database to investigate the socio-economic reasons for increased failure rates on the oral contraceptive pill; **(Code OG1)**

2. Please utilise the MEDLINE database to compare: caesarean sections being performed in planned and emergency situations from the standpoints of: maternal and foetal safety, infection control; statistical and economic data; **(Code OG2)**
3. Using the MEDLINE database please assess the importance of blood sugar levels and lipid profile in the cause of myocardial infarction within the male population; **(Code PH1)**
4. Utilise the database to determine some areas of increased and decreased efficiency within the N.H.S. since the introduction of formal clinical auditing. **(Code PH2)**

Subjects had access to the task scenario during a trial and were instructed that they could use the sheets to record information as they searched. Each scenario was given to the subject after they had reported that the preceding task had been completed satisfactorily, i.e. they did not have access to past search scenarios once completed. The subjects' notes; physical actions; search strategies; views of the problem to be tackled and search history were recorded. The tasks were presented to the subjects in a random order to standardise for learning effects. When questioned after the experiment, the subjects confirmed that the tasks were similar in style and complexity to ones they would normally use MEDLINE to solve. The experimental task used are the same as those reported in the empirical investigations reported in chapter 3.

5.3 Data analysis

This chapter explores the activities searchers performed for the four retrieval tasks. This is achieved by describing activity occurrence, the rate at which activity was performed and activity occurrence relative to the other activities performed in retrieval sessions. The investigation of the rate of activity occurrence aimed to indicate any effects user knowledge had on the speed with which a particular category of activity can be completed. A high level structure for the activity performed is used to test if alternative strategies are used on particular tasks or if elements of behaviour follow the same structure independently of the information need. Differences are explored for significance. The analysis investigated the inter-relationships between activities by describing the sequence of the activities performed using state transition diagrams. The structure of this analysis follows the methodology of Bakeman and Gottman (1986) using conditional transitional probabilities. To conclude this chapter the results are compared against the cognitive model described in chapter 4.

5.3.2 Mental behaviours

Protocol transcripts were analysed by matching mental behaviours to speech segments in accordance with Ericsson and Simon's (1984) method. Thirteen categories of mental behaviour were used in the analysis. This analysis produced a sequence of mental behaviours for each task and subject. The data was entered into an Access database along with user characteristics. This allowed the comparison and identification of patterns of mental behaviour sequences based on the user and task characteristics. The following categories were used:

Identify Concept	Articulation of new terms/ keywords. (present in the scenario or the device) e.g. ' so this wants myocardial infarction ......'
Identify concept relationship	Reports of concepts/terms with appropriate linking phrases. e.g. '... and within that...' or '... as a subset of that ...'
Formulate strategy	Verbalisations of planning within the search, and the indication of a pre-search map of activities to be performed. This occurs if the verbal protocols of a subject indicate they are trying to find a method of solving the problem. E.g. ' how are we going to ....'
Refine strategy	Amend or extend the existing plan of action for their current strategy. This category is dependent on the state of the current strategy changing. E.g. 'lets extend this by narrowing it down...' or 'lets attack this from a different perspective ...'. The verbalisation must indicate strategic decisions whereas reasoning about the problem is concerned with the structure of the search domain.
Reason about the problem space	Subject verbalises the search option and the reasons why they have chosen one over another. e.g. 'I could always use NHS but that will not help me as it can mean many different things so I'll leave that for now.....'
Explore synonyms for concepts, terms and their inter-relationships	Searchers report looking for lexical variations and synonyms for terms/ concepts already in the search problem space. Reasoning about possible terms to be used and how these should be inter-related. e.g. ' so that's oral contraceptive...or ... OCP.. or possibly contraceptive pill ...' the relationships are bounded by the scope of the concept being explored. Explicitly looking for variations on terms.
Select terms	Selection of terms for the current search. This only occurs as an activity if the searcher chooses a term from a previous verbalisation of possible search terms.
Formulate Query	The subject translates their information need into a query. This activity is observed when the subject produces a new search, which does not incorporate aspects of the previous search, but does refer explicitly to composition of query syntax. e.g. 1 inceas* Form Query 2 decreas* Form Query
Revise Query	The subject produces a search, which is an amendment to a previous query. This phase is the composition of previous queries or editing the existing query. e.g. 1 inceas* Form Query 2 decreas* Form Query 3 #1 AND #2 Revise Query
Execute actions to implement	Interaction with the computer to implement the search.
Evaluation of initial systems results	Evaluation of search results presented by the system is inferred. This activity has occurred if the verbal protocols of a subject indicate that they are evaluating their search based on the context in which the verbal report is found. e.g. ' so lets have a look at what this has found ...' or ' so that's 222 articles ...' This activity occurs also when a searcher removes dialogue boxes associated with a null search.
Evaluate the content of the results	Inferred from users observed actions and verbal protocols that indicate articles have been read to ascertain their importance to the information need. This may be based on the time spent reviewing an article; the display the subject has selected for the articles retrieved; moving through the results set using scrolling operators and by the subject paraphrasing the article's content. e.g. ' ...so this ones concerning ....'
Evaluate, report and diagnose search success/ failure	Inferred from verbalisations as subjects review the results/ articles retrieved. These reports may indicate that the results expected didn't occur and the search should progress accordingly. e.g. ' ... just checking to see if I've got enough information to ...' or ' ...this isn't quite what I want ... it doesn't ...' The verbalisations tend to precede a decision to reformulate or give up.

Protocol categorisation rules. During the search process many retrievals may be performed before the results set is inspected; furthermore, the searcher may return to previous results sets if they do not find the information required. If this was observed it was categorised as 'execute actions' to alter the current results set followed by subsequent evaluation.

5.4 Results

5.4.1 Impact of user device experience and task on the mental behaviours produced

A frequency distribution for the different behaviours was created from the category analysis of verbal protocols. The effects of task and users' device knowledge on the observed behaviour were investigated. The analysis considered three measures of activity: the average frequency, the average rate and the number of times a specific activity category occurred as a proportion of the total activities observed. The purpose of the last of these three measures is to indicate if the configuration of mental activities are affected by device knowledge.

5.4.1.1 Results

Table 5.1 highlights some interesting differences in searchers' activity frequencies. Expert searchers perform on average a third more transitions than novices (see table 5.1). The differences between experts and novices occur for: strategy revision, problem reasoning, synonym exploration, new query formulation, query revision, implementation actions and the evaluation of the number of articles retrieved, for all of which experts perform substantially more frequently than novices on all tasks. It is also interesting that the number of times article evaluation and search diagnosis are observed do not vary greatly with device knowledge for the same task, but do vary across tasks. The majority of differences in the frequency of activities for experts across the four tasks are restricted to the execution of the device and the results interpretation (i.e. identify concept, form query, revise query, evaluate initial results and evaluate the content of the results). The activities observed in the behaviour of novices show less variance for querying based activities than experts, with the exception of task PH2 where subjects appeared to give up.

Category	OG1		OG2		PH1		PH2	
	Expert	Novice	Expert	Novice	Expert	Novice	Expert	Novice
Identify concept	3.22	2.57	7.67	4.14	4.22	3.43	3.33	2.71
Identify relation	1	1	2.56	2.14	1.56	1.57	0.89	1
Form strategy	1	0.86	1	1	1	1	1	1
Revise strategy	4.22	1.86	4.89	2.43	4	2.43	3.56	1.71
Reason about problem space	3.11	0.86	3.89	0.71	3.44	0.86	3.22	0.86
Explore synonyms	5.44	2.57	6.11	1.43	4.67	2.14	3.67	0.86
Select terms	2.33	2.57	3	1.43	2.67	2.28	2	1
Form query	5.11	2.86	6.67	4.43	4	3.29	4	2.43
Revise query	8	6	10.78	7.7	6.89	6	5.67	2.86
Execute implementation actions	15.55	9.71	19.78	13	12.78	10	13	5.71
Evaluate initial results	12.56	8.86	17.22	11.85	11.44	9.14	9.44	5.43
Evaluate results content	5.11	4.71	7.44	7.14	3.67	3.86	4.67	2.71
Evaluate and diagnose search success/ failure	2.11	2.71	3.89	2.85	2.78	2.71	3.33	1.57
Average total activities observed	68.76	47.14	94.9	60.25	63.12	48.71	57.78	29.85

Table 5.1: Average frequency of activity types for device novice/ experts for each retrieval session. Averages are used to enable inter-group comparisons. The categories of activity in which more substantial differences between experts and novices occurred are shaded.

The tentative observations on the raw data were investigated for statistical significance (see table 5.4). The analysis tests for significant differences between novices and experts on each of the experimental tasks using unrelated t tests dividing the data based on experimental task, user knowledge (novice and expert) and behaviour category.

Category	Tasks							
	OG1		OG2		PH1		PH2	
	t	p<	t	p<	t	p<	t	p<
Identify concept	2.10	0.05	3.60	0.005	2.57	0.025	1.25	
Identify relation	0		0.52		0.03		0.37	
Form strategy								
Revise strategy	2.12	0.05	2.39	0.025	1.46		2.02	0.05
Reason about problem space	2.06	0.05	4.02	0.005	3.00	0.005	2.91	0.01
Explore synonyms	1.13		2.87	0.01	2.68	0.01	2.19	0.025
Select terms	0.22		1.65		0.49		1.66	
Form query	1.18		0.84		0.39		1.31	
Revise query	0.80		1.20		0.43		1.60	
Execute implementation actions	1.06		1.32		0.80		1.72	
Evaluate initial results	0.49		1.16		0.72		1.69	
Evaluate results content	0.20		0.24		0.16		1.71	
Evaluate and diagnose search success/ failure	0.87		1.10		0.08		2.69	0.01

Table 5.2: Significant differences between the frequencies of categories of activities for device novice/ expert df = 14. The direction of the differences is apparent in table 5.1

The data indicates significant differences between expert and novices exist in the number of concepts identified and this difference was not expected, but it occurred on three of the four tasks. The results in table 5.2 show that significant differences occur with experts performing more problem reasoning and development based activity during a retrieval session than novices (i.e. revise strategy, reason about the problem and synonym exploration) and in general these differences were common across tasks (except for synonym exploration on OG1 and strategy revision on PH1). The statistical significant of the inter-group differences identified in the activity observed (Table 5.1) indicate different approaches to retrieval by experts and novices; and are not attributable to subjects with extreme activity distributions. Task PH2 shows that in addition to the significant differences found for the other tasks the number of times search diagnosis is observed is significantly different between novice and experts; as experts produced double the number of activities performed by novices in this category activity on this task.

The configuration of the activities were analysed by determining what percentage distribution of all activities performed a particular category accounted for. A difference in the distribution of expert and novice behaviour would indicate a strategic difference in the approach to retrieval.

Category	OG1		OG2		PH1		PH2	
	Expert	Novice	Expert	Novice	Expert	Novice	Expert	Novice
Identify concept	4.68	5.45	8.08	6.87	6.69	7.04	5.76	9.08
Identify relation	1.45	2.12	2.7	3.55	2.47	3.22	1.54	3.35
Form strategy	1.45	1.82	1.05	1.66	1.58	2.05	1.73	3.35
Revise strategy	6.14	3.94	5.15	4.03	6.33	4.99	6.16	5.73
Reason about problem space	4.52	1.82	4.09	1.17	5.45	1.77	5.57	2.88
Explore synonyms	7.91	5.45	6.43	2.37	7.4	4.39	6.35	2.88
Select terms	3.39	5.45	3.16	2.37	4.23	4.68	3.46	3.35
Form query	7.43	6.07	7.03	7.35	6.33	6.75	6.92	8.14
Revise query	11.63	12.7	11.36	12.78	10.92	12.31	9.81	9.58
Execute implementation actions	22.61	20.6	20.84	21.58	20.25	20.53	22.5	19.13
Evaluate initial results	18.27	18.8	18.15	19.67	18.12	18.76	16.34	18.19
Evaluate results content	7.43	9.99	7.84	11.85	5.81	7.53	8.08	9.08
Evaluate and diagnose search success/ failure	3.07	5.75	4.1	4.73	4.4	5.56	5.76	5.26

Table 5.3 Configuration of activities by the device knowledge and task as a % of all activities recorded

Table 5.3 shows large expert-novice differences in the percentage of activity frequency across tasks for: reason about the problem, synonym exploration and article evaluation. Experts appear to reason about the problem and explore synonyms proportionately more than novices; whereas novices spend a greater proportion of their activities evaluating article contents. This indicates that novice strategies focus on evaluation and experts more on problem exploration and definition of search directions pre-search. These tentative conclusions are tested for statistical significance between the groups of searchers in table 5.4.

Unrelated t tests are used to investigate the significance of strategic differences between searchers with differing device knowledge. This analysis is performed for each of the tasks.

Category	Tasks							
	OG1		OG2		PH1		PH2	
	t	p<	t	p<	t	p<	t	p<
Identify concept	0.40		1.05		0.65		1.88	0.05
Identify relation	0.22		0.55		1.07		1.42	
Form strategy	0.75		3.22	0.005	1.38		1.98	0.05
Revise strategy	2.83	0.01	0.97		1.48		0.29	
Reason about problem space	3.06	0.005	3.00	0.005	3.33	0.005	2.44	0.025
Explore synonyms	2.49	0.025	2.73	0.01	3.20	0.005	1.18	
Select terms	0.54		1.00		0.42		0.72	
Form query	0.81		0.47		0.46		0.05	
Revise query	1.46		1.45		0.49		0.29	
Execute implementation actions	0.80		0.93		0.10		0.80	
Evaluate initial results	0.55		1.56		0.17		1.40	
Evaluate results content	2.86	0.01	2.20	0.025	1.26		0.45	
Evaluate and diagnose search success/ failure	2.47	0.025	0.34		0.96		0.58	

Table 5.4: Significant differences between the different tasks performed on each of the categories of activities for high and low device knowledge populations for the proportion of the total activities a specific activity accounts for in a retrieval session(Unrelated t test, 2 dp df = 14)

Table 5.4 shows that some differences in the configuration of expert-novice activities are significant. The differences support the different focuses for search activity; with experts favouring reasoning and exploration of synonyms and novices favouring evaluation activity (except on the PH tasks). PH2 also shows differences with novices spending a greater proportion of the total activities performed identifying concepts than experts. Significant differences in the configuration of the activities observed indicate that different strategies for information retrieval exist based on device knowledge.

To conclude this section data for the rate of activity (number of time a category activity occurs per minute) for novices and experts are analysed. Rate of activity occurrence is an indicator of skill for all but evaluation based activity.

Category	OG1		OG2		PH1		PH2	
	Expert	Novice	Expert	Novice	Expert	Novice	Expert	Novice
Identify concept	0.26	0.20	0.34	0.19	0.26	0.21	0.24	0.25
Identify relation	0.08	0.08	0.11	0.10	0.10	0.10	0.06	0.09
Form strategy	0.08	0.07	0.04	0.05	0.06	0.06	0.07	0.09
Revise strategy	0.34	0.14	0.21	0.11	0.25	0.15	0.25	0.16
Reason about problem space	0.25	0.07	0.17	0.03	0.21	0.05	0.23	0.08
Explore synonyms	0.43	0.20	0.27	0.06	0.29	0.13	0.26	0.08
Select terms	0.19	0.20	0.13	0.06	0.16	0.14	0.14	0.09
Form query	0.41	0.22	0.29	0.20	0.25	0.20	0.28	0.22
Revise query	0.64	0.47	0.47	0.35	0.42	0.37	0.40	0.26
Execute implementation actions	1.24	0.76	0.87	0.59	0.79	0.62	0.93	0.52
Evaluate initial results	1.00	0.69	0.75	0.54	0.70	0.56	0.67	0.50
Evaluate results content	0.41	0.37	0.33	0.32	0.23	0.24	0.33	0.25
Evaluate and diagnose search success/ failure	0.17	0.21	0.17	0.13	0.17	0.17	0.24	0.14
Total activity per minute	5.5	3.68	4.15	2.73	3.89	3	4.1	2.73

Table 5.5: Average rate per minute of activity occurrence by device knowledge and task

Table 5.5 shows a large difference in the total number of activities performed per minute for novices and experts; with experts performing substantially more activities per minute than novices. This is to be expected due to the effect of the users' device knowledge on activity completion. Rassmussen's (1983) view of the experts using pre-compiled plans of action and novices using knowledge based reasoning in determining interaction is demonstrated by the differences in rate data; as if an activity is pre-compiled it is likely to be completed quicker than if interactions have to be evaluated against the state of the interface. The data reveals a stark contrast for novice and experts for the number of times the following occurred per minute: strategy revision, reasoning about the problem, synonym exploration, formulation of new queries, query revision, device execution and the evaluation of the initial results.

Category	Tasks							
	OG1		OG2		PH1		PH2	
	t	p<	t	p<	t	p<	t	p<
Identify concept	0.91		1.52		1.44		1.01	
Identify relation	0.01		0.05		0.20		0.74	
Form strategy	0.87		1.95	0.05	0.15		1.33	
Revise strategy	3.07	0.005	1.97	0.05	1.77	0.05	0.34	
Reason about problem space	2.30	0.05	3.18	0.005	4.10	0.005	2.58	0.025
Explore synonyms	0.67		2.92	0.01	3.50	0.005	2.07	0.05
Select terms	0.01		1.22		1.07		1.11	
Form query	1.48		0.25		0.62		0.37	
Revise query	0.58		0.03		0.98		0.47	
Execute implementation actions	1.20		0.48		1.22		0.14	
Evaluate initial results	0.90		0.14		1.27		0.53	
Evaluate results content	0.62		1.84	0.05	0.18		0.53	
Evaluate and diagnose search success/ failure	0.05		0.23		0.54		0.07	

Table 5.6: Shows the significant differences between the different tasks performed on each of the categories of activities for high and low device knowledge populations for the rate of category occurrence in a retrieval session (Unrelated t tests 2 d.p df = 14)

Significant differences occur in the rate of behaviour between novices and experts for: strategy revision (except PH2), reasoning about the problem and exploration of synonyms (except OG1); see table 5.6. It is striking that these activities develop the information need and search approach as a pre-search activity. Significant differences in the rate of activity are expected because of the proceduralisation of knowledge as skill and thus increased speed with which activity completion as experts do not have to perform knowledge based reasoning.

5.4.1.2 Summary of the effects of user device knowledge on the activity performed in a retrieval session

- Experts perform more strategy revisions (OG1 $p< 0.05$, OG2 $p<0.025$, PH2 $p<0.05$), reason about the problem space more (OG1 $p<0.05$, OG2 $p<0.005$, PH1 $p<0.005$, PH2 $p<0.01$) and explore more synonyms (OG2 $p<0.01$, PH1 $p<0.01$, PH2 $p< 0.025$) than novices irrespective of the task.
- Experts appear to form more new queries (>50% more across tasks) and reformulate more old queries (>35% more across tasks) than novices.
- Most of the across task differences found in the activity of a subject group occur for device execution and the interpretation of results.
- Expert strategies focus on the exploration of problems; whereas novice strategies focus on the evaluation of the content of articles. These differences are significant for tasks OG1 and OG2. On PH1 and PH2 the percentage of activities performed used to evaluate articles was similar for expert and novices but problem manipulation and exploration differences are still significant (experts > novices).
- The rate of activity for experts is substantially greater than novices. So either activities are completed quicker indicating the effects of skill on behaviour or novices' behaviour is dominated by categories of activity which can take longer to complete (e.g. evaluating a large number of articles).
- Experts' rate of strategy revision, reasoning about the problem space, synonym exploration, new query formulation and query revision is much greater than novices. The differences in strategy revision (except on PH2), reasoning and synonym exploration (except on OG1) are significant ($p<0.05$).

5.4.2 Analysis of sequences of mental behaviours

The analysis describes the activity observed as state transition diagrams for the frequency of category to category transitions. The analysis uses transitional probability to further describe the observed activity. The transitional probability for a given state transition pair (category A to category B transition) is the likelihood that a particular state (category B) will follow the occurrence of another state (category A) given that this state (category A) has already occurred. Binomial test Z distribution is used to determine if a particular sequence of activity (category A to category B transition) occurs significantly more than expected (Bakeman and Gottman 1986).

5.4.2.1 Transitional frequencies for category to category transitions

The frequency of each, possible category to category transition are presented in this section. The analysis separates the data by task and subjects device knowledge. The frequency transition data observed for individual subjects can be found in appendix 5a. To enable inter-subject group comparisons adjustments were made to the novice data (experts 9 subjects, novices 7 subjects). The diagrams also include the transitional probabilities of a category to category transition. The transitional probabilities discussed in this section are derived using the following formulae for the conditional probability of a particular category occurring given a particular category of behaviour has already occurred (Bakeman and Gottman 1986).

$$p(category2/category1) = frequency(category1,category2)/frequency(category1)$$

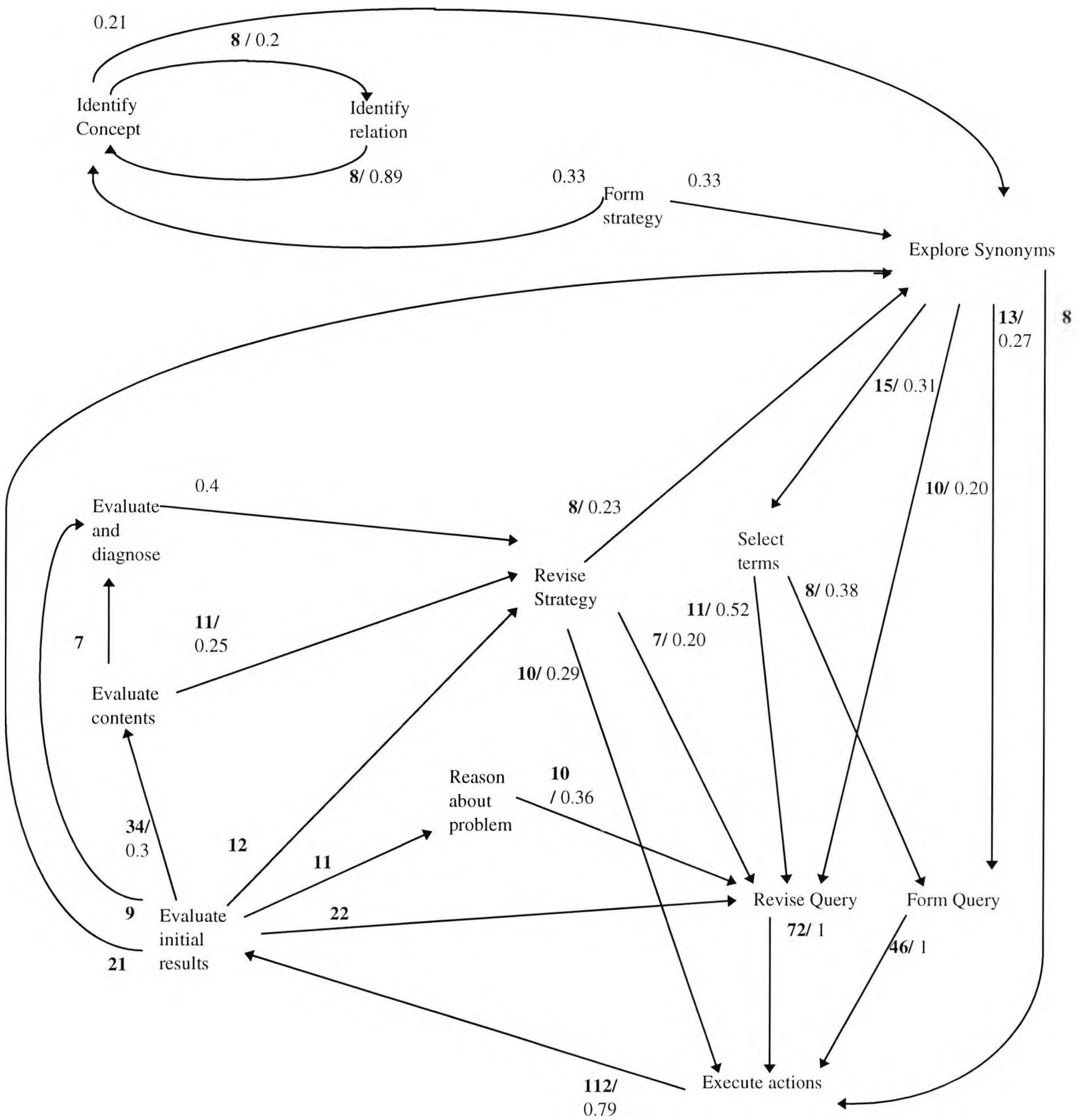

Figure 5.1: Transitional frequencies for experts on task OG1. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.1 shows the more frequent category transitions for experts on task OG1. The figure shows the expected cycle of identifying concepts and associated relationships. The figure shows an interesting transition cycle between query revision, execution and evaluation (evaluate initial results, evaluate article content and diagnose search success). Strategy revision may be triggered by judgements about the number of articles retrieved or the content of the articles. The outcome of strategy revision is either the reformulation of the current query, synonym exploration or device execution. The figure indicates experts make query and strategy refinement decisions based on information gained from the different levels of evaluation. The figure shows some interesting transitions from the evaluation of the size of the results; as experts do not necessarily proceed to article evaluation. The figure highlights a common transition from the searcher reasoning about the problem and revising the current query.

