DEVELOPING THE USER MODELLING FUNCTION OF AN INTELLIGENT INTERFACE FOR DOCUMENT RETRIEVAL SYSTEMS

Penny Jane Daniels

Submitted for the degree of Doctor of Philosophy

Department of Information Science,
The City University,
London.

December 1987

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank the following people for their contribution to this work :

[redacted], to whom I am extremely grateful: his advice, support, encouragement and patience enabled this thesis to reach completion;
Steve Robertson, who supervised the writing-up of this work;
Nick Belkin, whose ideas initiated some of this research, and who supervised the first two years;
The intermediaries and users at the University of London, who agreed to be interviewed and recorded;
[redacted] of City University, who agreed to be interviewed;
My parents, who have never failed to encourage, support and express enthusiasm for all of my work;

..., and Martin, who has supported me and this work with a great deal of tolerance and good humour!

Abstract

This research forms part of a larger project, the eventual aim of which is the design and implementation of an intelligent interface for document retrieval systems. A number of functions which must be performed by the human intermediary in order to successfully interact with the user have been identified. The research presented here is concerned with one function in particular: the user modelling function, which aims to describe and model various aspects of the user's background, personal characteristics, goals and knowledge. An assumption underlying this research is that an intelligent interface should simulate the functional behaviour of a competent human intermediary. Therefore the ways in which human intermediaries carry out user modelling and employ these models, have been investigated.

The primary method was to make audiorecordings of seven human user/human intermediary interviews in online search service settings, and to subject the transcripts to detailed functional discourse analysis. This analysis produced a specification for the User Model, and identified its components and the knowledge resources that are needed by the intermediary, whether human or automatic, to carry out the function of user modelling. This analysis was supplemented by the examination of a number of users' problem statements, together with their accompanying recordings, which had been collected for another project, and by interviews with three intermediaries. The discourse analysis revealed that the User Model interacts with the other interface functions, and this interaction was also investigated.

The results showed that the User Model comprises a number of subfunctions, requires extensive knowledge resources, and interacts with the other functions, in particular providing information necessary for the other functions' own processing. A formalism for representing the User Model in a computer system is suggested, and an attempt is made to validate the User Model by applying it to a new dialogue. The results of the validation suggested that the User Model is independent of the data on which it is based, and that the formalism can adequately handle a new interaction. The implications of these findings for the design and implementation of the user modelling function in an intelligent interface, and for the design and implementation of the interface as a whole, are outlined.

Chapter 1

INTRODUCTION

An information provision mechanism (IPM) in general consists of three elements - user, intermediary mechanism or device, and a knowledge resource. Examples of IPMs include bibliographic retrieval services, citizens advice bureaux, social security offices and students' advisory services. A spiritual medium could be regarded as an IPM: a person consults a medium, who uses her specialist "skills" and abilities to contact "the other side" on behalf of her client. The goal of an IPM is to provide advice, information and resources helpful in managing or resolving the user's problem. Various studies (eg, Alty and Coombs 1980, Coombs and Alty 1980, Pollack et al 1982) have revealed that **good** advice is more than just recommending a potential solution: it helps the user develop or debug a plan of action, and may help the user to revise or reformulate his/her view of the problem and its likely means of solution. Very few successful advisory interactions are mainly controlled by the advisor, or intermediary: good problem management seems to imply a **cooperative** process, with both user and intermediary negotiating and modifying their views of the problem and of each other.

The information systems that are being considered in the context of this thesis are bibliographic databases hosted by a commercially available online system, in an academic information retrieval service. What tends to happen is that users (usually academic users such as lecturers, research staff, postgraduate students) make an appointment to attend the search service, because they need to retrieve some bibliographic references to enable them to manage a problem, eg, to carry out a research project, to write a thesis or report, or to prepare teaching materials. At the service, a trained intermediary discusses the problem with the user, and having formed some sort of representation of the problem, the user's goals and requirements etc., logs onto the system and accesses the database.

Users generally approach an IPM because they have an ASK (Belkin 1977, 1980) - an ANOMALOUS STATE OF KNOWLEDGE. The nature of an ASK is a central concern of this thesis, and has motivated much of the related research: a user is generally able to recognize that his/her own knowledge with respect to a problem is anomalous, and may even be able to describe the problem itself, but is generally unable to specify what is required to resolve the anomaly, and almost certainly not in a way which is appropriate for search formulation. A major problem with existing information retrieval systems is their requirement that the end user be able to precisely specify the request in terms that the system, rather than the user, can understand, and this is one of the principal reasons why a competent intermediary device, whether human or machine, is essential.

The intermediary's task is to mediate between user and knowledge resource, having gained a sufficiently adequate representation of the user, her/his problem and requirements etc. to enable her to formulate an appropriate search strategy, choose suitable databases and terms, etc. Users typically are not familiar with the access mechanisms, structure and content of online database systems. Their contact with the systems is generally infrequent; they have neither the desire nor the opportunity to become more familiarised. What is perhaps even more important is their inability to specify their information requirements, which are often nebulous and ill-formulated, in ways suitable for search formulation. Competent human intermediaries are trained in the use of the knowledge resources, their access mechanisms, contents etc., and have knowledge about typical users, their goals, requirements, experience and characteristics.

The goal of the information interaction between user and intermediary is to formulate a search stategy which will retrieve bibliographic references useful to the user in managing her/his problem. Although the user generally cannot specify her/his anomalous state of knowledge, he/she engages in an active, cooperative dialogue with the intermediary in order to achieve their mutual goal. Given the essentially non-specifiable nature of users' information needs, together with the distinct roles and differing areas of knowledge and expertise of each participant, it becomes apparent that much of the intermediary's task in the information interaction is concerned with building **cognitive models**. These include models of the user, her problem, her mental model of the system, goals and intentions, output requirements, etc. which can be used as the basis for the subsequent task of database interrogation (Belkin and Windel 1984). This activity is highly complex, and requires a wide range of learning capabilities and sophisticated reasoning skills. Belkin (1984), defining an information system as the dynamic interaction between user and intermediary and knowledge resource, has identified at least seventeen different models that the various participants could possibly have of each other which, together with the extent of their interaction, and the nonspecifiable nature of the interaction, render the overall information system situation problematic.

It is clear that the intermediary is a necessary component of an IPM, and does not exist merely as some sort of translation device standing between user and knowledge resource. However, many users would prefer to carry out their own searches, and a major goal of current IR research is to develop direct end-user access to information retrieval systems, whereby naive and intermittent users can interact directly with the knowledge resource, without needing recourse to a human intermediary. There are many approaches which could be adopted: one is to develop an intelligent interface, ie, a computer system, which could carry out the tasks and functions currently performed by a human intermediary. Such an interface could be regarded as a supplement to, not a replacement for, existing human intermediaries, and could free human intermediaries from routine queries and requests. It could assist and advise on searches in subject areas or on systems with which the