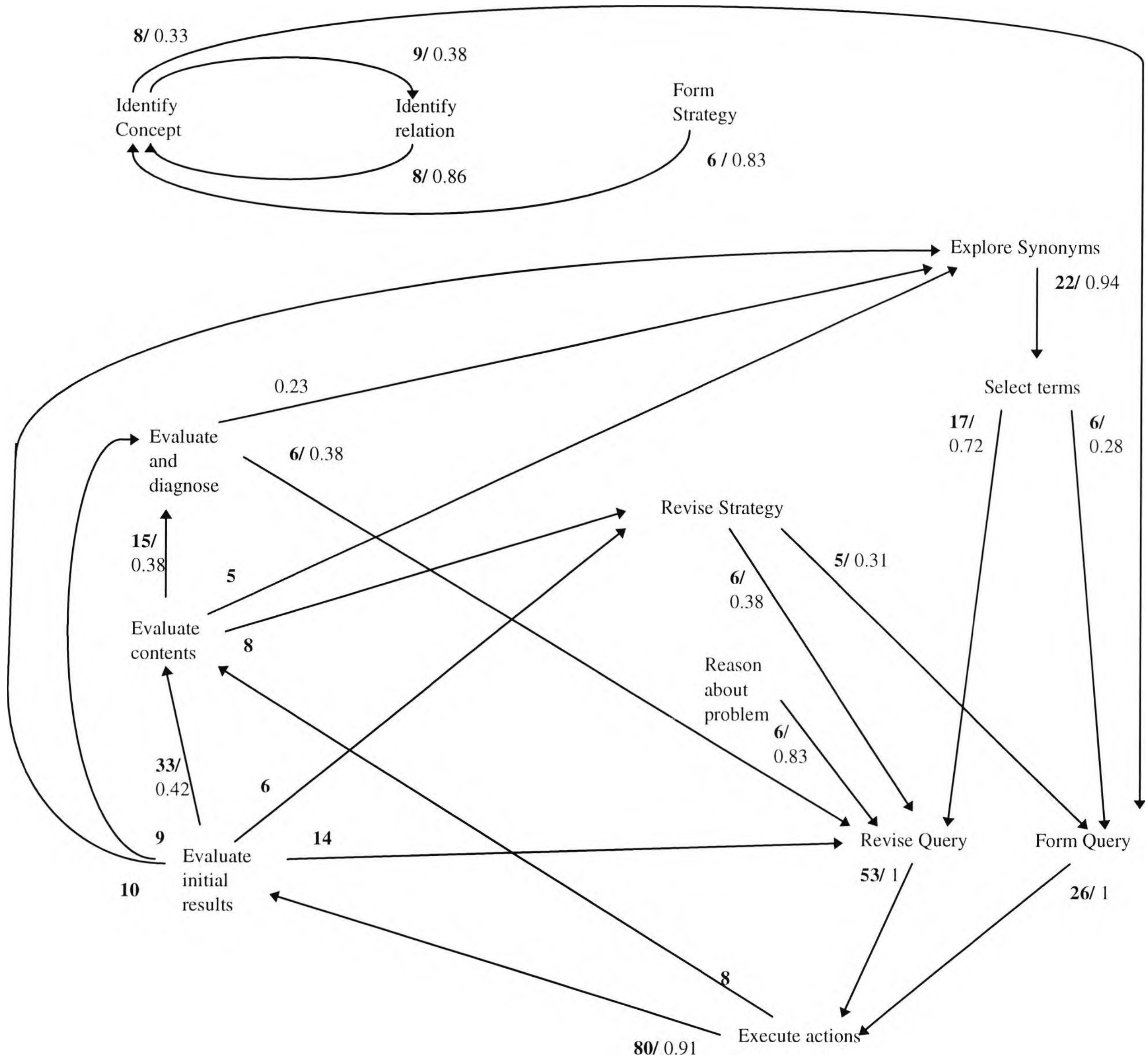

Figure: 5.2 Transitional frequencies for novices on task OG1. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.2 shows the most common category transitions for novice subjects on task OG1. The figure highlights the same cycle of identifying concept and relationships found in experts behaviour. The figure also indicates a frequent transition between strategy formulation and concept identification. The figure highlights three main outcomes for the evaluation phase of searching. The subjects either reformulate the current query, revise their strategy or explore synonyms. Strategy revision results in querying activity and this differs from experts, as they may also reason about the problem. It is apparent from a comparison of figures 5.1 and 5.2 that experts produce a greater variety in the sequences of behaviour occurring above the threshold of 1% of all transitions. This may indicate a richer set of IR strategies are being employed.

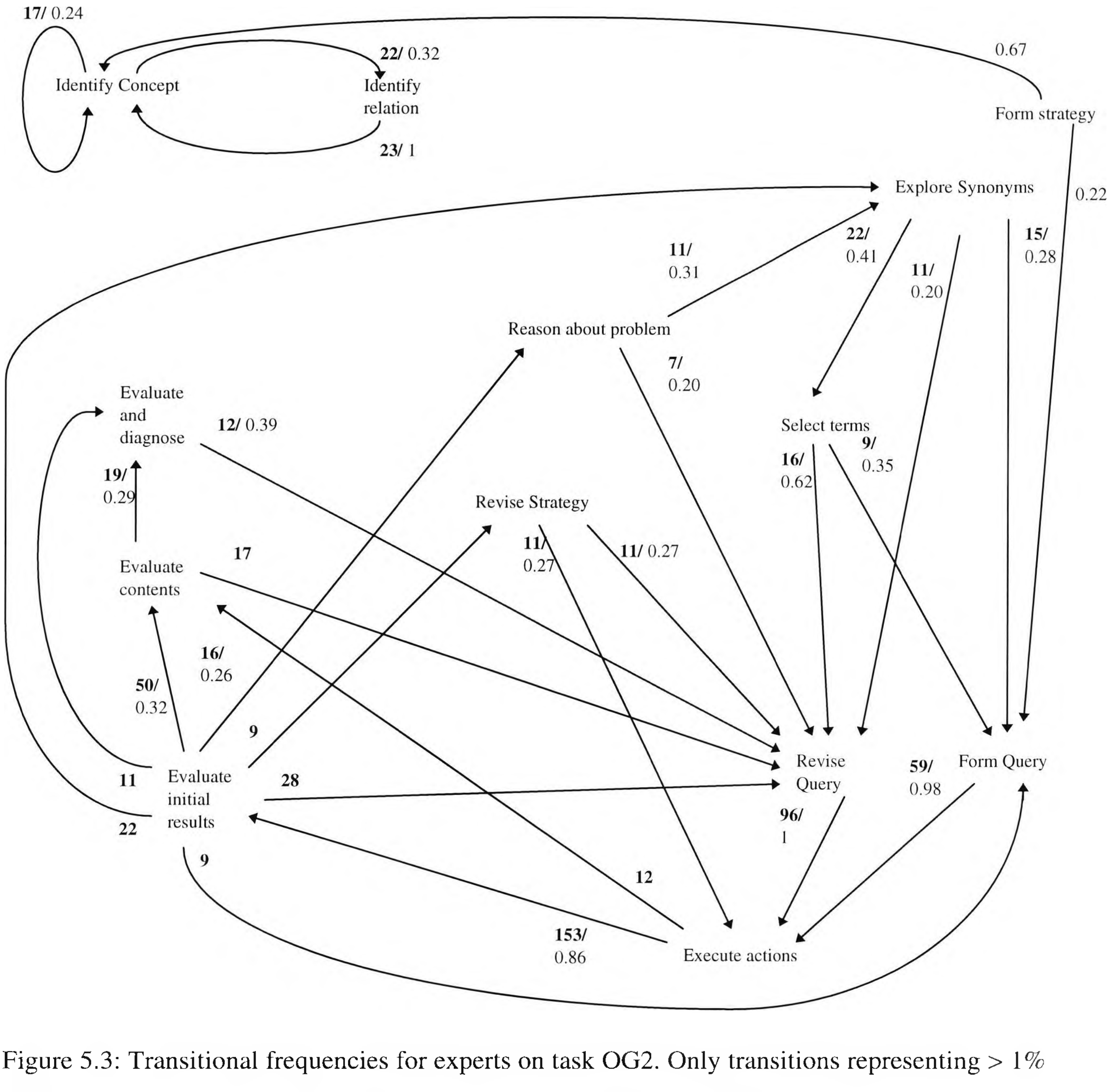

Figure 5.3: Transitional frequencies for experts on task OG2. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.3 shows the most frequent category transitions for experts on task OG2. The figure highlights the same identify concept to identify relationship cycle found on task OG1, but it also shows the iterative identification of concepts. The evaluate-query reformulation cycle observed on task OG1 is also present on this task from each of levels of search evaluation. The figure shows that reasoning about the problem can lead to either the exploration of synonyms or the reformulation of the current query. Strategy revision leads to device execution or query reformulation. On this task there appears to be a sequence of behaviour of producing a query, evaluate the number of results produced, adjusting the strategy (or reason about the problem) and alter the query.

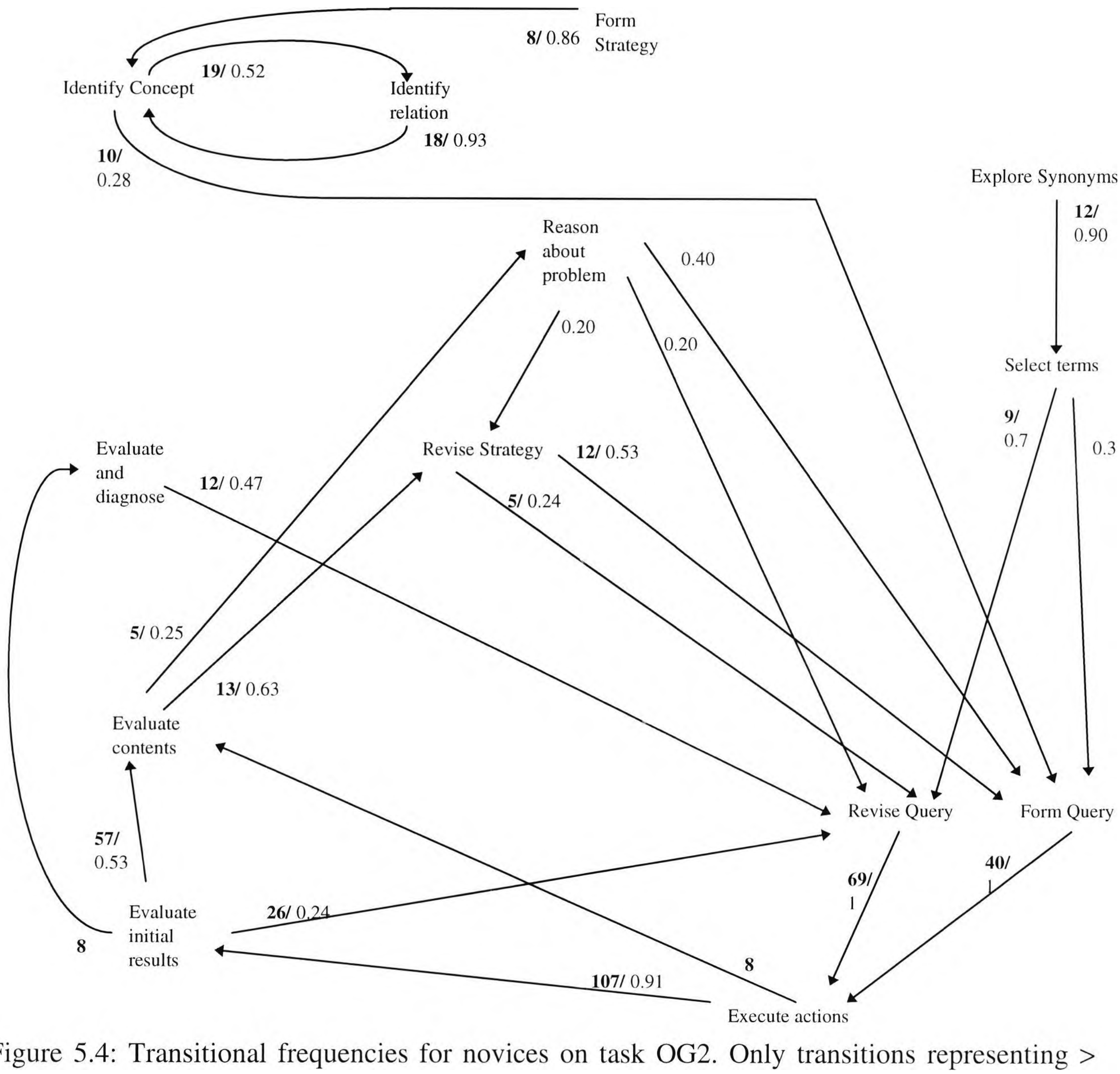

Figure 5.4: Transitional frequencies for novices on task OG2. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.4 shows the most common category transitions observed from the interaction and protocols of novices on task OG2. The figure shows the cycle of identifying concepts and relationships common to the other tasks. Interestingly the figure also shows a common transition between the identification of concepts and the formulation of new queries which is not observed in novice behaviour on task OG1. This may be caused by subjects trying to 'get a feel' number of articles associated for the many aspects of the need before determining the best approach to retrieval. The figure highlights the expected evaluate-query reformulate cycle found on the other tasks; but it also shows that novices make reformulation decisions either based on the number of articles produced or diagnosis of search success. Interestingly, article evaluation leads to either strategy revision or reasoning about the problem, but not directly to querying. However, strategy revision is followed by either the formulation of a new query or the reformulation of the current query; so strategy adjustments appear to focus activity on altering queries rather than leading to activity whose focus is the development of a richer need description using domain knowledge (explore synonyms, problem reasoning). It is notable that on this task novices follow strategy revision with the formulation of a new query much more frequently than query reformulation; whereas on task OG1 transitions from strategy revision to formulation or reformulation of queries were approximately equal. This may occur owing to the larger number of qualifying conditions present in this task whereas task OG1 had only one qualifying condition (socio-economic).

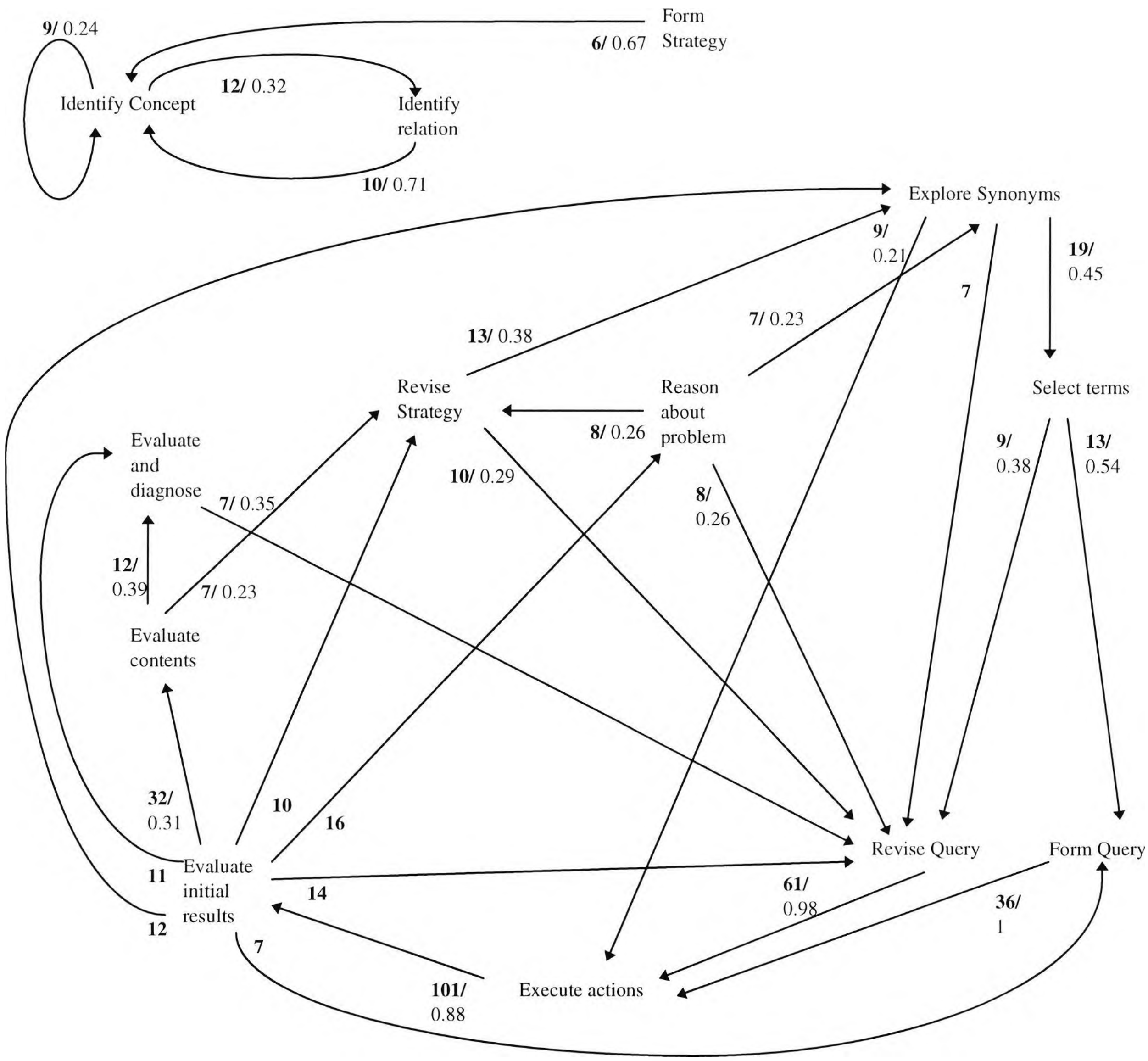

Figure 5.5: Transitional frequencies for experts on task PH1. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.5 shows the most common category transitions produced by experts on task PH1. Cycles between the identification of concepts and the identification of relationships are observed as in the other tasks. The subjects also iteratively verbalised a list of the concepts identified as on OG2. The expected evaluation-query reformulation cycle is shown, with the searcher adjusting the query based on the results retrieved. Interestingly the evaluation of article contents may lead to the revision of the current strategy, possibly in the light of the characteristics found in the results set (as on OG1). The revision of the current strategy tends to be followed by the reformulation of the current query or exploration of synonyms. Reasoning about the problem leads to the exploration of synonyms, strategy revision or the reformulation of the current query.

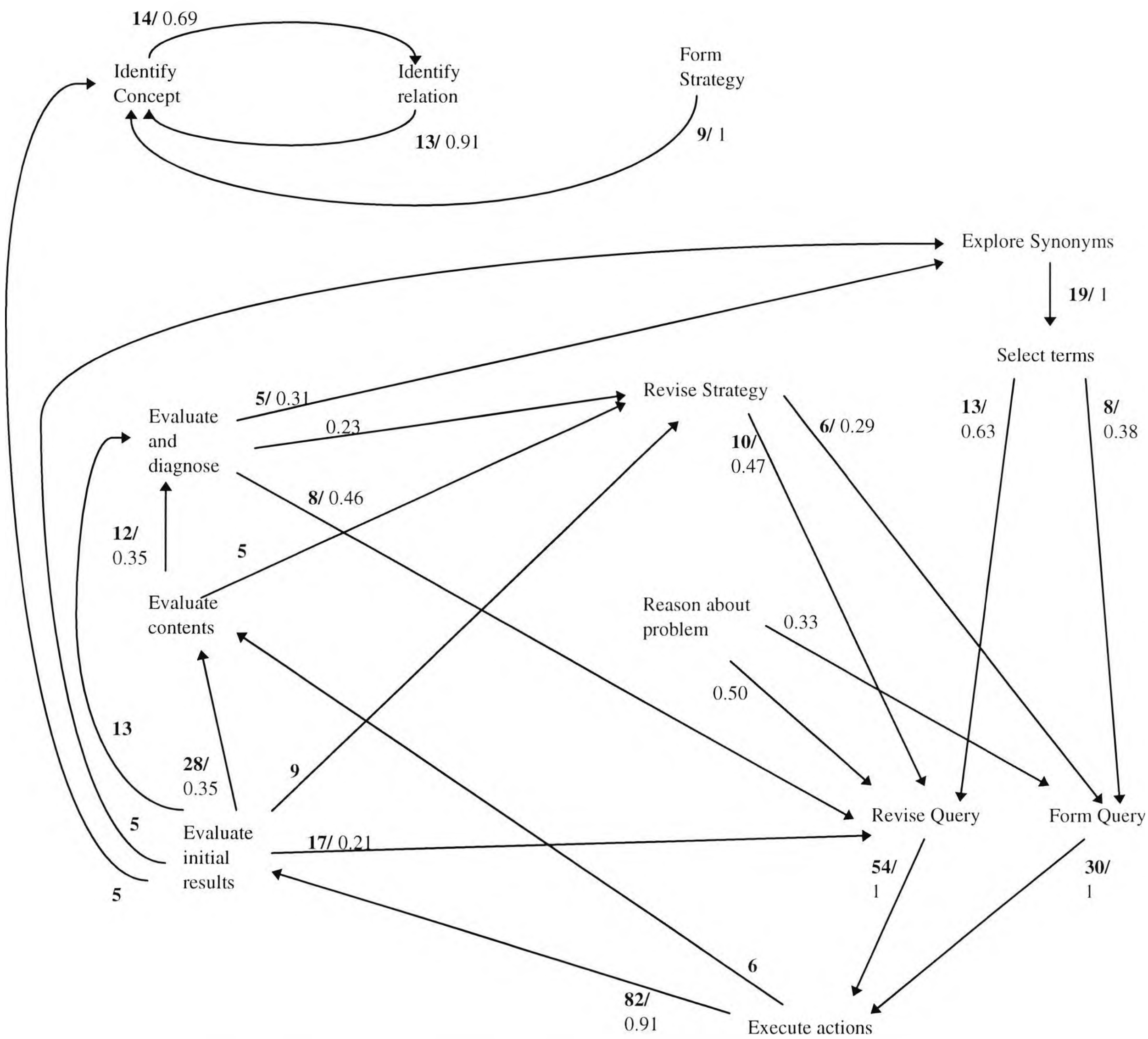

Figure 5.6: Transitional frequencies for novices on task PH1. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.6 shows the most frequent category transitions for novice subjects on task PH1. The cycles of identifying new concepts and associated relationships is the same as in the other tasks. Search evaluation and diagnosis leads to the reformulation of the current query or synonym elaboration. The revision of the current strategy tends to lead to querying activity; either the reformulation of the current query or the generation of a new query, as does reasoning about the problem.

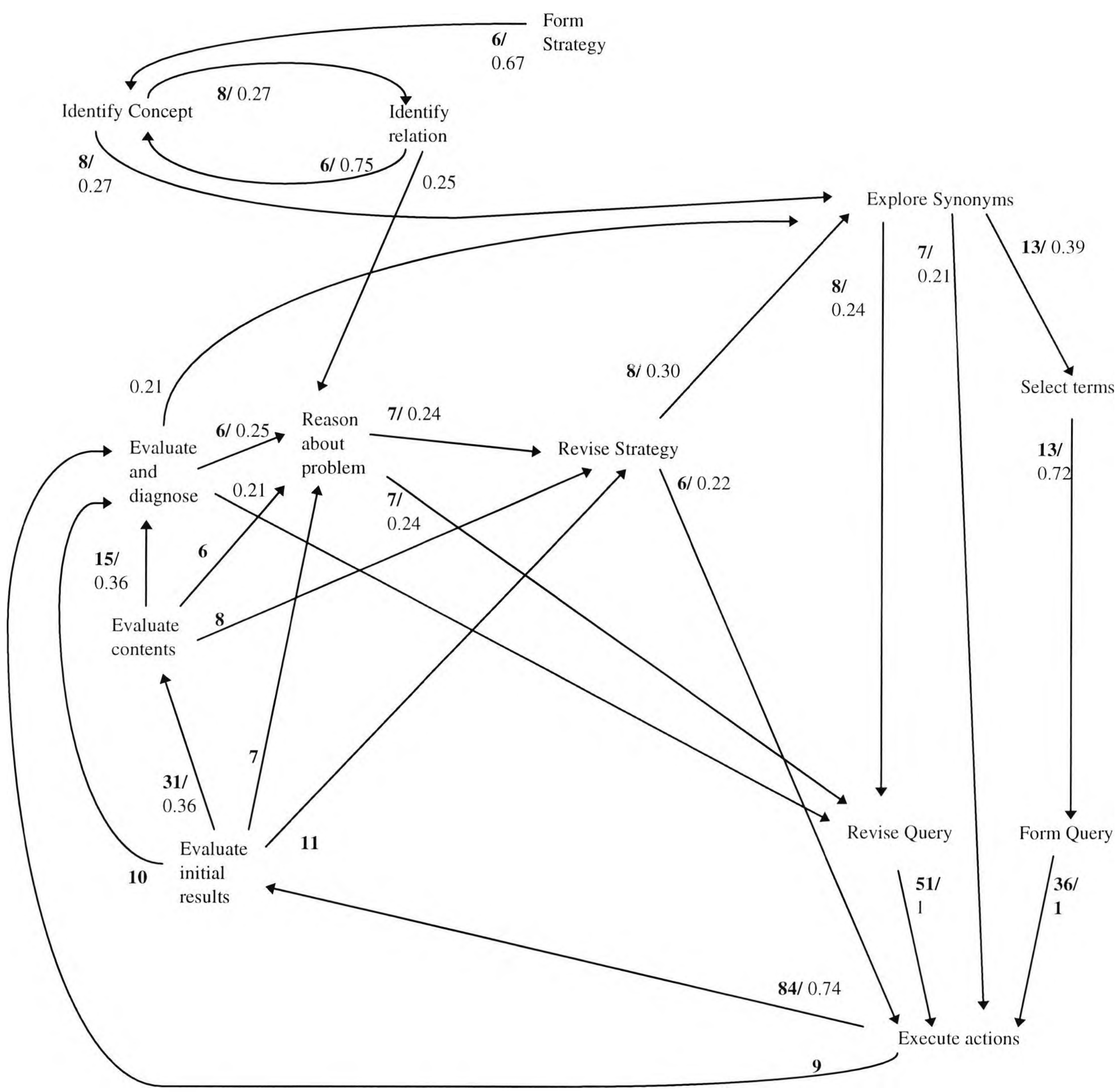

Figure 5.7: Transitional frequencies for experts on task PH2. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.7 shows the transitions observed from the behaviour of experts on task PH2. The figure shows the cycle between the identification of new concepts and relationships. The figure shows the evaluation-query reformulation cycle observed in the other tasks. Search diagnosis leads to reasoning about the problem and this in turn leads to query reformulation.

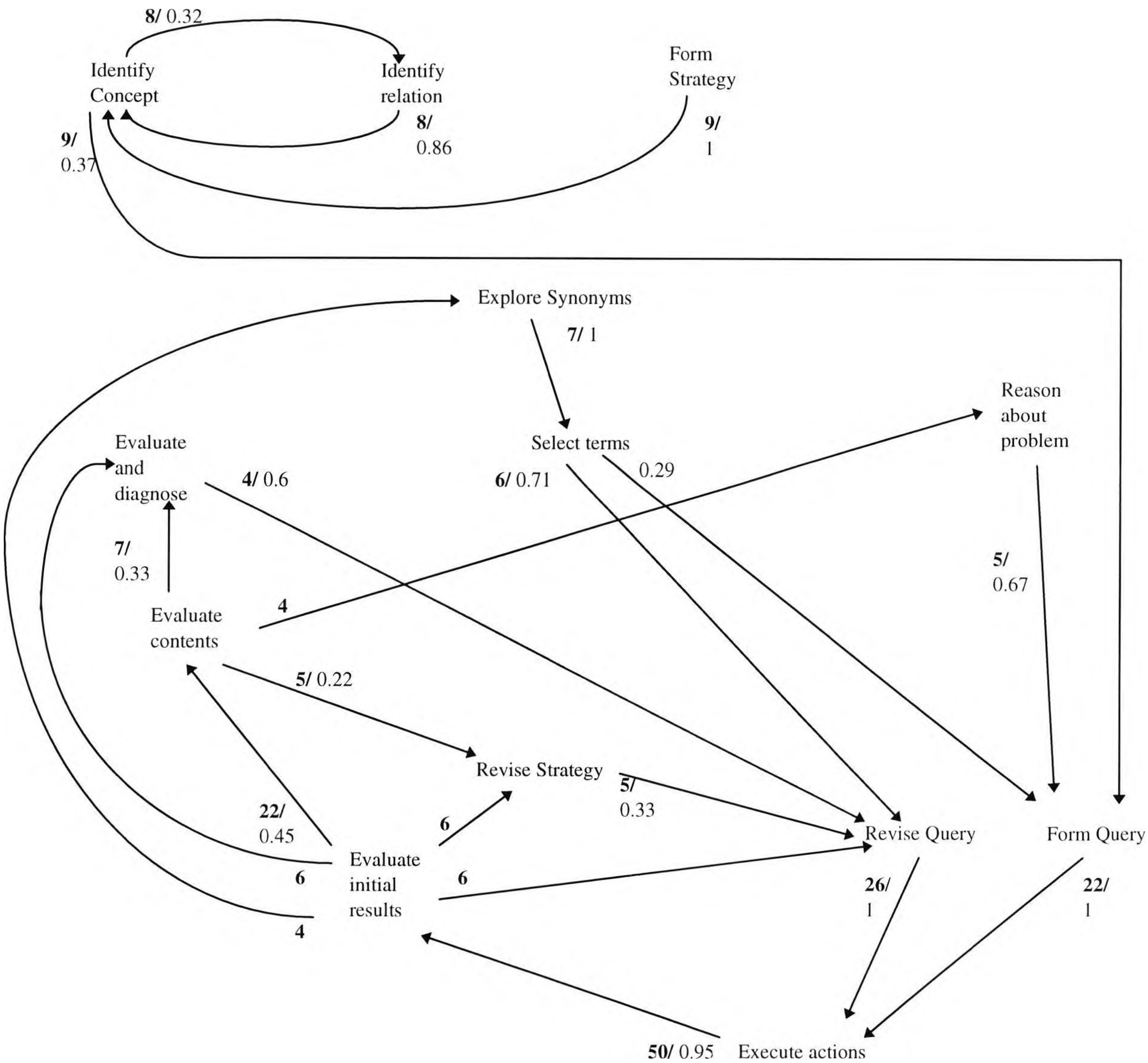

Figure 5.8: Transitional frequencies for novices on task PH2. Only transitions representing > 1% of all transitions (indicated in bold) or transitional probabilities > 20% are shown.

Figure 5.8 shows the transitional frequencies for novice subjects on task PH2. The cycle of identifying concepts and relationships is the same as on the other tasks. Interestingly the identification of concepts can also be followed by the formulation of a new query (as in task OG2) so it appears this transition is only observed frequently on the more complex tasks. Interestingly, the number of different transitions between evaluation centred activity and reformulation activity is greater on this task than on others. The subjects follow search diagnosis with querying based activity. The evaluation of the contents of articles is followed by either strategy revision or problem reasoning.

Summary

- Both subject groups produced the expected cycle of identifying concepts and relationships on all tasks.
- Experts frequently explore synonyms and revise the current query after strategy revision on the simpler tasks (OG1 and PH1), whereas on OG2 only query revision is pursued frequently and on PH2 subjects either explore synonyms or execute the device. Device execution after strategy revision indicates subjects are returning to the results of previous queries later in the retrieval session and so highlights a back tracking in their behaviour.
- Experts revise the current query frequently after reasoning about the problem on all tasks
- Experts follow the evaluation of the size of the results set with the exploration of synonyms and query reformulation on three of the tasks (OG1, OG2 and PH1) and they follow evaluation at this level with strategy revision and reasoning about the problem frequently on all tasks.
- Experts follow the evaluation of the content of articles with strategy revision frequently on three of the tasks (OG1, PH1 and PH2) whereas on task OG2 evaluation frequently leads directly to query reformulation.
- Novices follow the identification of a concept with the formulation of a new query relatively frequently only on the more complex tasks (OG2, PH2).
- For novices strategy revision tends only to lead to querying activity on all tasks.
- Novices follow the evaluation of the size of the results set with the exploration of synonyms on three of the tasks (OG1, PH1 and PH2) and they follow evaluation at this level with query reformulation frequently on all tasks.
- Novices follow the evaluation of the article contents with strategy revision frequently on all tasks (OG1, PH1 and PH2) and a frequent transition from article evaluation to reasoning about the problem is observed on the more complex tasks (OG2, PH2).

5.4.2.4 Statistical techniques for pattern comparison

The statistical analysis of behaviour sequences of behaviour used the Binomial Z distribution following Bakeman and Gottman (1986) method of using the transitional probabilities to calculate Z scores using following formulae:

$$Z = \frac{p(category2/category1) - p(category2)}{\sqrt{\left(\frac{p(category2)*[1 - p(category2)]}{(\sum_{l}^{Number\ of\ categories} category\ occurrences)*p(category\ 1)}\right)}}$$

$$p(category2) = frequency(category2)/ \sum_{l}^{Number\ of\ categories} category\ occurrences$$

$$p(category1) = frequency(category1)/ \sum_{l}^{Number\ of\ categories} category\ occurrences$$

$$p(category2/category1) = frequency(category1,category2)/frequency(category1)$$

A first order model of expectation represents the probability of a category pair occurrence based on the number of category occurrences observed. Bakeman and Gottman's (1986) methodology states that a first order model is preferable to an equiprobable model as it requires fewer assumptions about the data and uses the simple probabilities of events in its predictions. Z scores greater than or equal to 1.96 are considered significant ($P < 0.05$).

5.4.2.2.1 Results

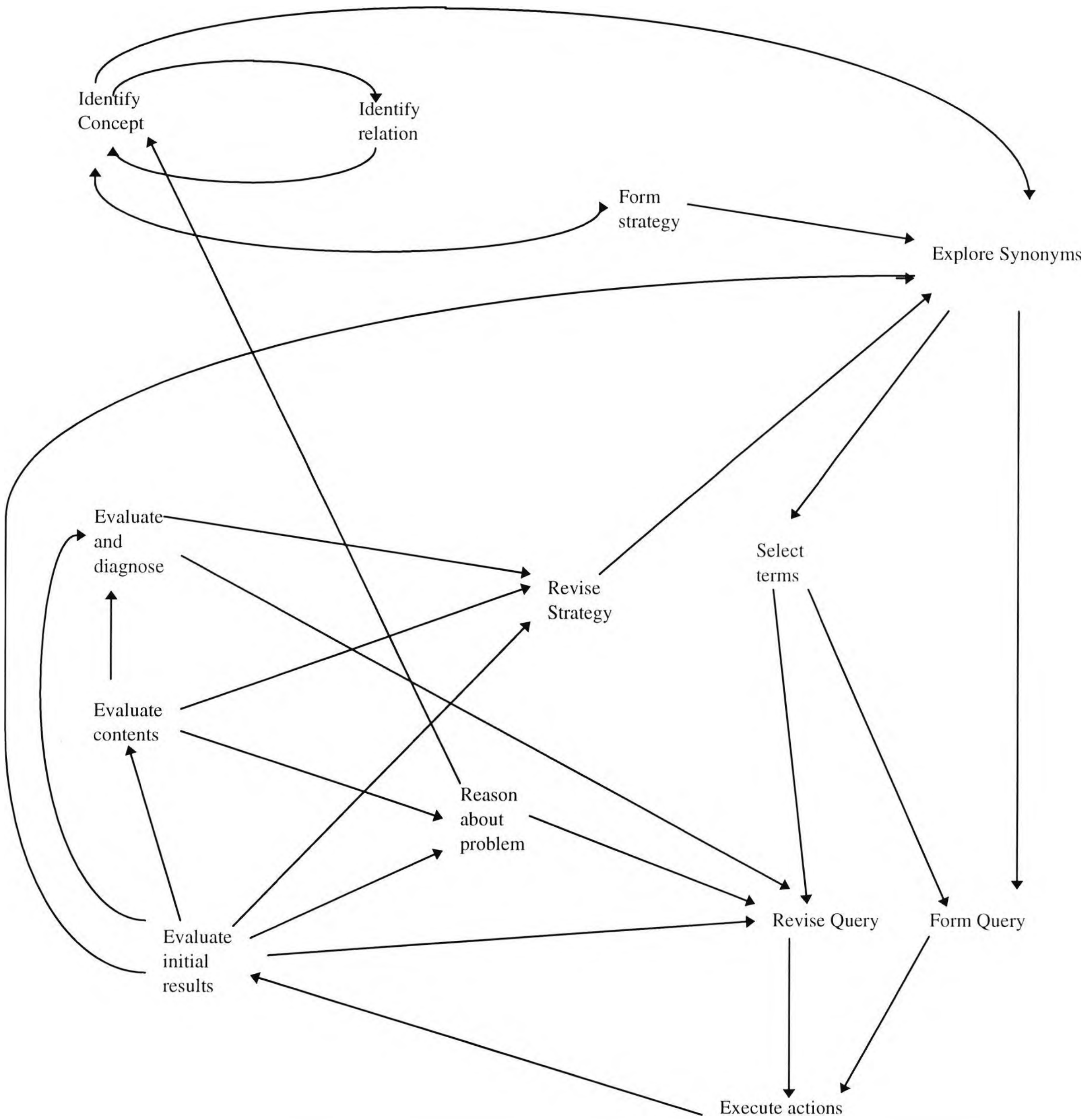

Figure 5.9. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of experts on task OG1. The significance transitions at $p < 0.05$ are shown. With a total of 27 significant sequences of behaviour.

Figure 5.9 shows the category pair sequences for experts occurring significantly more than expected for task OG1 based on a first order model of expected category pair sequence occurrence. The figure shows some of the cycles of activity to be significant sequences of behaviour (IC-IR, IR-IC; IC-FS, FS-IC). The figure indicates a cycle of producing search results (querying and device execution); evaluating those results and then adjusting behaviour based on these judgements.

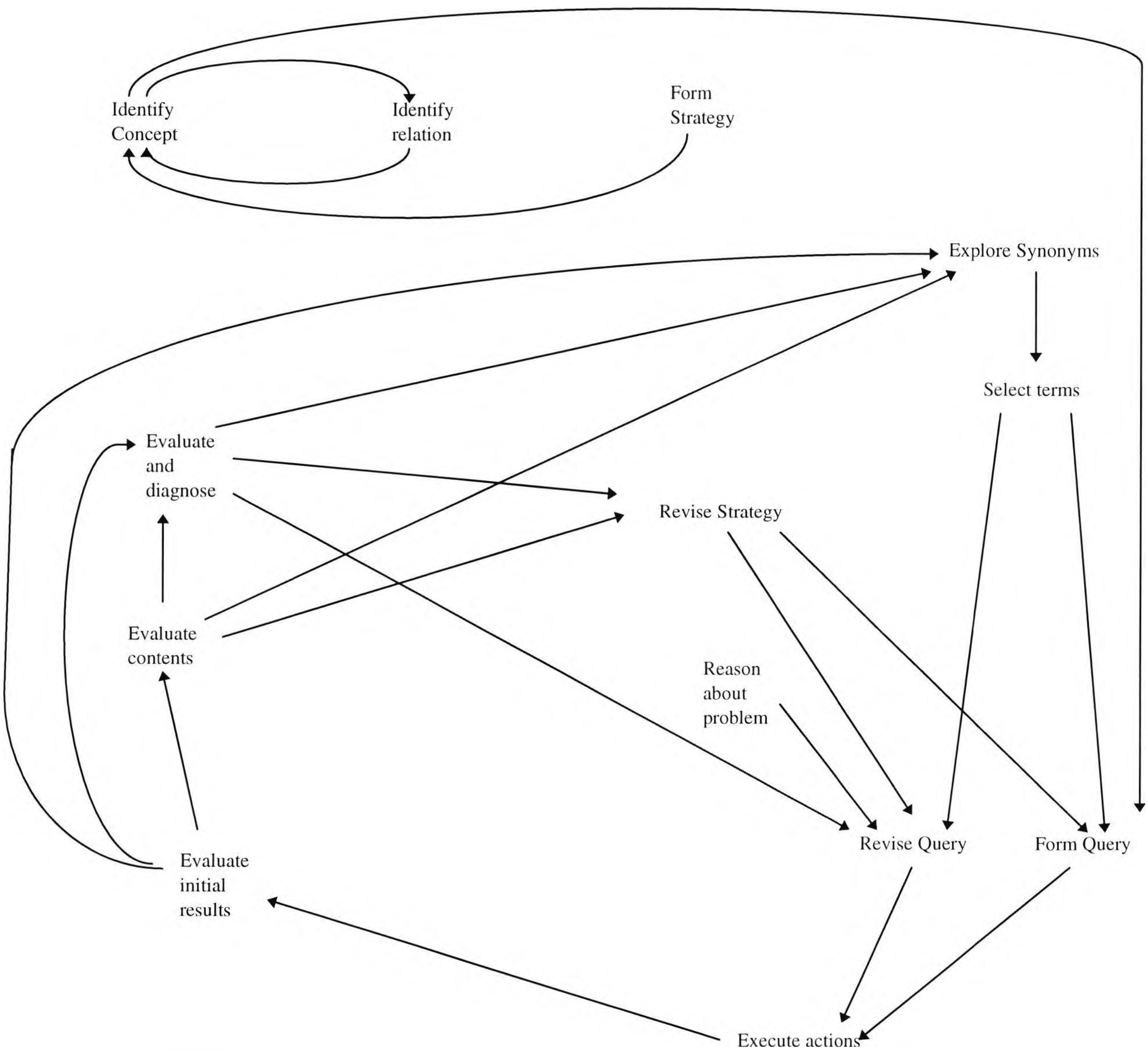

Figure: 5.10. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of novices on task OG1. The significance transitions at $p < 0.05$ are shown. With a total of 22 significant sequences of behaviour.

Figure 5.10 shows the significant transitions for novices. The figure highlights the expected cycle of identifying concepts and relationships, and indicates that novices focus less on pre-search planning and are more reactive; as adjustments to strategy or the exploration of synonyms only occur after the examination the searches results and not in the initial stages of searching (i.e. after concept or relationship identification; formulation of the initial strategy).

Both experts and novices use cycles of concept and relationship identification (IC-IR, IR-IC); but experts also follow a similar cycle between the identification of concepts and strategy formulation (IC-FS, FS-IC). The transitions from concept identification to synonym exploration is significant for experts but not for novices as they favour query formulation (see figures 5.9, 5.10). For experts the strategy formulation is followed by either the identification of new search concepts or the exploration of synonyms whereas novices only identify new search concepts. Interesting differences between experts and novices occur for the categories of behaviour following strategy revision and reasoning about the problem. Experts favour the identification of new search concepts or the exploration of synonyms; whereas novices proceed to querying (new query formulation or query reformulation). The other interesting differences in the behaviour sequencing of categories occur for the evaluation of results. Experts and novices both follow the evaluation of the results content and search success with strategy revision; in addition to this experts follow evaluation of the number of results with strategy revision whereas novices do not. Experts reason about the problem after evaluating the results (results set size or article contents) whereas novices do not. Interestingly novices proceed to explore synonyms based on the evaluation of the results set size, the content of articles or the success of the search; whereas experts only explore synonyms based on judgements made for the evaluation of the size of the results set. Experts and novices show similar sequences with the reformulation of queries following on from the evaluation of the search success, reasoning about the problem or term selection. In addition to these common inter-group sequences, experts also reformulate queries after the evaluation of the size of the results set; whereas novices reformulate queries after strategy revision. Experts and novices show the same progression through the sequence or results evaluation (ER-EC, ER-ED, EC-ED).

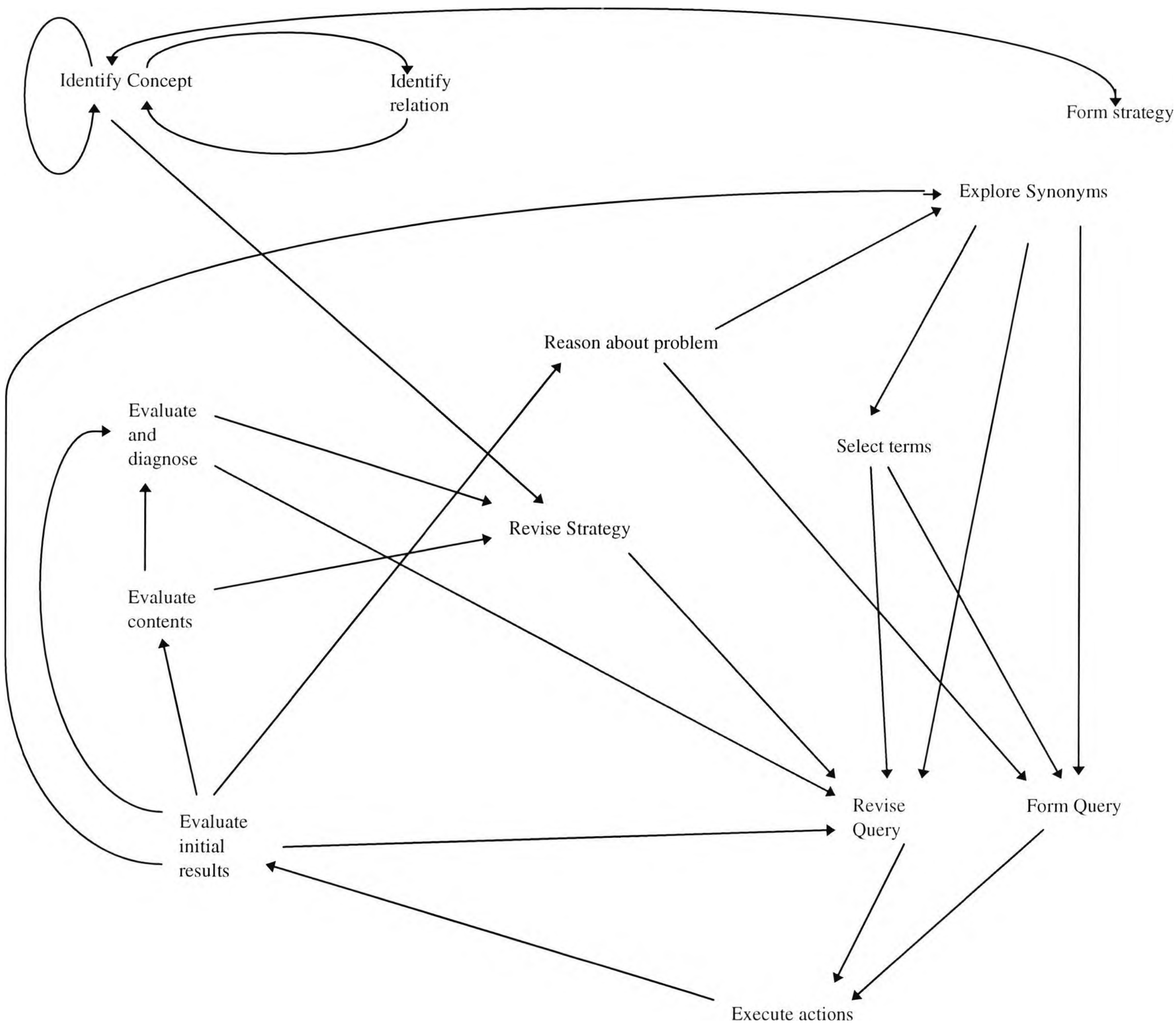

Figure: 5.11. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of experts on task OG2. The significance transitions at p < 0.05 are shown. With a total of 26 significant sequences of behaviour.

Figure 5.11 shows the behaviour sequences for experts for task OG2. As with task OG1 the figure shows some of the expected cycles of activity (IC-IR, IR-IC; IC-FS, FS-IC); but with iterative identification of concepts. This addition may occur due to the larger number of qualifying conditions for this task. The figure also indicates cycles of producing search results; evaluating results and then adjusting behaviour based on these judgements. Interesting differences between the sequence of behaviours produced on this task in comparison to those observed on OG1 are that on OG2 the transitions from the evaluation of the size of the results set to strategy revision, and article evaluation to reasoning about the problem are not significant sequences of behaviours as in task OG1.

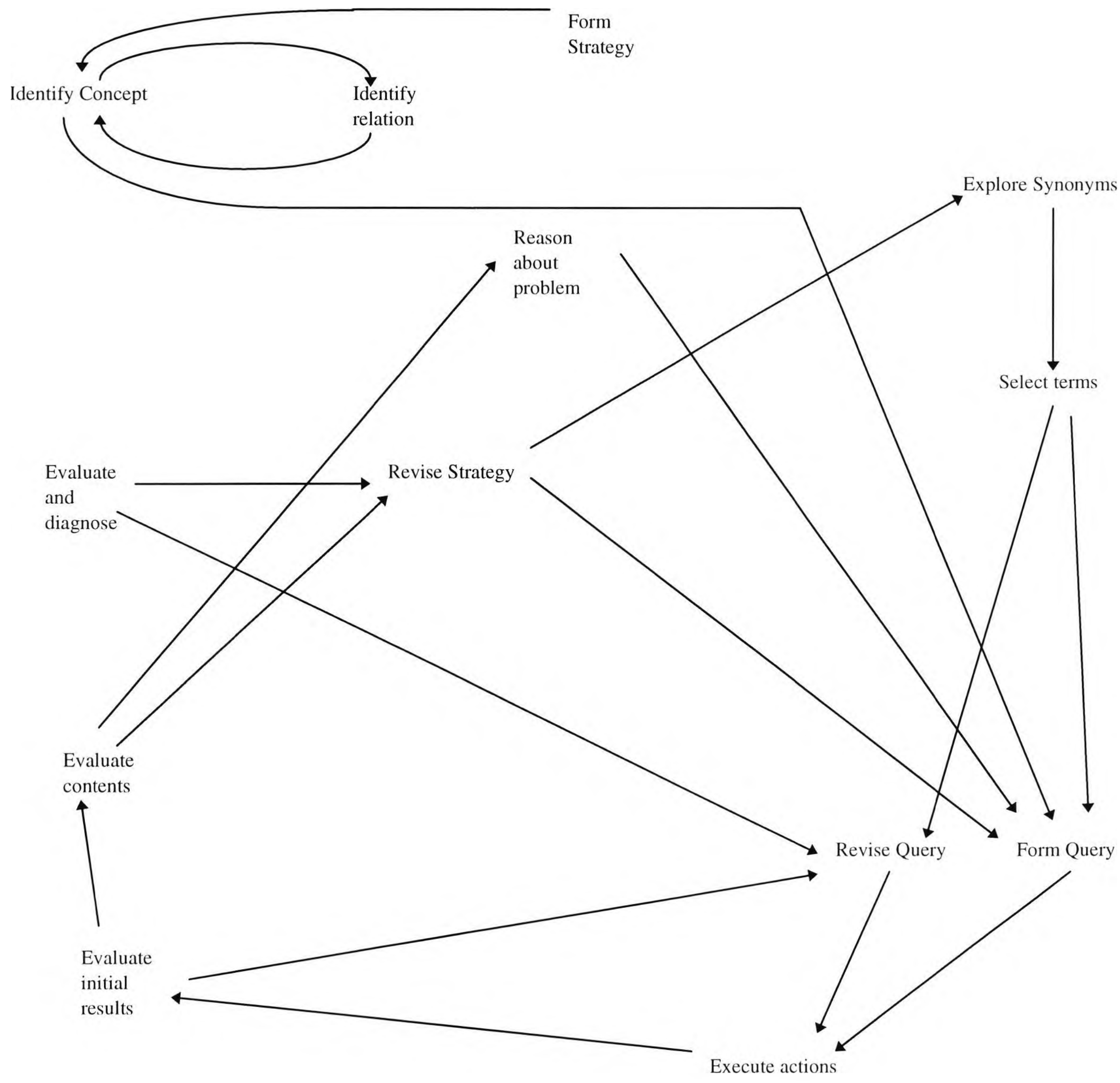

Figure: 5.12. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of novices on task OG2. The significance transitions at $p < 0.05$ are shown. With a total of 19 significant sequences of behaviour.

Figure 5.12 shows the behaviour sequences for novices for task. As in task OG1 it indicates a cycle of identifying concepts and relationships. The identify concept to query formulation transition is observed as in task OG1 which may indicate an investigative 'see how many articles are associated with a term' approach to searching. The table indicates that novices focus less on pre-search planning and are more reactive as adjustments to strategy or the exploration of synonyms only occur after examining the search results and not in the initial stages of searching (after concept or relationship identification; formulation of the initial strategy). Interestingly the sequences for search evaluation differ from OG1 as the evaluation of the initial results leads to query reformulation on this task. Also on this task no evaluation activity led to the exploration of synonyms and neither the evaluation of the size of the results nor article evaluation led to the diagnosis of the success of the search, as observed on task OG1.

As on task OG1 both experts and novices use cycles of concept and relationship identification (IC-IR, IR-IC), but experts augment this with identification of concepts and strategy formulation (IC-FS, FS-IC). Both experts and novices follow strategy formulation by the identification of new search concepts and strategy revision with querying activity (reformulation or the formulation of a new query); in addition, novices showed a significant sequence between strategy revision and the exploration of synonyms. Both subject groups showed reasoning about the problem and the formulation of a new query; and in addition to this experts also produced a significant sequence of behaviour between reasoning about the problem and the exploration of synonyms. Experts follow the exploration of synonyms with querying whereas novices do not. The only other differences occur in results evaluation where experts evaluate the size of the results set and then either reason about the problem, explore synonyms, reformulate queries, evaluate articles or evaluate search success; whereas novices show that the initial evaluation is followed by either query reformulation or article evaluation. It is noticeable these inter-group differences occur for the activities link with formulation, adaptation and amendment to the problem pre-searching. Thus it appears that experts exhibit richer and more sophisticated evaluation strategies even if novices do spend more of their retrieval sessions evaluating the results.

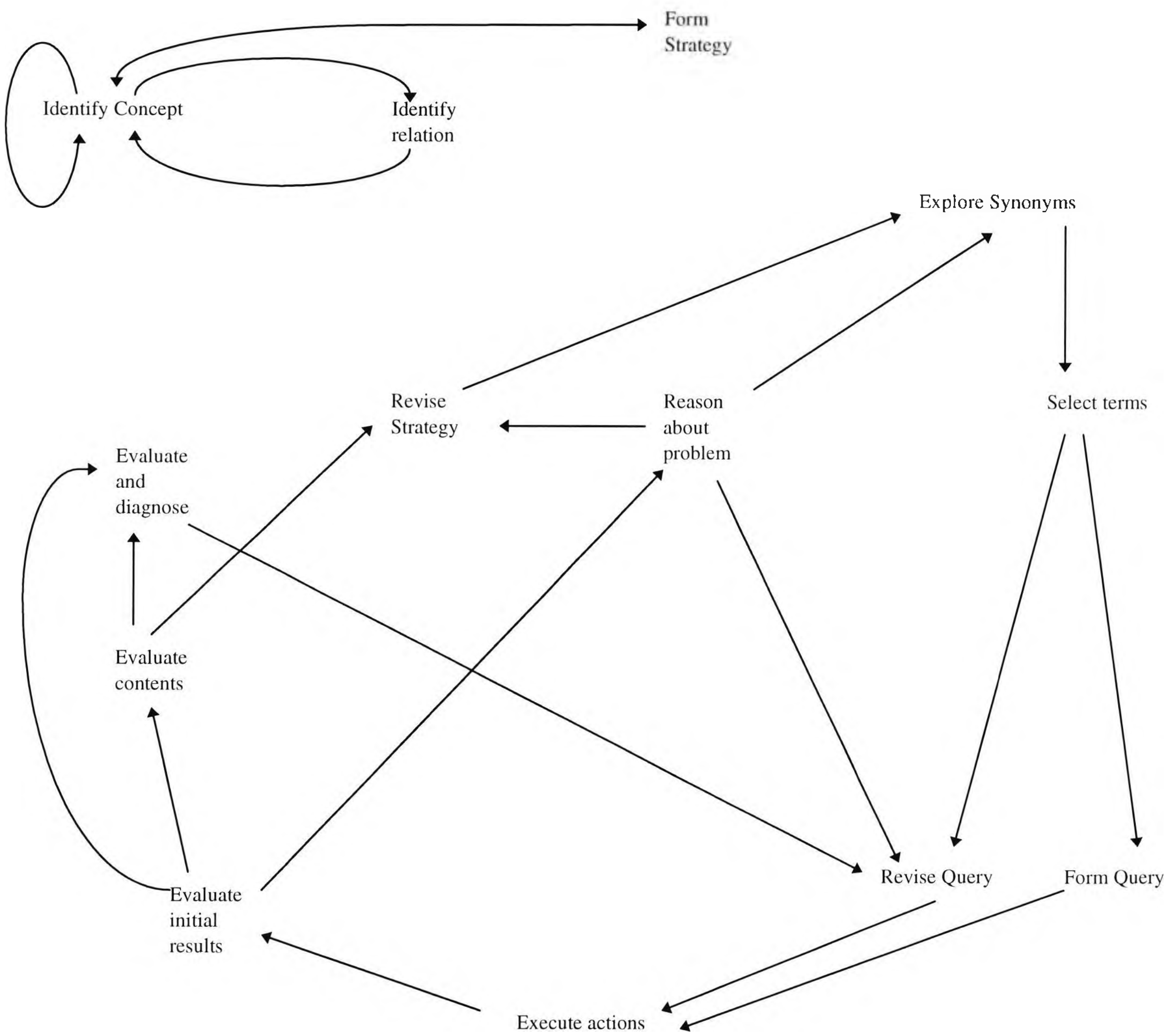

Figure: 5.13. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of experts on task PH1. The significance transitions at $p < 0.05$ are shown. With a total of 21 significant sequences of behaviour.

Figure 5.13 shows the behaviour sequences for experts for task PH1. As with tasks OG1 and OG2 the figure shows similar cycles of activity (IC-IR, IR-IC; IC-FS, FS-IC) but it also shows the iterative cycle of identifying concepts, also highlighted on task OG2. The figure shows fewer search reformulation sequences from results evaluation are used on this task (6 post evaluation behaviour paths) compared to the other tasks (OG1 11 post evaluation behaviour paths, OG2 9 post evaluation behaviour paths) and the post evaluation behaviour paths not present on this task are associated with the exploration of synonyms.

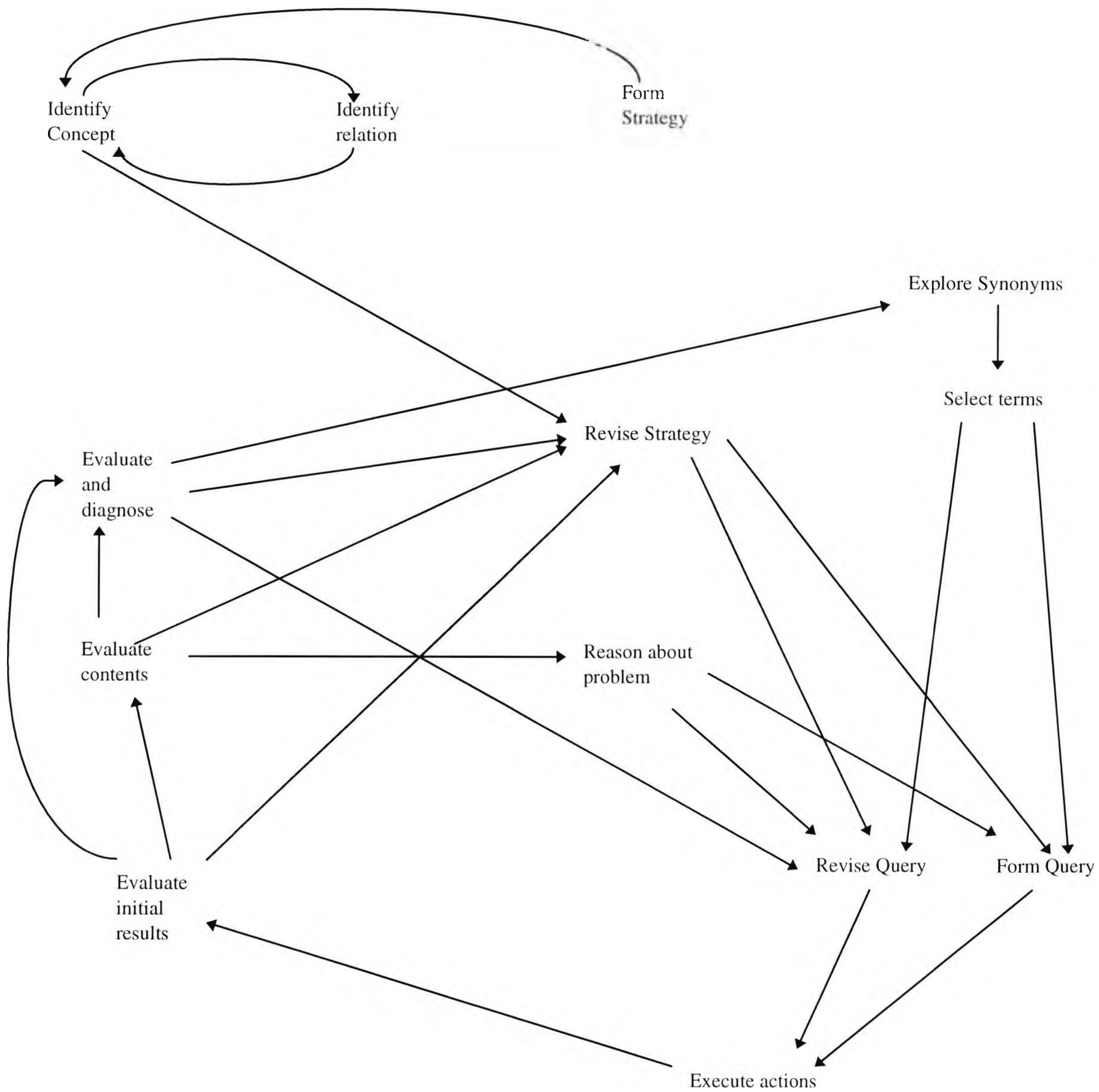

Figure: 5.14. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of novices on task PH1. The significance transitions at $p < 0.05$ are shown. With a total of 23 significant sequences of behaviour.

Figure 5.14 shows the behaviour sequences for novices for task PH1. The behaviour sequences from the identification of concepts, identification of relationships and strategy formulation are the same as on OG1 and OG2 apart from the IC-FQ sequence being replaced by IC-RS. Strategy revision and reasoning about the problem lead to querying as on OG1 and OG2. The behaviour sequences observed on all three tasks up to the evaluation of the results are similar. As in tasks OG1 and OG2 article evaluation and search diagnosis are followed by strategy revision; but in addition evaluation of the initial results produced is also followed by strategy revision. In common with task OG2, article evaluation is followed by reasoning about the problem. As in OG1 and OG2 search diagnosis leads to query reformulation. In common with OG1 the evaluation of the initial results and article evaluation are both followed by search diagnosis. As on the other tasks the evaluation of the size of the results set leads to article evaluation.

Both experts and novices show the sequences IC-IR, IR-IC, FS-IC but they also show differences as experts follow the iterative identification of concepts; whereas novices do not and novices show an IC-RS sequence whereas experts show a IC-FS. The interesting inter-group differences involve sequences with: strategy revision, reasoning about the problem and exploration of synonyms. Experts produced a cluster of transitions between strategy revision, reasoning about the problem and the exploration of synonyms (RPS-RS, RS-ES, RPS-ES) whereas novices do not. Novices exhibit more entry points (ER, EC, ED) for strategy revision from results evaluation than experts (EC). Novices also show a significant sequence between search diagnosis and synonym exploration.

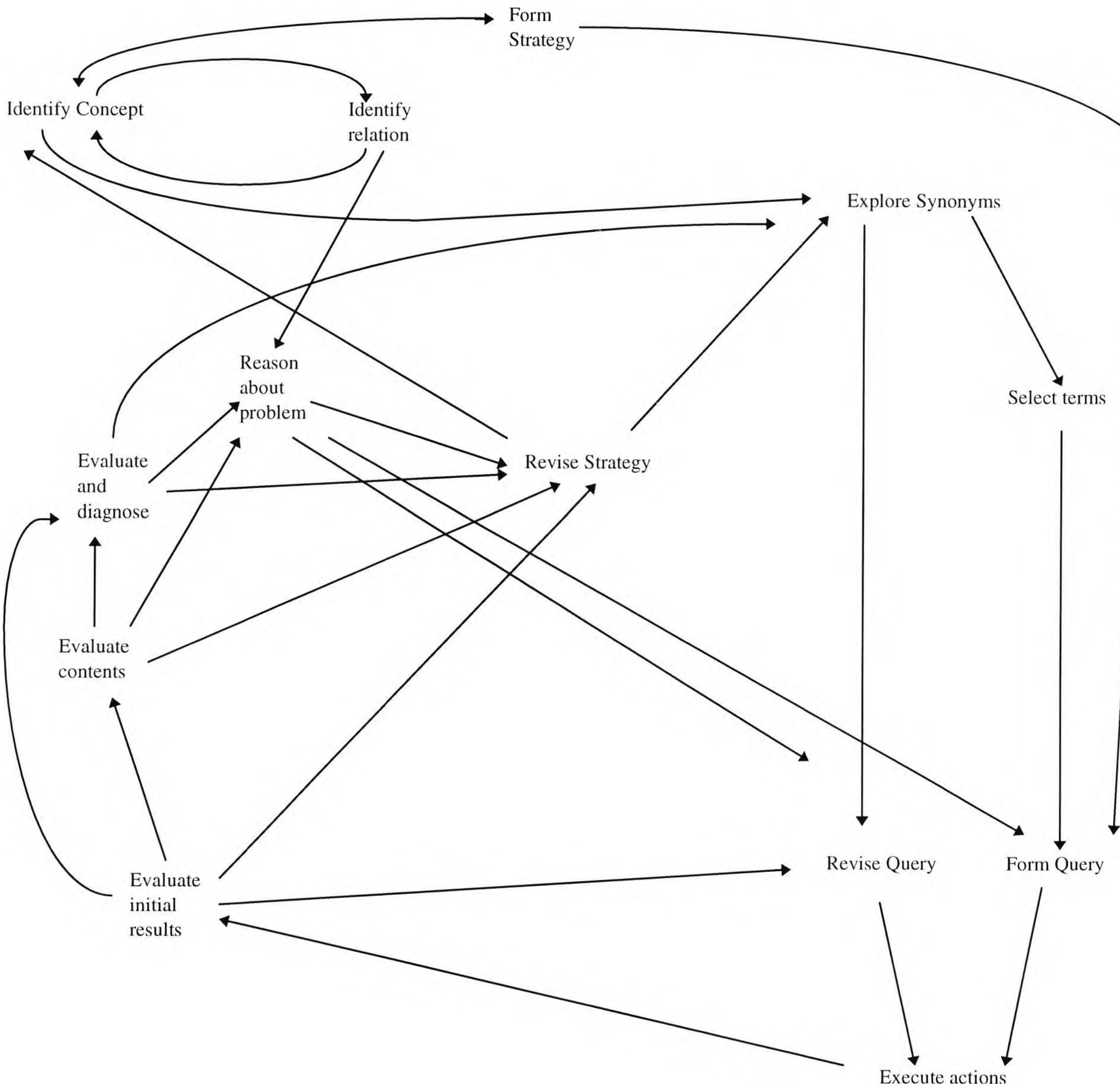

Figure: 5.15. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of experts on task PH2. The significance transitions at $p < 0.05$ are shown. With a total of 28 significant sequences of behaviour.

Figure 5.15 shows the category pair sequences for experts occurring significantly more than expected for task PH2 based on a first order model of expected pair sequence occurrence. The experts show the same IC-IR, IC-FS, IR-IC, FS-IC sequences observed on the other three tasks. The notable absence is the iterative identification of concepts found in tasks OG2 and PH1. The subjects also followed strategy revision with the identification of concepts, a sequence not observed on the other three tasks. The RPS-RS, IR-RPS sequence of behaviours is significant for PH2 but not for the other tasks. Search diagnosis to query reformulation is absent in the more complex tasks (OG2, PH2) but is present in the simpler tasks (OG1, OG2).

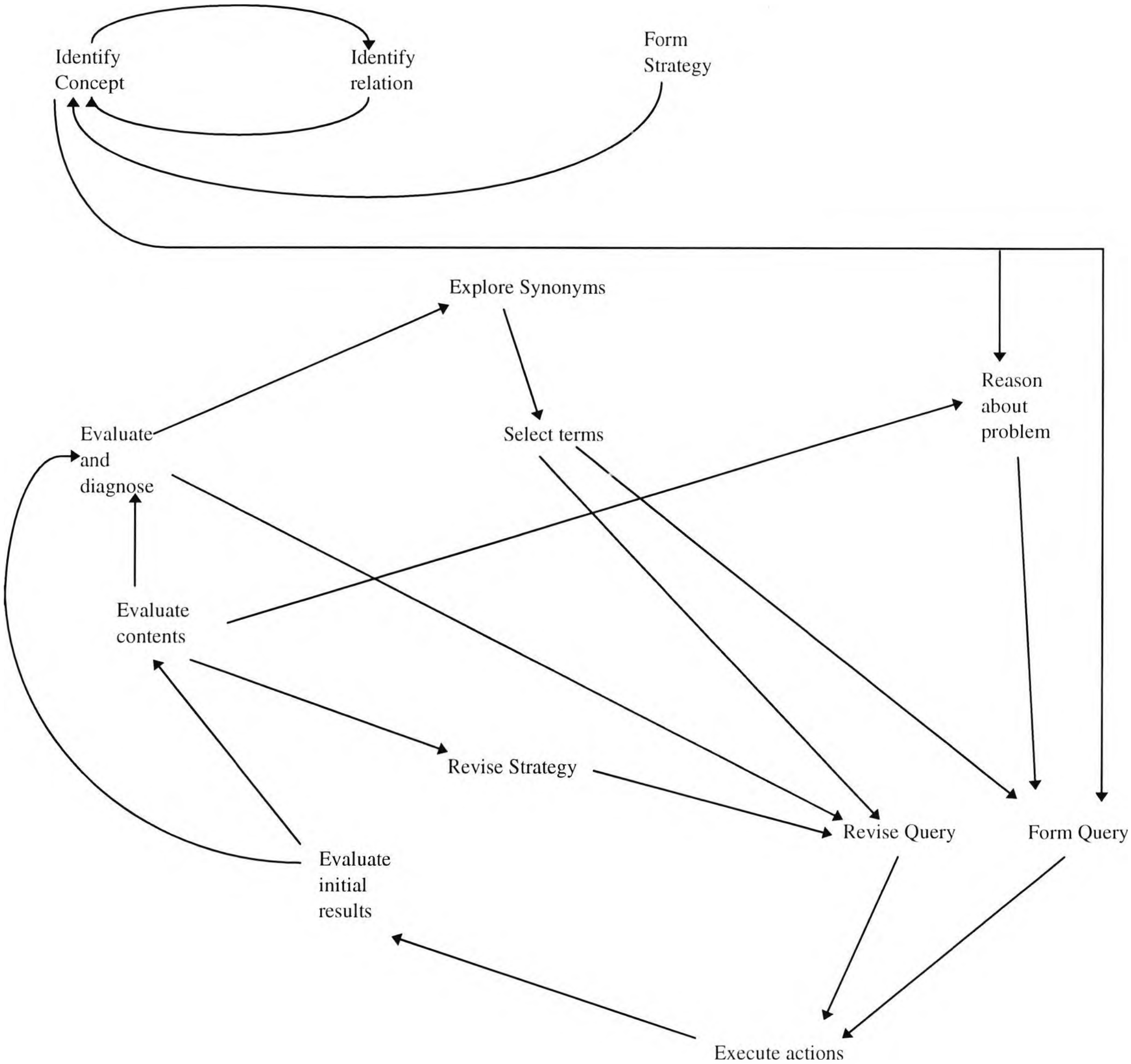

Figure: 5.16. Transition diagram illustrating the significant category to category transitions occurring in the behaviour of novices on task PH2. The significance transitions at $p < 0.05$ are shown. With a total of 20 significant sequences of behaviour.

Figure 5.16 shows behaviour sequences for novices for task. Novices show a IC-RPS sequence not observed on the other three tasks. The transition between initial results evaluation and query reformulation occurs only on the PH tasks and not on the OG tasks. Novices do not show the search diagnosis to strategy reformulation transition observed on the other three tasks.

Experts exhibited ten significant sequences not shared with novices; whereas novices showed only two significant sequences not exhibited by experts. In addition to the common expert-novice transitions experts follow the identification of concepts with strategy formulation or the exploration of synonyms; whereas novices favour reasoning about the problem. The main expert-novice differences occurred: pre-search planning (formulate strategy, revise strategy, reason about the problem and the exploration of synonyms) and reformulation (reformulation is triggered from the evaluation of the results quantity, article evaluation and search diagnosis). Experts follow strategy formulation with querying; strategy revision with the exploration of synonyms and reasoning about the problem with strategy revision whereas novices do not. Novices show a significant sequence between strategy revision and querying which experts do not show. Experts show many more routes into reformulation and changes in the current approach to the problem than novices. In addition to novice sequences experts follow the evaluation of the size of the results set with either strategy revision or query reformulation; and search diagnosis with either strategy revision or reasoning about the problem whereas novices show none of these sequences. Novices show a significant sequence between search diagnosis and query revision not found in expert behaviour.

5.4.2.4.2.Summary of the effects of user device knowledge on the significance of the occurrence of sequences of categories of behaviour

- Experts produce more significant transitions of behaviour on three of the four tasks (OG1 23%, OG2 37%, PH2 40% greater than novices). On PH1 the number of significant sequences of behaviour is approximately equal. Therefore it appears that expert behaviour is richer and more varied than novices.
- Most expert-novice differences occur in behaviour patterns for: pre-search problem manipulation and outcomes of results evaluation.
- Experts follow strategy revision with the exploration of synonyms whereas novices proceed straight to querying.
- Experts proceed to synonym exploration from more categories of behaviour (concept identification, strategy formulation or reformulation, reasoning about the problem, evaluation of the size of the results or search diagnosis) than novices who only tend to show significant sequences after the evaluation of search results. Therefore experts are exploring synonyms in a greater number of situations.
- The iterative cycles between the identification of concepts and relationships is shared by both subject groups in all tasks. Experts follow concept identification with the formulation of a strategy irrespective of the task whereas novices do not. Both subject groups follow the formulation of a strategy with the identification of the associated concepts on all tasks.

- The network diagrams show cycles of reformulation following evaluation either by exploring alternative approaches to the problem (strategy revision, reason about the problem or explore synonyms) or updating the query given the information encountered during the results encountered. Experts appear to show more courses of action following results evaluation than novices, except on task PH1 in which experts produce fewer sequences of behaviour from the evaluation of results than on the other tasks.

5.5. Summary and conclusions

The subjects behaviour highlighted that experts and novices identified a significantly different number of concepts; this was not expected as both subject groups had high domain knowledge. A complete explanation for this difference can not be given but it is possible that users only reported concepts intended to be used in queries and experts used more terms than novices. Experts revised their strategies, reasoned about the problem and explored synonyms substantially more than novices; and these differences are significant for the majority of tasks. This indicates that the retrieval activity of experts and novices match the strategic differences identified in chapter three. The differences may occur due to experts' strategic knowledge of the 'best practise' for retrieval in different contexts and thus activities which assist in a more complete need articulation occur more often; whereas novices may minimise the complexity of queries by favouring article evaluation.

Experts favour pre-search problem development and planning whereas novices favour evaluation, and this is indicated by the differences in the configuration of the total activities observed. Article evaluation and the diagnosis of search success account for a greater proportion of the activities observed for novices than experts for the majority of tasks. Strategy revision, problem reasoning and synonym exploration account for a greater proportion of expert activities than novices for the majority of tasks. Experts perform over 50% more new query formulations and over 35% more query revisions than novices.

Experts are observed to conduct many more activities per minute than novices irrespective of the task. The main subject group differences occur in the number of times per minute a strategy is revised, reasoning about the problem is performed, synonyms are explored and queries are formulated or reformulated. Experts are observed to conduct many more of these activities per minute than novices. This difference may occur for the following reasons: experts are more skilled and can thus complete activities quicker, or the activities favoured by novices tend to be more time consuming (i.e. article evaluation) than query revision activity, especially if the user group is more tolerant of evaluating a large number of articles.

State transition diagrams show the varied sequences of the activity observed in expert and novice behaviour. Users make reformulation judgements based on different levels of results evaluation. This was expected as the cognitive model specifies that different evaluation sequences will exist based on the users perception of the quality and completeness of the results produced. The sequences of novice behaviour appear to indicate their retrieval activity is simpler than experts. For example, novices always follow strategy revision with querying activity on all tasks; whereas experts may also explore synonyms or execute the device to return to a previous results set. Experts may revise their strategy, reason about the problem, explore synonyms or reformulate their query based the view of the acceptability of the results quantity; whereas novices will only explore synonyms or reformulate the query, but they will revise their strategy (on OG1, PH1 or PH2) or reason about the problem (for complex tasks) after evaluating articles. It seems that novices require more detailed information (i.e. article content judgements) than experts before committing to a strategy revision or reasoning about the problem. Experts, in addition to the article evaluation transitions observed for novices, may make strategic changes using a minimal amount of information (i.e. results set size). This difference may be due to a greater awareness of a greater variety of strategies and tactics by experts, or due to the effects of device expertise on users confidence to develop strategies and problems without a need to address the content of the articles retrieved. The state transition diagrams indicates that experts produced less absolute state to state transition frequencies above the 1% threshold on PH2 than the on the other tasks confirming this task may have been too difficult and searchers gave up. The complexity of novices state transition diagrams are approximately the same but they produced much less activity than experts. The other complex task (OG2) produced more transitions occurring above the 1% absolute frequency threshold than in other tasks, for experts (more different transitions above the threshold OG1 and PH2, more activity for the same transitions on task PH1).

The significant sequences of behaviour indicates users apply a definite structure and method in retrieval activity. This highlights that it may be possible to assist user retrieval by supporting the expected sequences of activity given the users current context (e.g. evaluation and reformulation advice given the results to a query). Experts are observed to produce more significant sequences of behaviour than novices on three tasks (OG1 23%, OG2 37% and PH2 40% more) and this emphasises that experts use a greater variety of search options than novices. This may indicate differences in the strategies and tactics users apply in a retrieval situation. The inter group differences in the behaviour sequences indicates a necessity for models of expertise and novice behaviour. Expert behaviours augment the sequences observed for novices, and this matches the view of retrieval given in the cognitive model in which the options available expand as device knowledge increases. The expert-novice differences for the significant sequences of behaviour occur either in the pre-search manipulation of the problem or the possible outcomes from results evaluation. Experts are observed to proceed to synonym exploration from many more categories of activity than novices. Thus the use of different approaches to searching will require different systems support to enhance and improve the effectiveness of retrieval behaviour.

The model described in chapter 4 can only be partially validated owing to the number of possible search contexts to be explained. The model is complicated as it must account for a large number of possible combinations of information need, user type, search results and system facilities. It is not possible to test all combinations and therefore only part of the theory can be validated. The model requires validation at multiple levels due to this complex and dynamic nature:

- process level (activity level),
- strategy and tactic level.

However, the protocol study presented in this chapter only addresses the occurrence and sequence of user behaviour at the process level. The validation thus operates at a higher level of granularity than the strategies and tactics identified in chapter 4; as it was not possible to accurately identify the strategy and tactic changes in user behaviour from verbal reports. For example, a user protocol may not include reports of all strategy revisions (or may do so retrospectively) hence, if attempts are made to match these to protocols, false correlations between the timing of user strategy reports, actions and activity may be drawn. The protocol study validates the behaviour sequences identified within the theory against those observed in user behaviour. To achieve this sequence differences in the behaviour of users with different levels of device knowledge have been discussed; see sections 5.4.1 and 5.4.2. This section is used to describe the relationship between the verbal report categorisation and the activities within the model; see figure 5.17. The model indicates the expected optimal behaviour; whereas observed behaviour is not as efficient. Novices' protocols recorded sub-optimal behaviour using inefficient approaches to compensate for the low knowledge of the device. Expert subjects did not articulate all strategies as their skilled knowledge is implicit so automated retrieval behaviour is not reported, as found in Rassmussens' (1996) studies.

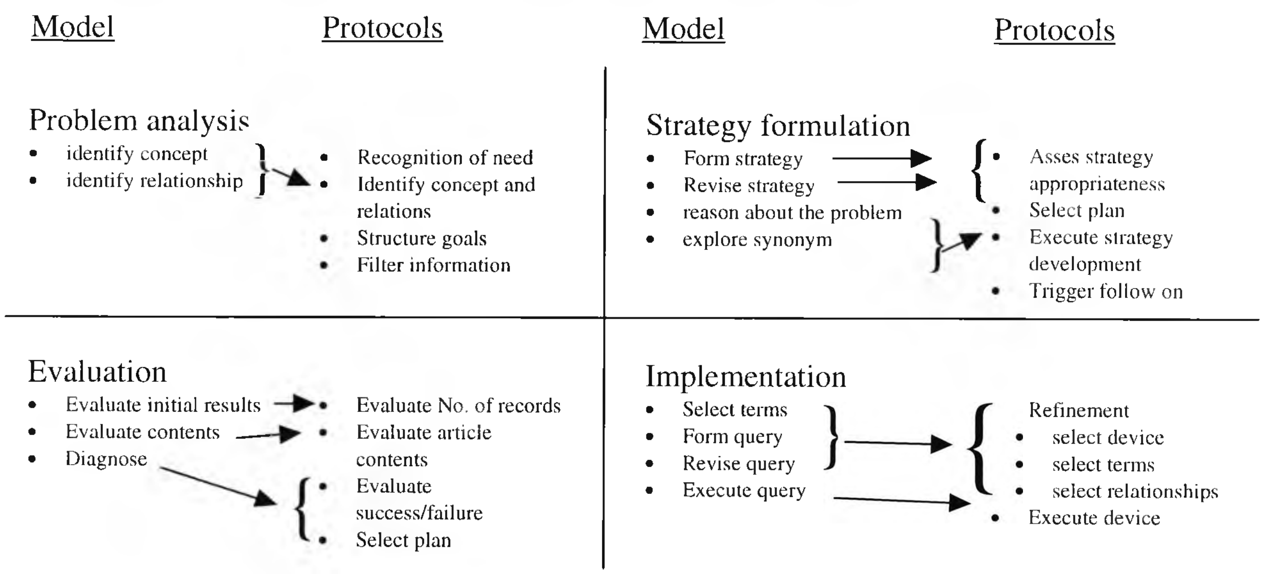

Figure 5.17 Mapping between cognitive model and verbal protocol categories. Within the main process activities verbal report categories appear on the left and model activities on the right.

The activities included in the model of user behaviour are more detailed than the verbal report categories used in the protocol study. The activities contained in the model can not all be directly mapped to verbal reports as the occurrence of some of the activities within the model could only be inferred given the other user activities e.g. the assessment of the different possible strategies for the current search context is inferred to have already occurred given a user reports a strategy for search completion has been selected. The differences between the categorisation of users' verbal protocols and the activities in the model occur because of difficulties encouraging users to report all activities during retrieval sessions and the incomplete nature of verbal protocols. To generate more complete activity traces it is possible to derive some activities based on the gaps in user utterances, the state of the system and users actions; but it would not be appropriate to include these in the categorisation as judgements concerning their occurrences requires the experimenter to make inferences when encoding protocols based on speculation rather than actual user behaviour, and would thus be inappropriate as part of model validation.

As an example of the link between the activity sequences within the theory and actual sequences of user behaviour the matching for the evaluation process is shown (see figure 5.18). The figure indicates that the expected loops of activity generated within the model are found in the protocol study; but in addition to the expected sequences, users showed more possible termination points than described in the model. The behaviour sequences in the model represent an idealised view of complete retrieval behaviour; whereas the protocols showed that actual users may terminate results evaluation much earlier than expected and proceed directly to strategy formulation and implementation without diagnosing the search success or failure. The model predicts sequences with complete activity traces whereas those found in actual user behaviour include behaviour short-cuts or user errors. However, even though the model fails to completely mirror all sequences of behaviour it does provide a correct explanation of the reasons for early search termination i.e. in evaluation cut-off criteria. The model does indicate the correct reason for users' activity but it does so through reasoning processes which the protocol subjects did not always report. The reason for this divergence may be that either users are performing this diagnosis activity and failing to verbalise it, or that their activity is less systematic and exhaustive than expected. Thus once a failure criterion is identified it is immediately reacted to instead of the user applying a more considered approach to identify the cause of the failure e.g. if too few articles are produced instead of investigating if this is attributable to specific query elements a more general tactic is applied. The model is able to determine the activities requiring support but its specification of process (activity) sequences need to be formalised by including probability based mechanisms within correspondence rules if predictive power is to be enhanced e.g. in a specific context the probability the outcome will be activity A is P=0.6 and the probability of activity B occurring is P=0.4.

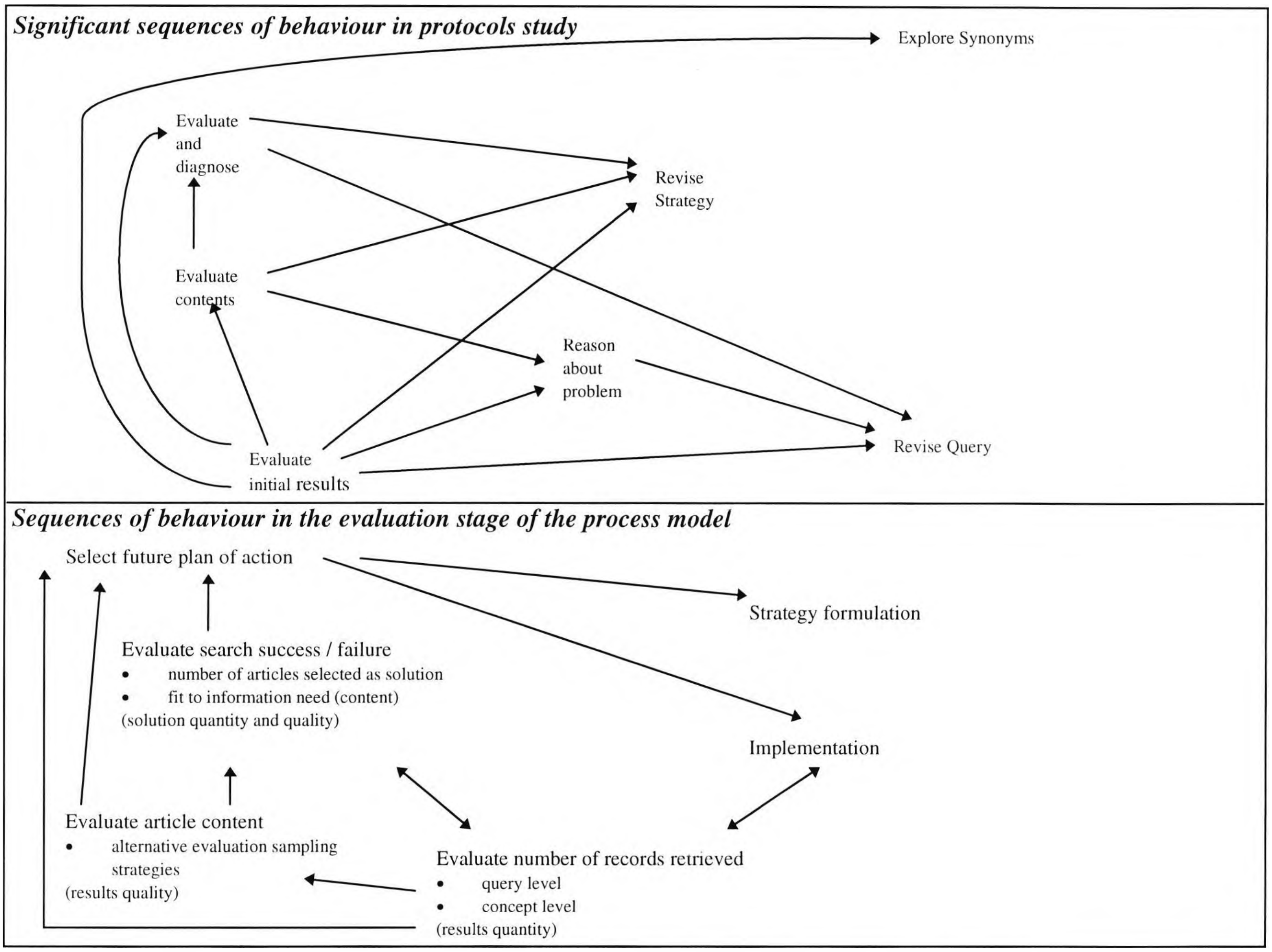

Figure 5.18: General process routes from the evaluation process a comparison of the model routes to those identified within the protocol study

The model also offers insight into the different process sequence pathways. Experts behaviour patterns augment novice behaviour with more detailed knowledge of strategies and tactics which may lead to success and more efficient retrieval. As a consequence, expert behaviour traces should be richer and more complex as they are able to perform a greater variety of activities in response to a search context. Novices will prefer a minimal use of the device whereas experts make more effective use of a wider range of system facilities. This is expected owing to the 'cognitive load' associated with learning and operating the system, users' confidence using the system, users' mental models of the facilities or methods of improving search efficiency (either device operations, strategies or problem development activities applicable). Experts are expected to formulate an accurate and complete description of the information required, will explore problems pre-interaction; whereas novices are likely only to explore alternative problem articulations after failure. Novices also avoid complicated Boolean expressions. Novices may not be aware of the problem exploration required for effective query articulation and thus recognise this need only if faced with search failure. Experts search activity will be quicker as it should be more automated than novice users whose activity is likely to be slower as they must attend to each behaviour step consciously.

To conclude differences in the composition and frequency of activity of expert and novices were found for some of the categories of activity. The study has further strengthened the argument that the device knowledge of a searcher effects the strategies, tactics and approaches to information retrieval. Some transitions found in the cognitive model were observed, and differences in sequence indicate that experts may produce a greater variety of activity sequences than novices for whom the pattern of search activities and retrieval behaviour is simpler. The study has emphasised the need for models based on user expertise and highlighted the need to support different strategies and sequences. The inter-subject group differences found are caused by the differences in the strategy and tactics used, unfortunately this protocol study could not identify these owing to the difficulties pinpointing the exact time and identity of users strategy and tactics. The different activity focuses and the proportion data indicate users with differing knowledge pursue alternative strategies but it is not possible to match these to the theory owing to identification problems. It is therefore only possible to state that differences expected between experts and novices were found at the process route level. The protocols confirms the basic sequence of user behaviour conforms with the model. However a more detailed analysis of individual traces is not possible because it is difficult to capture values user specific variables such as motivation.

Chapter 6

Implications for information retrieval systems design

This chapter discusses the user activity requiring support and how to provide this. A concept demonstrator is described with associated usability studies.

Chapter Six: Implications for information retrieval systems design

6.1 Introduction

This chapter explores the activities, strategies and tactics a retrieval system needs to support if it is to help improve performance and user confidence in the results produced. The chapter explores principles and properties of good design practice based on the empirical work already discussed (chapter three and five) and the model of user behaviour (chapter four) and its effects on design decisions made. The discussion centres around the task support facilities (TSFs) required within systems to support user retrieval activity. A full list of the possible TSFs with examples of systems in which they have been implemented can be found in the taxonomy of TSFs presented in chapter four. The purpose of the discussion of TSFs in this chapter is to ground the functionality offered by current retrieval systems within the cognitive model and systems design decisions.

An example application built for a specific category of user is then described which embodies context sensitive advice, user-system dialogue and some of the TSFs required to support the user. The systems architecture for the tool developed is discussed, with its associated functionality and interface. Two usability sessions with a total of 22 users are described to validate users' views of the interface and search dialogue provided. The usability sessions tested the interface on user groups with different backgrounds and on different task scenarios to determine the systems applicability in different contexts. To conclude this chapter, the effectiveness of the system interface, guidance and search dialogue are discussed.

6.2 TSFs and system support necessary to enhance user retrieval

Problem analysis support

Problem analysis activity is not directly supportable by the device as much of this activity occurs pre-interaction, but the system can provide facilities which encourage and assist error checking. This can be achieved through the problem representations used for formulating queries and query histories e.g. simple hypertext diagrams to show concepts and their relationships. The quality of the system representations is reflected in the ease with which users are able to decide on a course of action and adapt the query in the light of the search results produced. The information representations should indicate the limitations of information contained in the database as this will enhance the re-analysis process e.g. indicating if a particular concept occurs in a low number of articles (hit density maps Cugini 1997) or if no alternative terms are in the thesaurus.

Strategy formulation and reformulation support

Strategy formulation and reformulation are some of the major areas in which the system can provide advice and assistance to aid users produce more complete and targeted behaviour. The advice provided should prompt the user with appropriate strategies and tactics given their current retrieval context. For device experts this advice and prompting can act as an aide memoir; whereas for novices it will assist in training of efficient retrieval techniques, whilst not requiring them to actively acquire the in-depth retrieval knowledge necessary to attain effective performance.

Users should, apart from on precise queries, expand concepts by using alternative term representations for a concept. This leads to a requirement that systems should:

- promote concept expansion by the user-system dialogue and guide the user towards more optimal behaviour. This may be encouraged by structured dialogues, and by initiating automatic term suggestions from thesauri. The user may then be encouraged to consider term alternatives as a matter of course as queries are formulated; rather than just in response to search failure, as observed in the empirical studies (chapter three, trial and error strategies; chapter five, the common transitions between evaluation activity and synonym exploration). In general the use of richer query representations will lead to more effective retrieval; and it is important to raise user awareness of the need to expand concepts during query articulation, as users may assume the system performs associated term expansion automatically (informally reported in protocol study presented in chapter 5).
- System term suggestions must be presented in context i.e. concept or term displayed in a hierarchy of broader, narrower and synonym terms.

- the presentation of terms with the same base stem but alternative endings by thesaurus structures may help to correct users' false assumptions that IR systems match intelligently on the semantics of information.
- functionality to allow the user to re-visit the thesaurus structure for a particular concept and adjust the terms used is necessary; the effects of adjustments on the size of the associated results set should be immediately apparent thus facilitating recognition of the contribution individual query elements make to the results.
- For example:

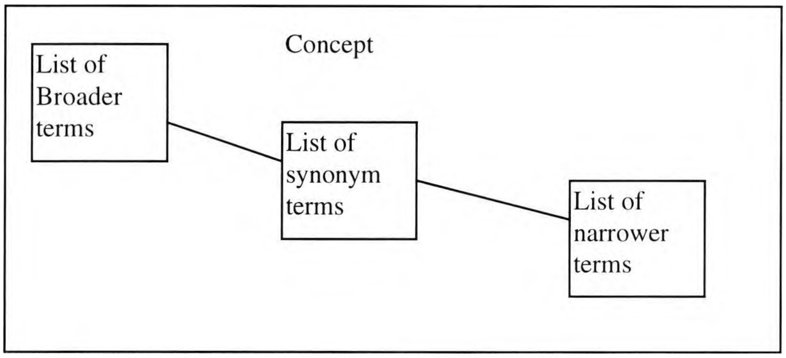

Figure 6.1: Example interface representation for term expansion

The representation of the current query and interaction aimed at assisting query development should aid the process of structuring information, determining the 'conscious need' (Taylor 1968) and confirming the scope of a problem in an attempt to minimise differences between the 'conscious need' and the 'articulated need'. This issue is especially pertinent for novice users with a limited understanding of either the domain or the device. The representations used for the current query, as part of reformulation dialogue, must

- permit the simple adjustment of query components (concepts, terms or relationships) especially for device novices who find difficulties in expressing complex Boolean relationships. This requirement originates from the use of the system to determine the granularity of queries developed through cycles of broadening or narrowing.
- Visual and graphical approaches to query representation and its development should have advantages over textual equivalents as they help to segregate query components as identifiable objects. The use of graphical representations over text is especially effective in simplifying the management and representation of

complex problems. For example (See figure 6.2) the use of direct manipulation, such as query by pointing, to adjust query element inter-relationships as opposed to constructing the syntax of a correct Boolean expression. The empirical results of chapter three and Sewell et al (1986) indicate that novices require a method of specifying relationships other than text based Boolean expressions as subjects neglected to use ‘OR’ relationships favouring multiple parallel queries using substitution and ‘AND’ relationships. Spatial metaphors, or highlighting, should be used to indicate the distribution of the articles retrieved in relation to query elements (Ahlberg and Shneiderman 1994). For example concept A has 20 articles associated with it whilst concept B has only 5. This gives immediate localised feedback about problem components and aids diagnosis by highlighting the concepts which require alternative articulations.

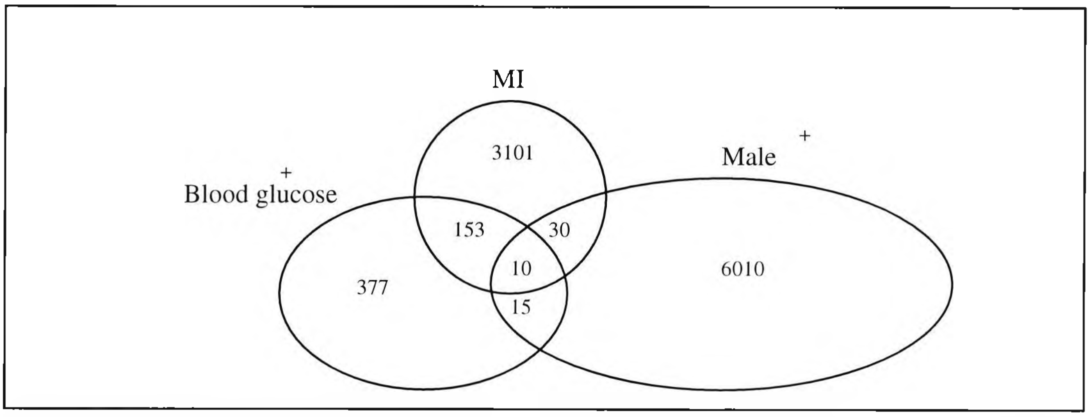

Figure 6.2: Example of a possible query representation. Male^{+} is a representation that concept male with the synonyms males; men or man are part of this set. This layered approach to query representation keeps it clear from clutter when the searchers query is complex.

- It is clear from the empirical study described in chapter five that users with differing knowledge of the device make reformulation decisions based on different levels of information and so it may be necessary to tailor advice to give novices the confidence to make these decisions earlier in the evaluation process.

The query history facilities must:

- support the re-use of strategies and queries for problems with similar characteristics. Re-use may operate at two levels: general problem templates and user specific query templates.
- support the user in the compilation of complex queries from component queries.

- support traceability and provide enough information to uniquely identify the query by displaying its component parts (e.g. a list of the key words, concepts and the number of results retrieved). Difficulties arise in the design of interfaces to search histories if the queries are complex and contain multiple nested clauses as this may overload the history representation leading to errors and difficulties. For example, a query may contain two concepts but use five terms in the expression of each of these concepts. If all of the query information is included in the query history the representation is likely to become overloaded and thus ignored by the user. If only concept information is provided then it may be difficult to uniquely identify the content of queries.
- Highlights can be used to focus user attention on causes of error (query syntax or typing error identified by null query terms), or the area of interest within the information need. The representations must show clear links between queries and results as this should assist the user in search diagnosis. The interface and presentation of this information should improve the efficiency and effectiveness with which users can re-analyse and amend queries. The navigation of a query history and transitions between different results allow the user to berry-pick the information required or backtrack if a strategy becomes ineffective.
- diagnosis guidance should indicate the possible activities and the problem areas believed to have caused search failure by relating error slips to the different strategy techniques applicable given the previous query and the results set.

The aim of any system guidance facility and the human computer dialogue used is to aid the efficiency with which information needs are satisfied given the characteristics of the user, information need and quality of the results, whilst improving user confidence that the results contain a high proportion of the relevant information in the database. The purpose of guidance is to:

- reduce the user effort associated with the use of a search method by placing this workload on the system.
- to enable the different methods, or strategies, for locating information to be identified and to determine the most applicable strategy and its associated tactics given the problem and a current search context; and user-system dialogue must support these given the characteristics of a searchers knowledge, patterns inferred

from the search history and the quality of the results retrieved. The query reformulation dialogue should suggest and guide, term and problem expansion (or contraction) based on the size of the results set, user judgements of the quality of the results and the previous search query. For example if a searcher uses a query producing a large number of results the system should indicate the different methods that may be used to narrow the results. This is necessary if advice is to support the possible reformulation situations and refinement actions identified during the evaluation process of the cognitive model (see chapter four).

- Facilities must target user task resource requirements: Browsing may use information maps or interaction dialogues which enable the development of a more detailed understanding of the structure of the domain by guiding the searcher through classifications in the system thesaurus. However, if this type of browsing facility is the default interface it may hinder a searcher with a clear view of the information required.

Implementation process support

Search implementation enables the 'conceptual need' for information to be translated and constrained into a 'device specific articulated need'. The system can assist this process:

- by reducing action slips through the provision of spelling and query syntax checking functionality during query generation and by providing short-cuts based on the nearest alphabetical match to user input (see figure 6.3). The user may then reduce the possibility of spelling errors by selecting the concept required from system alternatives.

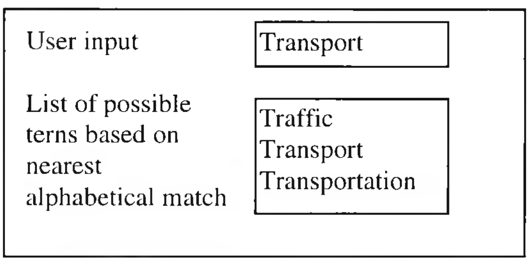

Figure 6.3: Example term inclusion short cut facility

- To minimise errors in query syntax intuitive interaction methods are necessary for specifying queries other than by textual Boolean expressions; as Boolean relationships are difficult for most users. Query articulation aids must be designed to reduce the effort associated with the production of queries. The aim is to remove Boolean query languages from the interface and replace this function by interpreter mechanisms based on diagrammatic representations of concept inter-relationships.

- Query representation facilities must enable the efficient extension of concepts (truncation) and inclusion of term alternatives without cluttering the query representation.

It is important that parts of a results set can be included in subsequent queries. For example if whilst reviewing a results set a term is observed which is thought to give added value to the query, then it must be possible to select and include it directly in the next query. The ability to merge two previous queries is essential and it is preferable that the elements of these previous queries can be inter-related.

Evaluation process support

The main aims of evaluation support and the representation of information in the systems are to:

1. Improve the user's ability to diagnose the extent to which the results match the information need and assist users to make these judgements;
2. Improve the appropriateness of tactics for reformulating searches given the diagnosis judgements made, the quality of the results and the users context. The aim is to assist in the recognition of triggers for the different types of reformulation and provide guidance through the user-system dialogue to enable users to apply the tactics necessary to update the query with a minimum of IR strategic or device knowledge.

Support is required to permit the evaluation of results at the different levels of granularity necessary to assist the different types of results evaluation, with the effects of query actions on the characteristics of the results being clear and directly associated with query components. For example the searcher may require information to be displayed to enable evaluation at the following levels:

- Quantity information: e.g. the number of records retrieved; the number of articles associated with individual components of the query

- Article characteristics: e.g. title, author, journal name; year of publication, number of times a particular query component occurs in an article.
- Gist information: e.g. highlighted keywords
- Articles content evaluation: e.g. reading the article to assess relevance

Users make judgements on the quality of the results retrieved based on different levels of information. The judgements may indicate the results should be evaluated in more detail or that the retrieval approach should be altered. Alterations in the retrieval approach can result in strategy refinement, query reformulation, reasoning about the problem or exploring synonyms. Information gathered during the evaluation may be used to update the query or strategy. For example the searcher must be able to select terms in articles and include these in the query without having to re-enter the term at the search interface.

The number of records associated with a concept should be displayed in close proximity to that term, to assist evaluation and query diagnosis. As an extension of this direct association between query representations and the results the effects of any alterations or development of a particular query element must be immediately apparent in the results. For example the system may highlight the null areas occurring between query components given the current query has retrieved zero articles. The advice of strategy and tactics must be provided given the current query, the users knowledge and the quality of the results retrieved (e.g. null set, article too general etc.). The results quality has to be elicited from the user; as these relevancy judgements are personal to a user and automatic relevance diagnosis is not possible. Reformulation in the evaluation process of the model (see chapter four) must be supported by dialogues to guide users through the process updating the query. For example given only a few records are retrieved by a query the system should suggest methods which broaden the search (i.e. term expressions; term alternatives; relationships alternatives and problem spaces at alternative levels of granularity).

Users are under the common misconception that the articles retrieved are ranked in relation to the query posed rather than on a first matched first displayed basis (informally reported whilst evaluating articles in the protocol study described in chapter 5). Some web search engines report results ranked in order of the percentage relevance to the query posed with the percentage relevancy displayed. This creates a user expectancy that the results retrieved rarely live up to as relevance is not only a function of term hits. Individual concepts in the query may be more relevant than others and so the concepts users determine to be important in the information need should be used to rank and order the results. This will assist searchers in the evaluation of the results as the information presented to them will be ranked based on their perception of the focus of the search. The presentation of articles should emphasise query terms or phrases, e.g. the results highlighting or underlining, in a attempt to reduce evaluation time by assisting article skimming techniques. In the empirical studies users reported strategies for evaluation which consisted of looking for keywords and reading around these to determine the relevance of a particular article before reading the article abstract, thus assisting the location of these keywords by highlighting may reduce the evaluation time. Percentage relevance figures for the articles should not be used as these figures tend to create unrealistic expectations from the results and reduce user confidence in the retrieval system; a subject in the usability session presented in section 6.4.2 stated ' I no longer use Altavista as it does not produced relevant results' and many subjects (PTRC usability session) in the discussions about the concept demonstrator indicated these figures of relevancy (in web based search engines) are not used due to their inaccuracy.

User group related issues

The diversity and expansion of the user population has caused a need to reduce the time and effort required to learn how to operate IR systems effectively. This causes search inefficiencies and reduces user confidence in the results. Interaction dialogues must guide searchers through the IR process if they are to be more effective.

Task support facilities (TSF) must be targeted at specific user groups in an attempt to:

- reduce the alienation of searchers with differing knowledge levels by providing guidance at an appropriate granularity;

- support users with different knowledge levels, thus the TSF's provided must be tailored appropriately to the activities and strategies being performed;
- provide targeted supportive dialogues which treat user-system interaction as a negotiation process rather than more abrupt query-answer interaction.
- the user should be provided with facilities which allow values for context variables to be specified, e.g. number of results wanted, as these factors alter their approach to retrieval. This additional information can be used to modify the interaction style and target retrieval behaviour by making the interaction dialogue and methods suggested sensitive to the database coverage required.

6.3 Design scenario

The multimedia broker demonstrator is concerned with the integration of Information Retrieval and Internet technologies, and developing support services to enable authors to find resources for successful multimedia publishing. The system is to support and advise users when searching and retrieving information from multiple, multimedia, heterogeneous databases. It will also support users in creating multimedia documents and applications, and will enable publishers to provide raw data to the Broker and charge users to access it.

The retrieval system is required to provide advice and assistance to users during retrieval activity to enable them to be more effective and efficient in two respects:

- query generation, development and problem articulation;
- results diagnosis and query reformulation.

The database containing document information is a publishers transport database (PTRC). The users of the database are central government officials, local government officials and researchers. The user population all have a background in the transport domain and are classified as domain experts. The intended users have a low knowledge of IR strategies and tactics and are thus device novices. The user requirements stated that the completeness of the solution, precision and recall, and the effectiveness of retrieval activity is more important than download times and cost. The characteristics of information needs of the user group vary dependent on the users' context i.e. not simply standard form fill query structures. The resulting application should provide access to information distributed throughout the web using one standard interface and no existing application exists.

6.4 Expert advisor application

The cognitive model adds the following high level requirements for the design of the application:

1. The systems' advice must be context sensitive based upon the users' characteristics, the task characteristics and users' perception of the results retrieved. The later of these will alter as the retrieval session progresses and is a function of the information expected and the results retrieved.
2. System dialogue should implicitly encourage the user education of 'good IR practice' but the user must retain the initiative and control the development of queries.
3. The presentation metaphors used for queries, the results presented and techniques used to encourage more complete need articulations (strategies, tactics and query development activities) must require minimal user training to be mastered.
4. The system must improve user confidence in the results provided through positive and context sensitive feedback. As such the advice provided must be adapted based upon the user's query history and the quality of results. The effects of pursuing a particular piece of systems advice should be directly associated with the query representations.
5. The system and its dialogue should not intrude on the 'natural progression' of the users information gathering task but it must also minimise the users' effort required to construct complete, and sometimes complex, queries.
6. The system should advise, but never dictate, allowing the locus of control to rest with the end-user.

6.4.1 Architecture

The design of the concept demonstrator required the development of a client side application and a server side application (see figure 6.4). The client side application is further sub-divided into the end-user interface and a database component used to drive the context sensitive advice and user-system dialogue.

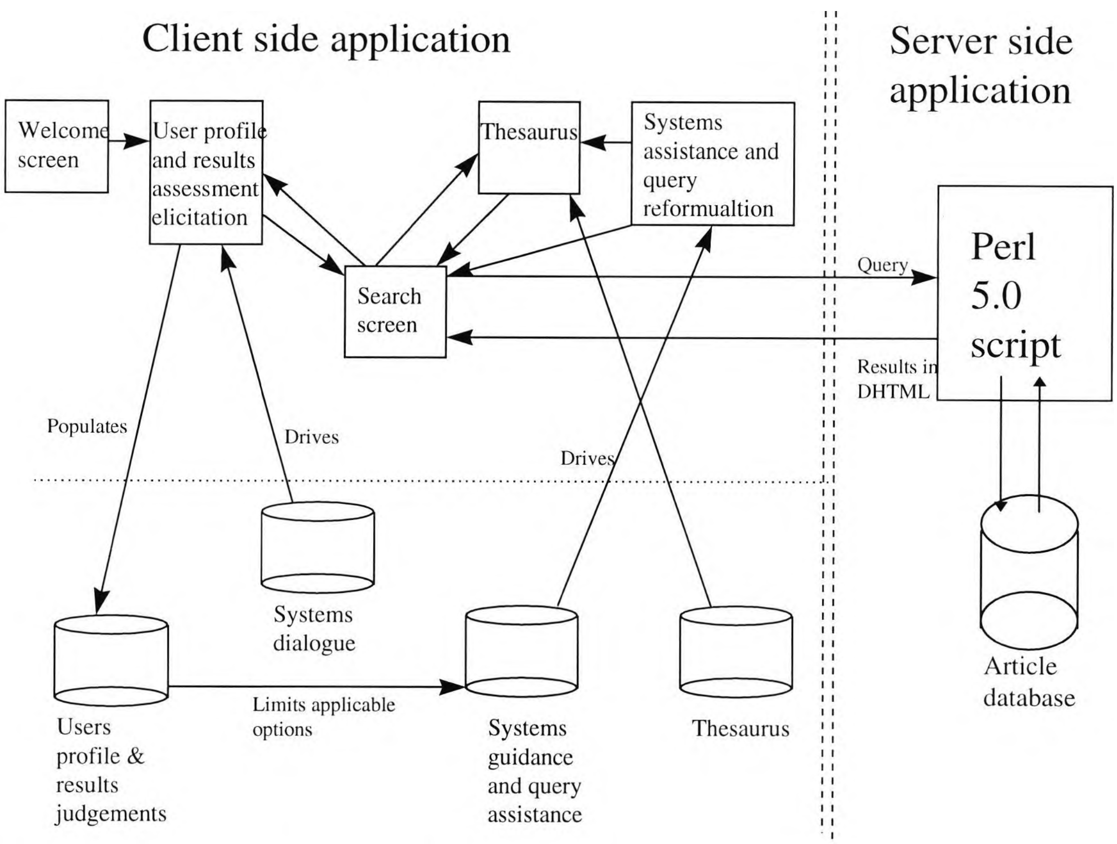

Figure 6.4: Systems architecture for the PTRC demonstrator application

6.4.2 Server side application

The server side application offers a skeleton text retrieval facility enabling the search for keywords, or part of keywords, within a document set. The database offers information on the title, author, country and the abstract. Keywords can occur in any of these fields. The retrieval system was developed in Perl and supports a reduced set of the possible Boolean relationships and nested queries (AND, NOT, OR). The output from the server-side application is a list of relevant articles encoded as DHTML. The server side application is not an important aspect of the concept demonstrator and the a implementation of the full server application is to be connected to the retrieval interface; but it is necessary to provide results to the users during querying. This is essential if the quality of the interface and the system advise are to be tested in live situations.

6.4.3 Client side application

The client side application is composed of two components: the interface and a local database application which controls the system-user dialogue and advice provided based on a user's context. The characteristics of the local database are discussed first, followed by the interface and systems dialogue.

6.4.3.1 Local databases, user profile and generation of context sensitive advice

A profile of the characteristics of the user enable context sensitive advice to be generated. The user profile is determined based upon the users answers to systems questions (see figure 6.5). The system supplies strategy and tactical options given the user's answers given to the system's questions. Subsequent user-system dialogues are adapted according to the system advice the user selects and the actions performed on the interface. The user can apply one or many of the strategies and tactics advised by the system. The user profile from which the advice is generated consists of information about the users knowledge, the characteristics of the current task, expectations about the information required and the users' perception of the quality of the current results. This information can be dynamically updated during a retrieval session. The information is elicited using dialogues for user and task characteristics; and results judgements. The user must provide information if the system is to provide context sensitive advice as the selection of the reformulation strategies and tactics applicable is based on a users perception of the results e.g. system can not determine if a set of articles are too general or specific as this is based on a users perception of the retrieved articles; but it can advise the user of different methods of adapting a query given a user has determined the articles are too general or specific. The system provides a number of strategy options based upon the current user profile. The users selection leads to the presentation of the tactics for achieving the high level strategy. The selection of the most appropriate tactic initiates a user-system dialogue which enables the manipulation of the current query. The dialogue guides the user through the adaptation of the query and tries to indicate the effects of the user's decisions on query development and the results set. Control over decisions is the user's; the system only advises and develops the query based on the users decision.

6.4.3.2 Interface and dialogue design

The mechanisms used to produce the advice for the second of the aims have already been discussed (see section 6.4.2) but in this section the interface provided to achieve these aims are discussed.

User profile elicitation

The provision of context sensitive advice relevant to the user's task and knowledge requires the user to provide information about their characteristics and those of the particular search task. A set of questions is used to elicit this information (see figure 6.5) at the search outset. Users can ignore any of the questions and the interface provides a reminder of the answers already given to questions. The information provided is used to populate a user profile in a database on the client side. The questions enable the user to specify their knowledge of the search domain, experience with this specific retrieval system, experience of retrieval systems in general, the general nature of the task, the complexity of the task, how well known the task is, the number of articles required, the completeness required and the format of the information required, see figure 6.5.

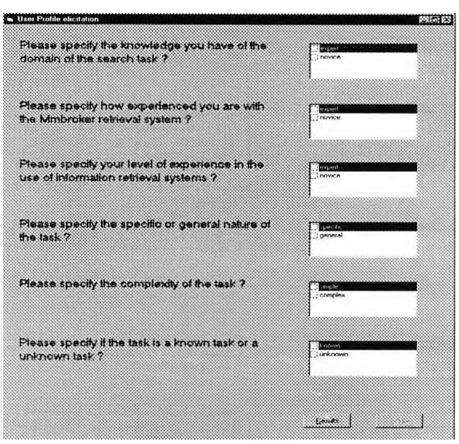

Figure 6.5: Question sets for user and need profile elicitation dialogue.

The system can now advise the user of applicable strategies before they proceed to the main search interface given their characteristics and that of the information need. Users can by-pass this advice if they wish and proceed straight to the main search interface.

Main search interface

All presentations of the query are displayed in the main interface and all sub-dialogue stems from it. As the user enters a concept the system provides the nearest alphabetical matches from the thesaurus. The user can incorporate any of these term options by selecting them. The submission of a concept to the system initiates a term expansion dialogue from which the user can add synonyms, broader or narrower terms from the thesaurus (see figure 6.7). A Venn diagram is used to represent the currently active query (traffic control and models in the screen shot). Each circle in the Venn diagram represents a concept in the users' information need. If synonyms are used for a concept they appear in a scrollable list box linked to the concept. To query the database a user must select an area in the Venn diagram. This triggers a concept level view of the query that appears for the area of the diagram selected. Multiple areas of the Venn diagram can be selected and included in a single query. The user then executes the search by sending the query to the database. The system constructs a syntactically correct Boolean query based on the terms, concepts and relationships associated with the area of the Venn diagram selected. The search history is updated with the concept level view of the query submitted. The results of the query are returned by the server side application and displayed as DHTML enabling them to be viewed at different levels (i.e. title and author or title, author, country and abstract). The articles retrieved can be viewed as a full screen or part screen as in figure 6.6. The user expands or collapses the view of a individual record and this enables a viewing history, or a view of the relevant articles, to be maintained by the user. The user can obtain strategy and tactical advice by answering questions about the results see figure 6.8.

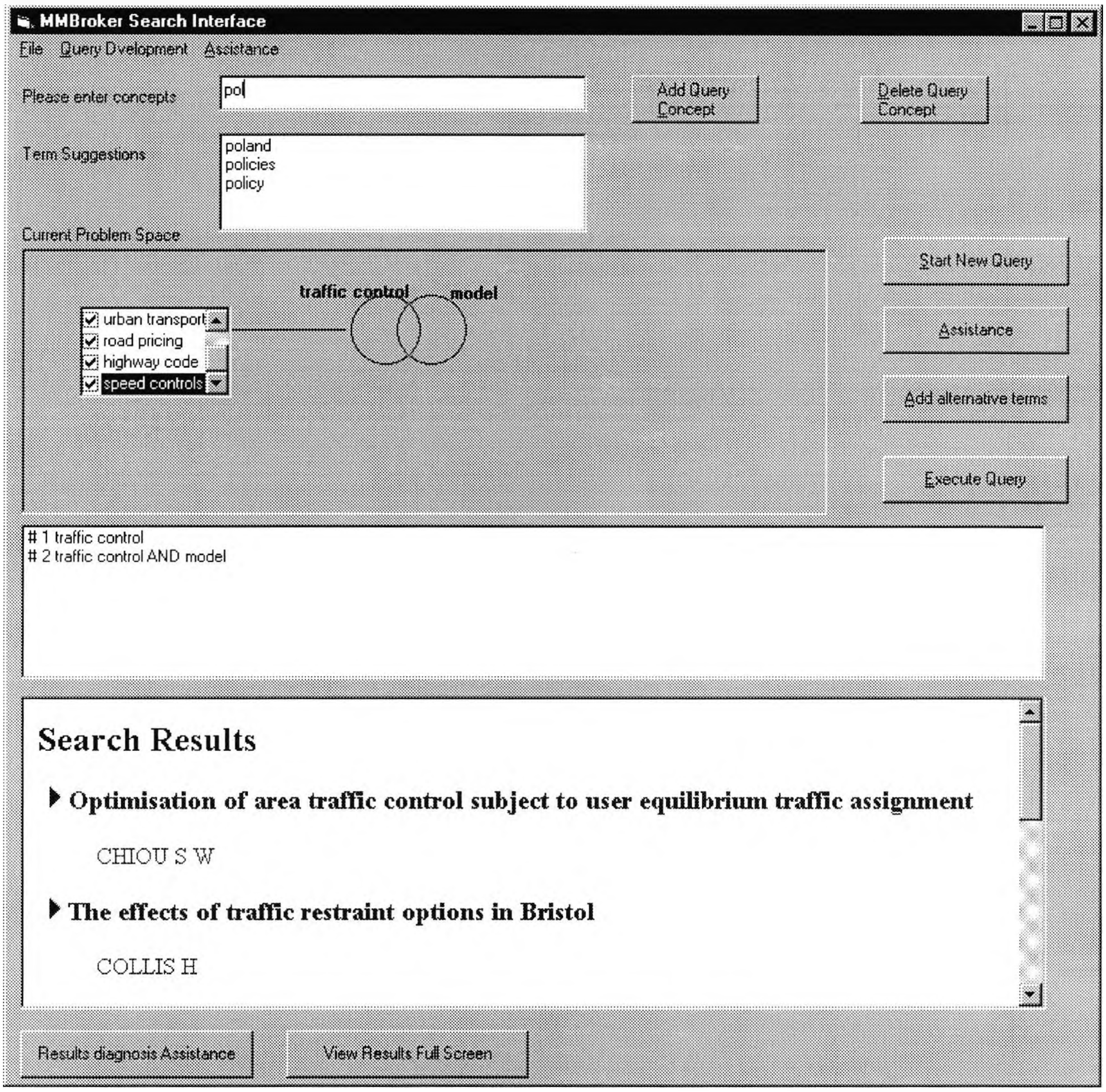

Figure 6.6 Main search interface for the concept demonstrator

Figure 6.7 shows the interface used to encourage the user to add alternative terms in query articulation. Magennis (1997) found this type of interactive query expansion (IQE) assisted in improving retrieval effectiveness. However, he found that term selection may be difficult for users inexperienced in term selection of search strategies. Thus systems may be required to provide support to assist in the effective use and selection of strategies and search terms. The screen shown in figure 6.7 is presented in two situations: if a new concept is added to a query or if the terms used to articulate a query concept are to be altered. The interface shows the sub-set of the thesaurus structure in relation to the concept. A user selects the terms necessary to accurately describe their information need and these are incorporated directly into the query. If the user is amending the query terms used as part of a reformulation tactic then this is achieved by deselecting terms they feel are causing the inaccurate results or selecting additional alternatives.

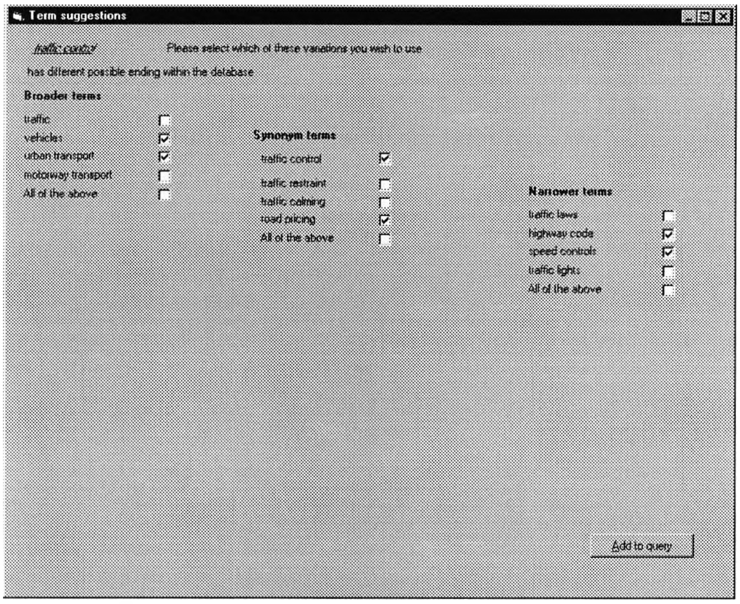

Figure 6.7 Term alternative relevant to the active query concept.

To provide context sensitive reformulation advice the user must describe the quality of the results produced by the current query, see figure 6.8. The questions operate at the different levels of evaluation described in the cognitive model (chapter four). Only a selection of these questions may be pertinent to the user's current situation as the relevance will depend on the evaluation activity performed. The user only needs to answer one of the questions to obtain advice.

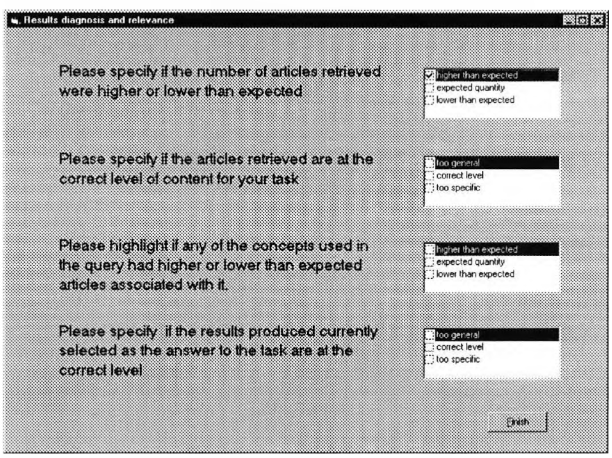

Figure 6.8 Quality of results elicitation dialogue.

Figure 6.9 shows the interface used to present the strategies and tactics relevant to the user's current situation. The system presents high level strategies based on the answers the user gave for the results profile. As a strategy is selected the relevant tactics are displayed in the list box. The selection of a tactic initiates a reformulation dialogue, which suggest the actions required to fulfil the tactic given the current search situation. User actions during this dialogue update the query and search history as appropriate. The system only offers advice at the users request and the quality and context sensitive nature of the advice given depends on the users answers to the questions.

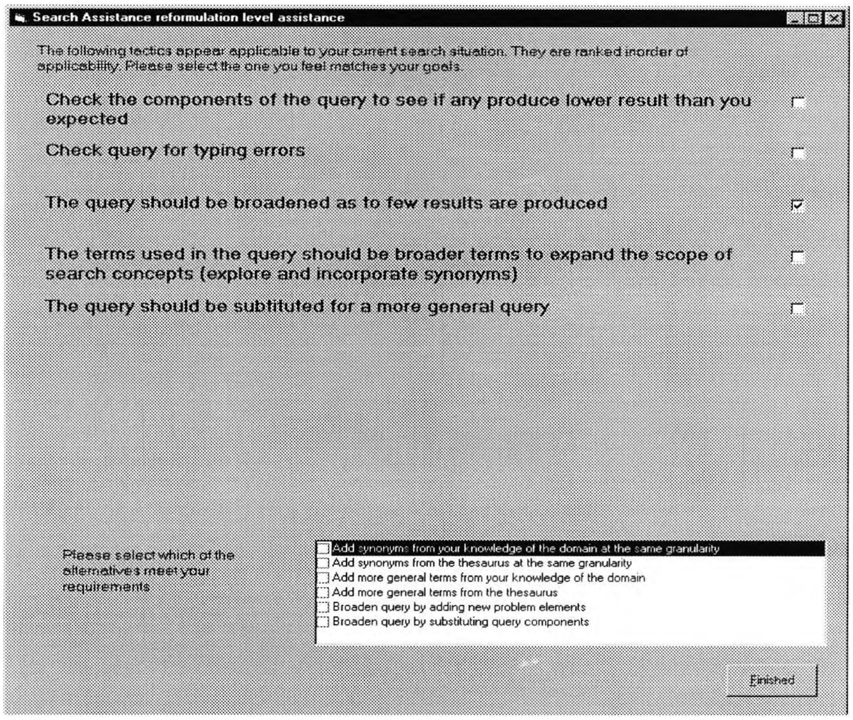

Figure 6.9 Strategy and tactic interface

Figure 6.10 shows an example of one of the interface configurations for reformulation dialogue. This interface is presented after the user has decided to add terms to an existing concept to broaden the results matched by this concept. Other reformulation situations enable concepts to be deleted, query relationships to be altered, terms to be deleted, new concepts to be added with the possibility of attaching synonym terms. In figure 6.10 the user must select the concept they require to alter and this will display a thesaurus screen (see figure 6.7) from which alterations or additions to the terms used can be made. Any changes automatically update the current representation of the information needed.

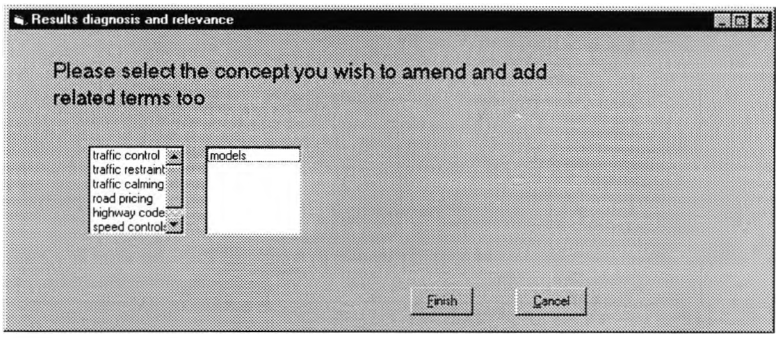

Figure 6.10 Initial interface enabling concept amendment and query reformulation

6.4 Usability sessions

Two usability sessions were conducted to enable different aspects of the application to be tested. The studies were conducted for two user group populations (Personnel of Liber a Swedish educational content provider. Users of a transport publication database produced by the company PTRC.). The sessions aimed to elicit feedback on whether the users felt the system would assist in the articulation of information needs and if it was felt that the dialogue and interaction structure promoted greater confidence in the results set produced.

6.4.1 Liber usability sessions

Twelve users participated in individual usability sessions. The format of the session involved a pre-session questionnaire aimed (appendix 6a) at eliciting the user's experience, a demo of the query advice tool, user verbal feedback and a post session questionnaire (appendix 6b). Each session took approximately one hour. All of the users had experienced problems searching for information on the WWW. Subjects were highly interested in obtaining help from systems in the reformulation of queries.

Task scenario

The scenario followed a script in the demonstration which presented the submission of a simple query without term alternatives. The query produced too few results and the system produced possible strategies for broadening the current query; and the option to broaden it by altering the terms used to articulate concepts was selected.

This presented multiple tactics for achieving this strategy from which the option to add alternative terms from the thesaurus was selected. Alternative terms were added to the current query and it was re-submitted to the database. Another concept was then added to the query to narrow the results but this included synonym terms and the query was re-submitted. The ease of altering the relationship between query elements was also demonstrated. The aim of the scenario was to introduce the subjects to the different information representations used in the concept demonstrator (query representation, thesaurus and term suggestion, strategy and tactic advice, results and query history) and to assess the usefulness of the advice given to assisting user retrieval. The subjects were publishers in the domain of education and not specialists in the domain of the target database.

Results

The results describe the level of user agreement with questions about the system and its facilities. The individual comments made by users about the system are then discussed.

Question	Average systems usability scale score
1) I thought the advice and assistance was represented in a useful and meaningful way	4.3
2) I found the interface tools effective for constructing and refining queries	3.8
3) I found the system unnecessarily complex	2.3
4) I though the system was easy to use	3.8
5) I would imagine that most people would learn to use this system quickly	3.9

Table 6.1 Average systems usability scale scores (copyright John Brooke digital corporation) from the post-session questionnaire on a scale ranging from strongly disagree (1) to strongly agree (5)

Table 6.1 highlights the level of agreement with statements made about the interface and the facilities provided. The system was generally well received with users finding the advice and assistance offered by the system to be meaningful and useful in the generation of queries. The number of facilities provided and the different nature of the interface from standard web based search engines created a possibility that the design would confuse users; but this was not the case as even though some of the subjects found the system complicated, in general subjects disagreed that it was too complex and felt that the system would be easy to use and require a minimum amount of time to become familiar with its functionality.

Key observations about the design and functionality of the system:

User recommendations have been clustered into four areas: the query representation, the thesaurus and synonym generation, the quality and clarity of the systems advice and general views about the application.

All subjects found the query representation to be a useful and clear way of viewing the information required. The representation was found to be easy to understand and some subjects highlighted that it provided good visibility of the entire query and appreciated not having to construct complex Boolean expressions using 'ANDs' and ORs'. One subject expressed the belief that system would be useful as an aid for people to know what they are looking for or as a teaching tool. However, searchers wanted some aesthetic changes to the Venn diagram view of the query and one subject felt that more feedback on the query representation was required.

Many users requested functionality to enable them to add their own synonyms, broader terms and narrower terms from domain knowledge or other information sources as an expansion of the thesaurus structure. The majority of subjects felt it was useful and helpful to be presented with synonyms and prompted to expand the terms used for concepts. The three categories of alternatives terms and the structure of terms were also seen to assist in decisions about which terms to include in queries. One subject commented that the thesaurus structure could be shown as a graphical presentation.

In general the reformulation adv ice was considered helpful and useful but it could be clearer. Users understood the advice and felt it would be good for novices but some of the subjects felt more illustrations and examples were needed to help digest the advice. A subject also felt that broadening advice could be useful in the beginning but perhaps not so much later on. Two subjects felt the system could take greater control with the advice being executed automatically. However, this requirement is neither realistic nor desirable as multiple strategies and tactics may be applicable in any given situation and only the end-user is able to distinguish which is the most appropriate; but having selected a tactic for query reformulation it is possible to automate query adaptation more. A subject commented that the advice helps in identifying mistakes; whereas another subject felt the advice was useful but it may be a bit too advanced with too many options for certain users.

The majority of user comments were positive but subjects felt the interface could be more interesting and tastefully designed. The system was seen as a far more comprehensive way of searching especially the support for narrowing or broadening, and many subjects stated they would use the application in preference to other web based search engines.

6.4.2 PTRC usability session

Ten users participated in the usability session. The format of the session involved a pre-session questionnaire aimed at eliciting the user's experience (appendix 6a), a slide presentation of the application and its interfaces, a demo of the application, a post session questionnaire (appendix 6c) and a focus group discussion about the interface; its applicability to users work contexts and the ease with which it was felt the system could be mastered. Additional questions to the Liber system usability score questionnaire attempted to assess the impact of this type of system on the 'natural' retrieval activity of the intended users for this specific application.

Task scenario

The scenario required the submission of a simple query without term alternatives. The query produced too few results and the system produced possible strategies for broadening the current query; and the option to broaden it by altering concepts in the query was selected. The scenario substituted the current concept for an alternative concept and this alternative concept had associated synonyms. This produced too many results and so the query was narrowed by removing some of the term alternatives. Another concept was then added to the query to narrow the results but this included synonym terms and the query was re-submitted. The ease of altering the inter-relationship for query elements was shown. The aim of the scenario was to introduce the subjects to the different information representations used in the concept demonstrator (query representation, thesaurus and term suggestion, strategy and tactic advice, results and query history), the application of different types of strategy, the different methods for amending query relationships and the dialogue necessary to produce a user profile. The subjects were specialists in the domain of the target database.

Results

The results presented describe the acceptance of the system and its facilities by the user population.

Question	Average systems usability scale score
1) I thought the advice and assistance was represented in a useful and meaningful way	4.4
2) I found the interface tools effective for constructing and refining queries	4.6
3) I found the system unnecessarily complex	2.2
4) I though the system was easy to use	4
5) I would imagine that most people would learn to use this system quickly	4.3
6) I thought the system restricted my natural approach to the retrieval of information and this made my retrieval less effective	2.4
7) I though the system may improve the effectiveness of my retrieval activity	3.9

Table 7.2 Average systems usability scale scores (copyright John Brooke digital corporation) from the post-session questionnaire on a scale ranging from strongly disagree (1) to strongly agree (5). On questions 3 and 6 for the system to be deemed successful users must produce answers at the opposite end of the SUS to the other questions. This is used to prevent users entering into a pattern of giving positive responses, and is thus an attempt to encourage users to think about individual questions and the systems usability scores.

The intended user population reacted even more positively to the application than in the Liber usability study and they also found the system advisor and the query interface to be useful, understandable and clear. The subjects found the interface easy to understand, felt it would be easy to learn and felt that it was not too complicated for use in usual work tasks. The subjects also felt this design and advice would assist and enhance their retrieval activity and not hinder or restrict their natural approach to the retrieval of information. The subjects believed the system would improve the effectiveness of the queries and strategies they would use.

Key observations about the design and functionality of the system:

As with the Liber usability studies, users recommendations have been clustered into four areas: the query representation, the thesaurus and synonym generation, the quality and clarity of the system's advice and general views about the application. Subjects in this usability study provided less variation in feedback about the design than subjects in the Liber study; but user comments about the system and its functionality were positive.

All subjects in the study liked the graphical query representation and found it useful. One subject commented on usefulness of including several terms for each concept. Two subjects felt the query representation was good but that they wouldn't use it; unfortunately they did not qualify their reasons for this.

All subjects found the suggestion of synonyms was useful and five subjects commented that the structure of term suggestions would assist in the determination of the ones to use. All subjects understood the different representations and a number of subjects stated that they would want to use the full release of the application.

Subjects found the advice useful but they also wanted to be able to restrict the results based on date, journal type and location of authors through simple switches on the main interface as opposed to in the queries submitted. Some subjects felt the advice may need to be accompanied with more explanation. Subjects also felt that the way advice effects the query and the results produced was clear.

6.5 Summary and conclusions

The application was well received by the user population. However, the application only addressed some of the design issues arising from the model due to implementation cost of certain features and the available usability study opportunities. The subjects felt the application was an improvement on the systems and interfaces available at present. In general subjects found the system easy to understand and felt the advice and information representations were clear. The majority of changes required to the interface were aesthetic rather than changes to the advice and functionality provided. The users did highlight a desire to add terms to the thesaurus structure as synonyms and to have the updated structure stored as a personal thesaurus. This was not expected but it is encouraging the subjects accepted the system and its functionality to an extent whereby they wished to customise the system based on their own domain knowledge. The acceptance of the advice and the belief this would be useful in the development of queries partially confirms the appropriateness of the advice, tactics and reformulation dialogue in assisting and enhancing retrieval. Aspects of the model relating to browsing and navigation based strategies still need to be tested.

Chapter 7

Conclusions and future work

This chapter summarises the thesis research and concludes with a discussion of possible future directions.

Chapter 7: Conclusions and future work

7.1 Introduction

This chapter summarises the work reported in this thesis and emphasises the benefit of the approach taken to improving the effectiveness of information retrieval systems design. This research proposes the use of cognitive task models of user behaviour as a means of enhancing users' retrieval activity and improving the design of systems for the different IR tasks with which users are faced. The advent of the World Wide Web and the increase in the numbers of users with regular requirements to use retrieval systems has challenged the traditional view of information retrieval being facilitated by an expert intermediary translating a users 'conceptual need' into a 'articulated need'. Therefore, the removal of this expert intermediary has produced a requirement for systems to try and replace the functions they used to offer. To achieve this IR systems designers must provide interfaces, search dialogues and guidance that support the search process. The deliverables of the research are:

- An empirical study of the search process with a typical system which demonstrated the need for improvements in retrieval systems interfaces (chapter three) assessing the effectiveness of the system in assisting and supporting user activity. The study identified strategies, activity focus (querying or evaluating), facility use, term usage in problem articulation, Boolean relationship use and system usability. The investigation considered these aspects of retrieval behaviour for device experts and novices on a variety of search tasks. This empirical work motivated the need to investigate the cognitive activities users perform in the retrieval process as a method of improving the match between users behaviour and system designs.
- A cognitive model of the search process is proposed to assist in the identification of the search activities, strategies and tactics applicable for different search contexts. The model dynamically predicts the strategies and tactics to be used in a given context based on the characteristics of the user, characteristics of the task and the retrieval system. The model includes a representation of the information retrieval system to be used and results encountered. The model indicated the

system functionality required to support user behaviour at different stages of the search process.

- The validity of this approach was demonstrated by a protocol study of the retrieval process which elicited the search activities performed during retrieval sessions and a computational implementation as an expert advisor. The protocol study is used to investigate the extent to which the model is a representation of behaviour. The protocol study only acts as a partial validation due to the inherent incomplete nature of verbal protocols and due to the limited categories of task and users examined.
- Attempts to improved the effectiveness of retrieval activity by encouraging 'good practice', and embedding strategy and tactical advice within a system was tested by implementing a expert retrieval advisor; and subsequent usability studies. The design and development of a query advisor was informed by the empirical studies and the cognitive model. User perceptions of this approach to enhancing retrieval behaviour and the provision of the appropriate TSFs to assist user retrieval were positive.

The thesis concludes with a review of future research directions.

7.2 A cognitive modelling approach

The research was motivated by the need to minimise the gap between users need's and the functionality provided by systems (Dervin 1977 and Borgman 1985). A cognitive view of information retrieval has already been shown to explain retrieval behaviour in a greater depth (Ingwersen 1996) than the previous observational modelling approaches (Bates 1989). It is for these reasons the research focused on the use of a cognitive view of IR to explain and predict behaviour; and determine the support systems must provide. The diversification of the user population and the types of tasks performed has prompted a need to produce more user centred information retrieval systems which support the activity performed in the retrieval process.

In the following section, problems with existing approaches to modelling user retrieval behaviour and empirical work addressing the different factors affecting users' retrieval behaviour introduced in chapter two are reviewed, then the role of cognitive models in overcoming these problems are discussed.

7.2.1. Problems of existing approaches

Existing information retrieval systems provide powerful mechanisms with which an experienced user can locate information but current designs are far from optimal. Furthermore the diversification of the user population, the use of IR systems in a greater variety of tasks and increases in the size of information repositories necessitates improvement in the designs of the information systems if they are to support end-users. To support user activity the design of IR systems must attempt to support the activities they need to perform, and guide and assist their behaviour in strategy and tactic choices. To achieve this aim it is necessary to have a more complete model of both expert and novice behaviour. Without a cognitive model of the information searching task, decision over the design of task support functionality's will be based on inspired guesswork and personal experience. No thorough task analysis of information searching has been undertaken prior to this thesis. While a standard task analysis might have produced a description of user behaviour it could not predict behaviour for different types of user in a variety of contexts, hence a theoretically based cognitive task model was developed.

There is a long tradition of modelling user behaviour in the search process (Markey and Atherton 1978, Bates 1979; 1989, Ingwersen 1982; 1996, Borgman 1985, Belkin 1982; 1993, Kulthau 1983, Marchionnini 1995). Unfortunately models of user behaviour fail to include the effects of aspects of the design and the results produced on searcher behaviour. The requirement for more complete models of user behaviour and the view of IR as dynamic and adaptive has recently been addressed by Ingwersen's (1996) model in which information needs effect the nature of the strategies used. However the effects of facilities provided by IR systems and the quality of results on the strategies and tactics applicable have not yet been incorporated into models of the retrieval process. Much research has been conducted into retrieval activity observed for specific contexts; but this research needs to be synergised into a more complete model of the retrieval process applicable for a wider variety of contexts. The cognitive model presented in chapter four uses the results of previous research in the field in conjunction with representations of information needs, the results produced and a model of the IR system to identify the contexts in which particular strategies and tactics are most applicable. The activities, strategies and tactics are used to specify how user guidance can be targeted and how search dialogues can be structured to encourage efficient and effective retrieval.

The cognitive model presented in chapter 4 incorporates many aspects of previous models of the retrieval process. The dynamic and iterative view of the IR process put forward by Markey and Atherton (1978) in which a previous query effects the subsequent queries is included in the proposed model. The model uses search tactics to enable the adjustment of queries, as proposed by Bates (1979) and Harter et al (1985). The tactics are used within the model as approaches to advise users of appropriate search tactics in different results situations. In the research of Bates and Harter et al, the tactics are given as a set of opportunities for effective query modification but the research does not explore how or when in the search process they are applicable; or for which types of tasks certain tactics are more suited. The tactics offer a summary of expert knowledge of best practise but end-users are rarely aware of these tactics and favour less optimal approaches thus use alternative approaches e.g. long evaluation times. The search tactics need to be placed in the context of a model of actual retrieval behaviour and expertise must be delivered to users with tools to apply the advice. The model proposed in chapter 4 enables the appropriate advice to be delivered at each stage of the task.

Ingwersen's (1982) communication models of information retrieval highlight the formal steps expert intermediaries take during the elicitation of information needs from an end user. The model highlights the process of need translation from the 'conceptual need' to the 'expressed need'. Ingwersen's model of the retrieval is based on the expertise of librarians and does not include a system model or discuss the effects of the search results on the IR process. However these aspects of retrieval are necessary in an explanation of the retrieval behaviour of end-users searching as the judgements of the quality of the results effect the subsequent direction of the search. The decomposition of the retrieval task into sub-steps is an aspect of Ingwersens research included within the model presented in chapter 4, but the actual steps are different as his model is based on modelling the communication between the end-user and the expert intermediary and the focus of the model proposed is end-user behaviour.

Belkin (1982) proposed the view that users are not always faced with tightly specified and complete information needs and thus users possess an anomalous state of knowledge when understanding the search problem and it indicates that needs change during the search. This view of variations in specificity and complexity for user tasks are incorporated into the model proposed in chapter 4 as part of the information need type classification.

Bate's (1989) highlights a retrieval strategy of using multiple searches to locate information with users acquiring results from different searches as part of the solution to the information need. This strategy is included within the model presented in chapter 4 through the use of multiple searches and sub-strategies in the satisfaction of an information need. This strategy is particularly useful if the user is faced with complex and complicated information needs.

Kulthau (1983) breaks the retrieval process down into a number of observable stages (Information Seeking Process) and recognises the iterative and none linear nature as a transition between these stages. However this model does not explore the situations in which the different stages are applicable or the cognitive activity underlying these observable actions. The decomposition and iterative nature of the activities within the retrieval process are included in the model proposed in chapter 4, but in addition to this the strategies and tactics resulting from the different activities are added and the model predicts the activities underlying these search steps. The proposed model also includes representations of the system, the results, and indicates their inter-relationship with search processes and the structure of user behaviour. Kulthau includes the extraction of terms within her stages and this aspect of retrieval behaviour is not included within the proposed model as the rate and terms extracted requires detailed understanding of the results semantics to be predicted.

Marchionini's (1995) model of electronic information seeking defines a number of steps in the information seeking process and makes distinct the inter-relationships between the different activities. However his model does not consider the effects of different strategy and information need types, or user knowledge, on the activities and transitions occurring in the search process. Furthermore it does not consider the effects of the system or the results retrieved on users' activity.

Ingwersen (1996) discusses the elements of a cognitive theory of IR and describes the search process as a communication between the users model of the world and the systems model of the world. His discussion does not explore the complex interactions between these two models, and does not address the effects of these models on the activities a user performs in translating between the two information models. Ingwersen's model does, however, indicate the link between different types of information needs (Well defined - ill defined, stable - variable) and strategies (search loops- berry picking), and advances the understanding of IR behaviour through its dynamic view of information needs, problem states, knowledge representations and that multiple context variables influence activity. The limited scope of the strategies linked to information needs requires further definition if it is to form the basis of user assistance.

The model presented in chapter 4 contains many of the aspects of previous models and previous research has contributed to the models components and mechanisms that allow it to react to different search contexts. In addition the model includes representations of the systems facilities and the results retrieved by the search. The model identifies the activities which may occur based on the quality of the results and includes strategy and tactic selection rules for different results situations and information need types extending Ingwersens' (1996) loose association between need types and strategy. This extension of need types allows the expansion of the number of strategies explained by the model. The model also highlights that many strategies and tactics are applicable in different search contexts and indicates the effects of searcher knowledge on behaviour and the applicable strategies. The model proposed considers retrieval activity in terms of processes, activities and rules dictating the appropriate strategies and tactics thus implementing the associations highlighted by empirical studies (Efthaniades 1993, Allen 1994) between the results retrieved and user strategies and tactics.

7.2.2 Contribution to HCI Theory and Cognitive Modelling

The model proposed in this thesis predicts user behaviour with a range of tasks that may be broadly described as information searching. One view of the model is a cognitive task model, in that it describes and predicts human mental as well as physical behaviour. It therefore represents an intermediate position between a standard, descriptive task model, such as might be produced by Task Knowledge Structures (Johnson-laird et al 1991) and an general model of human cognition (e.g. Interacting Cognitive sub-systems Barnard 1991). Barnard (1988) provides a method of describing user behaviour within the interacting cognitive sub-systems architecture. This enables a description of tasks and predicts human behaviour at a low level of detail but is not suitable for modelling more complex and flexible tasks. While ICS can predict usability problems and errors that are caused by general resource contention problems in human information processing, e.g. selective attention to active windows in user interface; it can not predict failure in larger scale tasks unless it is augmented with detailed additional knowledge bases describing the users task and domain knowledge (Wilson et al 1988). These domain models may be difficult and expensive to acquire for IR due to the variety of situations and types of task encountered; and the models may be dynamic within the retrieval process based on the results retrieved. The cognitive task of information seeking, on the other hand, can make wider ranging predictions within the domain of information retrieval, although it does not explicitly account for cognitive phenomena such as working memory or selective attention. The cognitive task model may benefit from integration with cognitive principle models of architecture (e.g. Barnard 1991) or learning (e.g. ACT * Anderson 1991). Other task description methods such as GOMS (Card et al 1983), CCT (Kieras and Polson 1985) offer similar low level descriptions of tasks but these are poor at describing dynamic and reactive tasks in which the systems responses and dialogue is essential in determining the course of activity. Norman's (1986) model of action provides descriptions of the cycles of action and evaluation observed in interaction, however a greater level of detail is required to model the IR task; and strategies and tactics also impact on the course of interaction. The proposed model fits between these different levels of modelling as it tries to describe behaviour at the action level but also explain and predict this behaviour given the underlying search context using rules and generalisations about the user and their environment. The proposed model thus explains activity and strategy without using detailed domain knowledge representations to determine the course of user behaviour.

The increasing variety of innovative search functionality and information presentation facilities provided by IR researchers (e.g. probabilistic retrieval Robertson 1994) and user interface design (e.g. starfield displays and dynamic query filters Alhberg and Shneiderman 1994, hyperbolic browser Lamping et al 1994, 1995, hit density maps Moran 1995, query by sketching Charles et al 1990, query by pointing Zloof 1978) raises the issue of how to combine these different task support facilities into configurations for effective and efficient retrieval. The advances have not been accompanied by a consideration of the applicability of these facilities within the context of different retrieval tasks e.g. browsing and/or searching. Each facility may solve a specific user interaction problem; but this leads to the need to investigate the suitability and scaleability of novel interfaces within a larger view of the IR task. The optimal design may be a combination of many of these novel interaction techniques and system functionality each supporting different aspects of the retrieval task; hence it is important to understand when and in what search contexts these facilities offer added benefit to user retrieval. In chapter four the model identified the different task contexts in which the facilities may be used. As such it is possible to highlight the different search situations and tasks for which particular facilities are needed to support the users activity. One possible future application of the model is to search service configuration on the Web. Users will require tailored set of functions to suit their search needs, however, search engines are increasing their range of functions, thus leaving users confused by excess choice and little advice, a development also observed in word processing systems (Fischer 1996). The model can help designers develop configuration facilities that might help users set up their own profile and need, then the system could download an appropriate set of facilities form a server and configure them on the user workstation.

Chapters three and five demonstrated that expert and novices require differing systems support based on the observation of their behaviour and search strategies, and thus retrieval systems must take this into consideration in the advice and interface support provided. Novices have knowledge of fewer strategies and tactics for the different search situations, and only use a small number of strategy and tactics due to their low IR strategy knowledge. Novice searchers encountered difficulties expressing Boolean inter-relationships in agreement with previous findings (Sewell et al 1986 and Marchionini 1989). In chapter 3 it was found that users follow multiple strategies during retrieval behaviour and that novices are able to compensate for their lack of knowledge by evaluating a greater number of articles. Experts and novices show different strategies and patterns of activities, as novices use simple queries and spend more time evaluating the results whereas experts submit more complex queries and strategies such as narrowing and broadening. Experts explored terms for more concepts and use more terms in queries than novices. If a searchers' retrieval included characteristics of good query construction and need articulation coupled with careful evaluation then in general recall was improved; but the choice of specific terms can have dramatic effects. Semantic distance between a searcher's articulation of information needs and the search system index or document description, is one of the perennial problems of information retrieval (Brook 1995, Ingwersen 1996). Thus assisting searchers in strategy and tactic choices will only go part of the way to improving the performance and effectiveness of users' retrieval behaviour. The lack of correlation between effective strategies and search success has been noted in ecological studies of on-line searches with intermediaries (Smithson 1994). In chapter 3 users highlighted a desire for assistance in query generation and need articulation; and thus research must determine the type of assistance to provide and the context in which this assists retrieval. It is apparent from the empirical studies and the cognitive model that novice searchers require substantial and targeted systems support, which they are currently not receiving from commercial systems. Improving system functionality and interfaces will assist users with the appropriate IR knowledge have more effective retrieval; however, the average user does not possess this knowledge, so retrieval systems must encode this expertise as intelligent advisors.

The validation of the model in chapter 5 confirmed differences in the strategic approaches used by experts and novices with experts favouring pre-search problem development and planning while novices favoured evaluation. The lack of substantial differences in recall performance suggests that different strategic approaches can lead to relatively similar performance success; however, overall performance was poor. The protocol study indicated that the different reformulation outcomes expected and the patterns between results evaluation (at its different levels) and query revision, strategy revision or problem development agreed with the model representation. The study showed the number of alternative activity transitions experts perform to be greater than novices indicating the effects of device and IR knowledge on the courses of action they pursued. The sequences of novice retrieval behaviour were simpler than experts, and this matched the view of retrieval given in the cognitive model in which the options available increases with expertise. The regular sequences of behaviour indicate users apply a method in retrieval activity so the expected sequences of user activity may be supported by system facilities matched to probable task stages. The inter-group differences confirm the necessity for models of expertise and novice models. The empirical study showed experts to perform 50% more new query formulations and 35% more query revisions than novices. The validation confirmed the effects of skill on the rate of activity completion (Rassmussen 1993, Solomon 1993) irrespective of the task. Task complexity reflected in the information need also effected user search behaviour as found in previous studies (Borgman 1985, Large et al 1994), and these effects were significantly different between experts and novices. The poor performance observed on task PH2 may have been caused by motivational effects as users appeared to give up finding this task too difficult.

Alternative methods of providing expert IR systems as computer intermediaries in the search process have been proposed and implemented by Marcus 1983, Vickery et al 1987, Croft 1989. Each of these systems followed different approaches to providing assistance, different levels of automation of query adaptation and achieved varying levels of success. The CONNIT system (Marcus 1983) enabled the formulation of queries to multiple distributed databases through a common interface. The expert system converting the user input into the language of the specific retrieval system to be searched. The experiments conducted compared the performance of the system over that in which a human intermediary was used and found that although recall was improved this was at the expense of longer retrieval sessions. Vickery et al's (1987) PLEXUS generates a formal problem statement from an informal user input. The success of the system depends on the quality of the systems domain representation and so the generality and scaleability of this approach to larger domains than gardening needs to be investigated as domain models may become intractably complex. Furthermore the domain bound approach inevitably suffers from a knowledge acquisition bottleneck, therefore more effort has to be expended in developing the expert advisor as the size and diversity of the database increase. The system is required to make decisions on the quality of the results in relation to the query; however it is likely that the systems success in this will be limited as the process of relevance assessment (Saracevic et al 1990, Su 1994) is a complicated issue which are difficult to unpack. Crofts' (1989) I^3R system used powerful statistical retrieval techniques and browsable displays with expert systems techniques based on condition and action rules. The system experts are used to assist query formulation and reformulation, search and user evaluation. However even though performance was good it required significant computational resources and can place more demands on the user than conventional systems (Croft 1989). The effectiveness of such query refiners depend on the quality of the domain knowledge base which can be expensive and difficult to acquire. The expert intermediary approach to supporting user strategy and tactic choices for different retrieval scenarios has been demonstrated to improve the quality and richness of the queries articulated but normally at the expense of user effort. Marcus (1983) concluded that to achieve more effective guidance, search advisors must be based on more comprehensive models of expertise and the search process. The expert advisor described in chapter 6 differs from other expert systems intermediaries in that it does not operate from explicit models of the domain and it always leaves control over search direction with the end-user. The advisor provides advice and guidance which is sensitive to the users' knowledge, task and current retrieval situation as it provides options for developing the query. The system guides the query reformulation based on the strategy and tactic options chosen by the user and will advise of possible terms to include but not automatically select them. The advisor incorporates the users' term choices within the current query and automatically generates the correct Boolean syntax for their inclusion in the query.

7.2.2 Benefit of the approach to IR system design

The benefit of a more detailed understanding of user behaviour should help to improve user confidence that retrieval systems can help them to access a high percentage of the information within the database. The comments of subjects presented in chapter three indicated a lack of confidence that the majority of the relevant information in the database had been retrieved. The subjects in the usability study of the exemplar application (chapter six) reported that they had low confidence in the facilities of some web based retrieval systems (Alta-vista was the example given by user). The usability studies for the expert advisor also indicated that providing targeted advice to support user activities can go part of the way to improving user confidence in retrieval systems, as the subjects felt that this type of IR system design would improve the effectiveness and efficiency of their retrieval behaviour.

It may not be acceptable to require users to explicitly acquire knowledge of the strategies and tactics of efficient and effective retrieval; especially for discretionary and occasional users. The system must take a greater role in structuring retrieval behaviour and assisting searchers make the correct strategy and tactic choices. This is likely to have the additional benefit of educating novices in more effective methods of locating the information required efficiently and effectively. However, even experts require advice as a aide memoir in the query formulation process as they sometimes neglect to pursue correct and fruitful search directions (Martin 1973, Oldroyd and Citroen 1977). It is important systems leave the user in control of the direction in the retrieval process as they may not accept the results retrieved unless they understand how query articulation and results are related.

The expert advisor described in chapter six demonstrated how targeted support can be provided. However there is a cost in providing this targeted advice; as users have to answer questions about their knowledge, information needs, expectations and the results produced. This is necessary because automatically making these judgements requires extensive domain knowledge and artificial intelligence. The expert advisor also encouraged 'good retrieval practise' and error checking through the user-system dialogue and information representations used. The usability studies performed on this system have produced encouraging results with subjects giving a positive response to this type of system initiation in structuring and assisting retrieval behaviour.

7.3 Limitations of the cognitive modelling approach to information retrieval system design

The use of cognitive modelling in retrieval systems design and the provision of targeted advice addressed only some of the issues to be considered for improving the effectiveness of retrieval behaviour. The model was based on general need types and user types but in the future it may be necessary to extend these classifications to improve the specificity of the advice given. The model is able to offer advice but requires the user to provide relevance judgements on the results. As IR research produces a greater understanding of relevance, and user's attribution of this, it may be possible to make some of these judgements on the users' behalf or assisting the user come to these judgements thorough more comprehensive result and query representations.

The model may benefit from the use of more complicated TSF models by making explicit associations between different information representations and their effectiveness for improving user searches. The extension of TSF representations from a functional view of facilities to include a richer representation of the methods used to present information could enable evaluation of different metaphors (e.g. fisheye, cone tree) success in supporting the retrieval process. The model specifies the activities (especially diagnosis, evaluation and query adaptation) users must perform; and from this the effectiveness of functionality and representation provided to support the user is predicted. Further research is necessary to establish the effect of different information visualisations on the effectiveness of user searching and methods of browsing. The model is, at the moment, limited in this respect as it can only indicate a browsing activity is necessary but it can not state which of the variety of methods for visualising information is the most appropriate, or effective, in a specific retrieval situation.

7.4 Summary

To summarise, cognitive models can be an effective way of targeting the design of information retrieval systems so they meet users task requirements, and assist in the supply of context sensitive information to users as they progress through a retrieval session. The thesis has explored some of the issues that are necessary if the effectiveness of information retrieval systems in meeting user requirements are to be improved. The model has helped to clarify the situations when specific sets of information retrieval facilities should be deployed to support user behaviour. The thesis has drawn on research from many sources for the modelling of IR and user interaction, and has argued the need to amalgamate views of IR into a more holistic view of the retrieval process so behaviour in different contexts can be predicted. A novel approach to providing support for user behaviour using expert advisors based on generalisations about the user and the systems results has been described; and it appears that users found this to be useful in improving the effectiveness of their retrieval. The empirical studies conducted have confirmed the findings of other researchers in relation to the query syntax used and the effects of term choice as well as strategy on the effectiveness of retrieval. The thesis has also added to the understanding of the IR task in terms of the different strategies and patterns of user activity. Promoting good practice and guiding users during retrieval were considered by users to be effective and acceptable ways of improving their behaviour. There are limitations to this approach as only general rather than domain specific advice can be given unless substantial domain knowledge is incorporated in systems. The use of domain knowledge is an alternative to the approach followed in this thesis; but one which can not be justified at present owing to the problems of constructing and maintaining domain knowledge base for information retrieval systems that cover large and ‘real’ retrieval systems domains. The model is to be used in IR systems design to inform the facilities required to support the users retrieval task. The following steps indicate how the model is run and used in design:

1. Input a need and user type
2. Model produces a list of appropriate strategies
3. One of these strategies is selected
4. The information need is developed given the strategy selected
5. The users activities, strategies and needs are matched against the system model (TSF->TSR)
6. A description of the results is input into the model
7. The model selects a results set diagnosis
8. The model produces a list of appropriate reformulation tactics based on the diagnosis of the results set
9. etc. until cut-off reached (eg. Time, motivation etc.)

7.5 Future work

It is necessary to update the prototype IR advisor tool based on the changes highlighted during the usability studies; and the system requires further validation. An empirical study comparable to that presented in chapters three and five must be conducted to determine the success of the system in supporting different users for different task situations. The study should concentrate on improvements in the effectiveness and performance of users with retrieval systems.

Further developments of the cognitive model are required to extend the classification of information needs. The model also requires a stronger definition of the effects of browsing on the behaviour produced. As part of this model development, further empirical studies are required on a greater diversity of task types and information need categories.

The eventual aim of this work is to embed the automatic customisation of the facilities incorporated into a system based on the user's profile. The extension of this is to allow the user to specify the services they require based on their information needs;

and then the system will combine these facilities into an application rather than just providing context sensitive advice and dialogue to promote good practise.

However a number of different task support facilities may support the same user requirements and a personal choice may be more appropriate in some cases. For example one user may prefer a hyperbolic browser to explore a thesaurus over a tree based structure and either could support the same task resource requirement as far as the model is concerned; so individual user preference should dictate which of the representations to use. The model would only indicate that the facility is required.

References

Ahlberg C. and Shneiderman B,, (1994), Visual information seeking: Tight coupling of Dynamic query filters with starfield displays, CHI 94 Boston

Allen B.L. (1991), Cognitive research in information science: implications for design, *Annual review of information science and technology*, Vol. 26, pp. 3-37.

Allen B.L., (1994), Perceptual speed, learning and information retrieval performance. In Proceedings of 17th SIGIR Conference, Dublin, ACM Press, pp. 71-80.

Anderson J.R (1990), The adaptive character of thought. Lawernce Erlbaum, Hillsdale, NJ.

Apted S.M., (1971), General purposive browsing, *Library Association Record*, Vol 73, No.12, pp 228-230

Bakeman R. and Gottman J.M., (1986), Observing interaction: an introduction to sequential analysis, Cambridge, UK: Cambridge University Press.

Barnard,P. Wilson,M. Maclean,A. (1988), Approximate modelling of cognitive activity with an expert system: A theory based strategy for developing an Interactive design tool, *Computer Journal*, 31(5), pp 445-456

Barnard P., (1991), Bridging between basic theories and the artefacts of Human Computer Interaction. In Designing Interaction, Psychology at the Human Computer Interface, Ed Carroll J.M. Cambridge University Press.

Barry C.L., (1994), User defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, Vol. 45(3), 149-159.

Bassili J.N and Regan D.T., (1977), Attributional focus as a determinant of information selection, *Journal of Social Psychology,* Vol. 101, pp. 113-121

Bates M.J., (1979), Information search tactics, *Journal of the American Society for Information Science*, Vol. 30, No. 4, pp. 205-214.

Bates M.J., (1989), The design of browsing and berrypicking techniques for the on-line interface, *On-line Review*, Vol. 13, No. 5, pp. 407-424.

Bates M.J., (1985), Idea tactics, *Journal of the American Society for Information Science*, Vol. 30, No. 5, pp. 280-289.

Beaulieu M.,(1997), Experiments on interfaces to support query expansion, *Journal of documentation*, Vol. 53, No. 1 pp 8

Belkin N.; Cool C., Stein A. and Thiel U., (1993), Scripts for information seeking strategies, In: Case-Based Reasoning and Information Retrieval - Exploring the

Opportunities for Technology Sharing, Papers from the AAAI Spring Symposium Series, Technical Report SS-93-07, pp. 8-17.

Belkin N.J., Oddy R.N. Brooks H.M.,(1982), Ask for information retrieval: Part I, background theory, *Journal of Documentation*, vol. 38(2).

Belkin N.J. (1987), Information concepts for information science. Journal of Documentation, 34, pp. 55-85.

Bellotti V., (1993) Integrating theoreticians and practitioners' perspectives with design rationale. In Proceedings of INTERCHI-93, Ed Ashlund S., Mullet K., Henderson A., Hollnagel E. and White T., ACM Press, pp 101-106.

Benyon D. and Hook K., (1997), Navigation in information spaces: supporting the individual, In Hunam Computer Interaction: Proceedings of Interact '97, ed. By S. Howard, J. Hammond and G. Lindagaard, London: Chapman and Hall.

Borgman C.L., (1985), The users mental model of an information retrieval system: an experiment on a prototype on-line catalogue, *International Journal of Man-Machine Studies*, Vol. 24, pp. 47-64.

Brindle E, (1981), The relationship between characteristics of searchers and their behaviour while using online interactive retrieval systems, PhD Dissertation Syracuse, NY, Syracuse University

Brooks T.A. (1995) Topical subject expertise and the semantic distance model of relevance assessment. *Journal of Documentation,* Vol. 51, No. 4, pp. 370-387

Bystrom K. and Jarvelin K. (1995). Task complexity affects information seeking and use. Information Processing and Management, Vol. 31, No. 2, pp. 191-213

Card S., Moran T, and Newell A., (1983), The psychology of human computer interaction, Hillsdale, NJ, Erlbaum

Card S., Robertson G., Mackinlay J., (1991), The information visualiser, an information workspace, In proceedings of CHI '91, New Orleans, April 28-May 2, 1991, pp. 181-188, New York: ACM

Carrol, J.M. (1990). The Nurnberg Funnel: Designing minimalist instruction for practical computer skill. Cambridge, MA: MIT Press.

Carroll J.M., Kellogg W.A. and Rosson M.B., (1991), The task-artefact cycle, In Designing Interaction, Psychology at the Human Computer Interface, Ed Carroll J.M. Cambridge University Press.

Charles S and Scrivener S., (1990), Using queries to search pictorial databases, Human computer interaction-Interact '90, pp. 493-498.

Croft W.B., (1987), Approaches to intelligent information retrieval, *Information Processing and Management*, Vol. 23(4), pp. 249-254.

Cugini J, Laskowski S and Piatko C, (1997) Document clustering in Concept Space: The NIST information retrieval visualisation engine (NIRVE), CODATA Euro-American workshop Visualisation of information and data: Where are we and where do we go from here?, 24-25 June, Paris.

Danilowicz C, (1994), Modelling user preferences and needs in Boolean retrieval systems, *Information processing and management*, Vol. 30, No. 3, pp. 363-378.

Davidson D., (1977), The effect of individual difference of cognitive style on judgements of document relevance, *Journal of the American Society for Information Science*, Vol. 28(5), pp. 273-284.

Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R.A. (1990), Indexing by latent semantic analysis., *Journal of the Society for Information Science*, Vol. 41, No. 6, pp.391-407.

Derr R., (1983), A conceptual analysis of information needs, *Information processing and management*, Vol 19, No. 5, pp 273-278.

Dervin B, (1977), Useful theory for librarianship: Communication, not information, *Drexel Library Quarterly,* Vol. 13, No. 3, pp. 16-32

Efthamiades E.,(1993), A user centred evaluation of ranking algorithms for interactive query expansion, 16th ACM-SIGIR Conferences Proceedings, Pittsburg, pp. 146-159.

Elkerton J. and Williges R.C., (1984), Information retrieval strategies in a file search environment, *Human factors,* Vol. 26, No. 2, pp. 171-184

Ellis D, (1992), The physical and cognitive paradigms in information retrieval, *Journal of Documentation*, Vol. 49, No. 1, pp. 54-64.

Elvins T, Nadeau D., Schul R. and Kirsh D., (1998) Wordlets: 3D Thumbnails for 3D Browsing, CHI '98, pp. 163-170.

Ericsson K and Simon H, (1984), Protocol analysis: Verbal reports as data, *MIT press*

Eysenck M.W. and Keane M.T., (1990), Cognitive psychology, LEA Publishers.

Faulkner C., (1998), The essence of Human-computer Interaction, Prentice-Hall.

Fidel R., (1991), Searchers' selection of search keys: I the selection. II Controlled vocabulary or free text searching, III Searching styles, *Journal of the American Society for Information Science*, Vol. 42, No. 7, pp. 490-527

Fischer G., (1993) Beyond Human Computer Interaction: Designing Useful and Usable Computational Environments Invited Papers Proceedings of the HCI'93 Conference on People and Computers VIII 1993 p.17-31

Fishkin K and Stone M, (1995), Enhanced dynamic queries via movable filters, CHI '95, pp. 415-420

Furnas G. W. and Bederson B., (1995), Space-scale diagrams: understanding multiscale interfaces, CHI '95, pp.234-241.

Furnas G.W., (1997), Effective view navigation, In proceedings of CHI 97 Atlanta.

Hancock-Beaulieu M., (1992), Query expansion: advances in research in on-line catalogues, *Journal of Information Science*, Vol. 18, pp 99-103.

Hancock-Beaulieu, M, Fieldhouse, M and Do, T, (1995), An evaluation of interactive query expansion in an on-line library catalogue with a graphical user interface, *Journal of documentation*, Vol. 51 , No. 3, pp 225-243.

Harter S.P. & Rogers-Peters, A, (1985), Heuristics for online information retrieval: A typology and preliminary listing, *Online Review*, 9(5) pp. 407-424.

Herner S., (1970), Browsing, In Encyclopaedia of Library and Information Science, Vol 2, pp. 408-415, New York: Dekker.

Hiltz S. and Turoff M., (1985), Structuring Computer-Mediated Communication Systems to Avoid Information Overload, Comm. ACM, vol 28 no. 7, pp. 680-689, July, 1985

Iivonen M, (1995), Searchers and searchers: differences between the most and least consistent searchers. In Proceedings of 18th ACM-SIGIR conference, Seattle, ACM Press.

Ingwersen P., (1982), Search procedures in the library- analysed from the cognitive point of view, *Journal of Documentation*, Vol. 38, No.3, pp. 165-191

Ingwersen P, (1993), The cognitive viewpoint in IR, *Journal of Documentation*, Vol. 48, No. 1, pp. 60-64.

Ingwersen P., (1996), Cognitive perspectives of Information retrieval. Interaction elements of a cognitive IR theory, *Journal of Documentation*, Vol. 52, No.1, pp. 3-50.

Irwin F.W. and Smith W.A.S, (1957), Value, cost and information as determiners of decisions, *Journal of Experimental Psychology*, Vol. 54, No.3, pp. 229-232

Jacobson T. and D. Fusani, (1992), Computer, system and subject knowledge in novice searching of a full text multifile database, *Library and Information Science Research*, Vol. 14, No. 1, pp. 97-106.

John B.E. and Kieras D.E. , (1996), Using GOMS for user interface design, Which Technique. ACM Transactions of Computer Human Interaction, Vol. 3(4), pp. 287-320.

Kieras D.E., Wood S.D. and Meyer D.E., (1995), Predictive engineering models using the EPIC architecture for a high performance task. In Proceedings of CHI-95, Human Factors in Computing Systems, ACM Press.

Kitajima M & Polson P, (1996), A comprehension based model of exploration, Common ground CHI 96, Vancouver, pp. 324-331

Kuhlthau C., (1988), Longitudinal case studies of the information search process of users in libraries, *Library and Information Science Research*, Vol. 10, No. 3 pp 257-304

Kulthau C., (1991), Inside the search process: information seeking from the users perspective, *Journal of the American Society for Information Science*, Vol 42, pp.361-371.

Kuhlthau C, (1993), Seeking Meaning, New York

Kwasnik B.H., (1992), A descriptive study of the functional components of browsing. In proceedings of the IFIP TC2 G2.7 Working Conference on Engineering for Human-Computer Interaction, pp. 191-202

Lancaster F.W., (1972), Evaluation of on-line searching in MEDLARS by biomedical practitioners, Urbana, University of Illinois, Graduate school of Library science (Occasional paper No. 101)

Lancaster F.W. (1977), Towards paper-less information systems, New York: Academic Press

Lamping J, Rao R and Pirolli P., (1995), A focus+context technique based on hyperbolic geometry for visualising large hierarchies, CHI '95, pp. 401-408

Large A., Besheshti J., Breuleux A. and Renaud A., (1994), A comparison of information retrieval from print and CD-ROM versions of an encyclopaedia by elementary school students, *Information Processing and Management*, Vol. 30, No. 4, pp. 499-513.

Lewis, C. Polson, P. Wharton C and Reiman J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces, in J. R. Chew and J. Whiteside (eds.), Proceedings CHI-90. ACM Press, 235-241.

Magennis M and Van Rijsbergen C.J, The potential and actual effectiveness of interactive query expansion, In proceedings of SIGIR 97, Philadelphia PA, USA, pp 324-331

Marchionini G., (1989), Information seeking strategies of novices using a full text electronic encyclopaedia, *Journal of the American Society for Information Science*, Vol 29, No. 3, pp 165-176

Marchionini G. and Liebscher P., (1991), Performance in electronic encyclopaedias: Implications for adaptive systems, Proceedings of the 54th Annual meeting of the American Society for Information Science, Vol. 28, pp. 39-48

Marchionini G., (1995), Information Seeking in Electronic Environments, Cambridge University Press.

Marcus R., (1983), An experimental comparison of the effectiveness of computers and humans as search intermediaries, *Journal of American Society for Information Science*, Vol. 34, No. 6, pp. 381-404.

Markey K. and Atherton P., (1978), ONTAP: Online training and practice manual for ERIC database, Syracuse University, ERIC Clearing House on Information Research, Syracuse, N.Y. , 1978 (ED106109).

Martin W., (1973), A comparative study of terminal user techniques in four European countries on a large common online interactive information retrieval system. First European congress on documentation systems and networks, pp. 107-167.

Maybury . M.T. (1990), Custom explanation: Exploiting user models to plan multi-sentential text. In proceedings of 2nd International workshop on User Modelling, Hawaii, USA.

Michel D.A., (1994), What is used during cognitive processing in information retrieval and library searching? Eleven sources of search information, *Journal of American Society for Information Science*, Vol.. 45, No. 7, pp. 498-514.

Neilsen J, (1993), Usability Engineering, Academic Press.

Neisser U., (1966), Cognitive psychology, New York: Appleton-Century-Crofts

Newell A. and Simon H., (1972), Human problem solving, Englewood Cliffs, NJ: Prentice-Hall.

Norman D.A. (1986). Cognitive engineering. In: User centred system design: new perspectives on human-computer interaction, ed. by D.A. Norman and S.W. Draper. (Hillsdale, New Jersey: Lawrence Erlbaum Associates).

Oldroyd and Citroen, (1977), Study of strategies used in online searching, *Online review*, Vol 1, pp. 295-310

Payne S.J., (1984), Task action grammar, In Shackel B. (ed), Proceedings of Interact '84, pp. 139-144

Payne S.J. and Green T.R.G, (1986), Task action grammar: A model of the mental representation of task languages, *Human Computer Interaction*, Vol. 2, pp. 93-133

Pirolli P, Schanck P, Hearst M and Diehl C, (1996) Scatter/ Gather Browsing communicates the topic structure of a very large text collection, CHI '96,Vancouver, Canada.

Quarterman J., (1990), The Matrix Computer Networks and Conferencing Systems Worldwide, Digital Press, Bedford, Massachusetts.

Raskutti B and Zukerman I, (1997), Generating queries and replies during information seeking interactions, *International Journal of Human Computer studies*, Vol. 47, pp. 689-734.

Rasmussen J., (1983), Skills, Rules, and Knowledge; Signals, signs, and symbols, and Other Distinctions in Human Performance Models, *IEEE Transactions on Systems Man. and Cybernetics*, Vol 13, No.30, pp. 257-266

Rasmussen E, (1996), Interface Design for information retrieval: Evaluation Issues, ACM Digital Libraries workshop

Roberston S.E. and Hancok-Beaulieu M., (1994), On the evaluation of IR systems. *Information Processing and Management*, Vol. 28(4), pp. 457-466.

Robertson S.E. and Sparck Jones K, (1994), Simple, proven approaches to text retrieval, Cambridge, University of Cambridge, Computer Laboratory, Dec 1994, Tech Report No. 356

Rouse W. and Rouse S., (1984), Human information seeking and design of information systems, *Information Processing and Management*, Vol. 20, No. 1-2, pp. 129-138.

Salton G., (1992), The state of retrieval system evaluation, *Information processing and management*, Vol. 28, No. 4, pp. 441-449.

Saracevic T. and Kantor P., (1988), A study of information seeking and retrieving III Searchers, searches and overlaps, *Journal of the American Society for Information Science*, Vol 39, No. 3, pp 197-216.

Saracevic T, Mokros H. and Su L., (1990), Nature of interaction between users and intermediaries in online searching: a qualitative analysis, ASIS proceedings, pp.47-54.

Sarantinos E. and Johnson P.; (1992); Explanation dialogues: A theory of how experts provide explanations to novices and partial experts. Research Report, Dept of Computer Science, Queen Mary Westfield College, University of London..

Schank R. and Abelson R., (1977) Scripts, Plans, Goals and understanding, Hillsdale, NJ: Lawrence Erlbaum.

Searle J., (1984), Intentionality and its place in nature, Syntheses, Vol 61, pp 3-16.

Sewell W and Teitelbaum S, (1986), Observations of End-user on-line searching behaviour over eleven years, *Journal of the American society for information science,* Vol. 37, No. 4, pp. 234-245.

Smithson S. (1994). Information retrieval evaluation in practice. *Information Processing and Management*, Vol. 30, No. (2), pp. 205-221

Solomon P, (1993), Children's information retrieval behaviour: A case study of an OPAC, *Journal of the American Society of Information Science*, Vol. 44, pp. 245-264

Suchman L.A., (1987), Plans and Situated Action, The problem of Human-Machine communication. Cambridge University Press.

Sutcliffe A.G. and Springett M. V., (1992), From User's problems to Design Errors: Linking evaluation to improving design practice. In People and Computers VII, Eds Monk A., Diaper D. and Harrison M.D., Cambridge University Press. pp 117-135.

Sutcliffe A.G., Bennett I., Doubleday A., and Ryan M., (1995), Designing Query Support for Multiple Databases. In Proceedings of INTERACT-95, Lillehammer, Norway. Ed Gilmore D.

Su L., (1994), The relevance of recall and precision in user evaluation. *Journal of The American Society for Information Science*. Vol. 45, pp. 207-217.

Taylor R., (1968), Question negotiation and information seeking in libraries, College and Research libraries, Vol. 29, pp. 178-194.

Timbie M. and Coombs D., (1969), An interactive information retrieval system; Case studies on the use of DIALOGUE to search the ERIC document files, ERIC Clearing house on educational media and technology.

Tou F, Williams M, Fikes R., Henderson D. amd Malone T., (1982), RABBIT: an intelligent data base assistant, Proceedings of AAAI, Pittsburg, Pennsylvania.

Van Rijsbergen C, (1986), A new theoretical framework for Information retrieval, In Proceedings of ACM SIG-IR Conference, Pisa, Italy, pp. 194-200.

Vickery A, Brooks H.M., Robinson B.A. and Stephens J: (1987) Expert system for referral, Library and Information Research Report 66, The British Library.

Wendt D., (1969) Value of information for decisions, *Journal of Mathematical Psychology*, Vol. 6, pp 430-443

Williams M. and Tou F, (1982), RABBIT: an interface for data base access, Proceedings of the ACM, Dallas, Texas.

Zloof M, (1977), Query by example: A database language, *IBM systems Journal*, Vol 16, No. 4, pp. 324-343.

Bibliography

Ennis M and Sutcliffe A.G, A cognitive model of search behaviour: a first step towards improving interface design for IR, In proceedings of the IR and HCI workshop, British Human Computer Interaction Group, The British Computer Society, September, 1996.

Ennis M. and Sutcliffe A.G., Cognitive models in IR: predicting task support for users needs, Presented at the 1st Cogent workshop, Birbeck College, December 1996.

Sutcliffe A G and Ennis M, Towards a cognitive theory of Information Retrieval, In Interacting With Computers special edition on Human Computer Interaction and Information Retrieval, 1998

Ennis M., Providing task support and interfaces to enhance IR, In proceeding of HCI 97, Doctoral consortium, Bristol, UK.

Ennis M. and Sutcliffe A.G., Towards a predictive model of information seeking, International conference for Information seeking in context, Sheffield August 1998

Ennis M. and Sutcliffe A.G. Cognitive Task Models for Improving the Design of Information Searching User Interfaces, In HCI Letters, Vol 1, Issue 1, pp 37-42, 1998

Sutcliffe A.G, Ennis M. and Watkinson S.J., Empirical studies of end-user information searching, In review by Journal of the American Society for Information Science.

Glossary

Concept	A concept is a component of the query relating to specific domain area. A concept can be articulated using a number of synonym terms.
Term	Terms are used to articulate concepts within queries. A concept may use one or more terms in its articulation.
Strategy	A strategy is a high level plan of action determining the approach taken to retrieval. Strategies can be sub-strategies e.g. sub-goal and formulate a strategy for each goal.
Tactic	A tactic is a lower level operator than a strategy and allows a query to be amended in a specific search situation.
Information need	A information need is a users requirement for information to allow them to satisfy or solve a problem.
Activities	Activities within the search process are the actions the user performs as part of the different search processes.
Process	Processes are higher level operations describing a class of activities.
Search context/ situation	The context of a search are the characteristics of the situation with which the user is faced.
Information indexing policies	These are the classification and information structuring methods used to represent the information within the database.
IR cycle	An IR cycle is composed of the activities, processes, strategies and tactics required to complete one cycle of the generic process model.
Mental model	This relates to the information and expectations a user has about the device based on experience.
Device knowledge	Consists of knowledge relating to the current IR system, knowledge of IR systems in general and strategic IR knowledge.
Domain knowledge	Domain knowledge constitutes how well the user knows the specific domain of the current information need
Gold standard queries	Gold standard queries are the queries developed by a mediation between a domain expert and a device expert determining how to develop the problem
Gold standard solution set	Gold standard solution set constitutes the set of all articles which are relevant to the information need.
Document description	A document representation are the characteristics and information contained within the retrieval system about each article.
Search options	Search options are the possible courses of action in terms of decisions, strategies and activities in any given situation.
Query level evaluation	These are judgements based on the quantity of results retrieved by the query.
Term level evaluation	These are judgements on the quantity of results retrieved by the different components of the query (term or concept)
Article level	These judgements are based on the quality of the articles

evaluation	retrieved by a query in relation to the information need.
Solution evaluation	These are judgements on the success of a retrieval session up to that point in satisfying the information need.
Search diagnosis	These are diagnosis decisions made in relation the various levels of evaluation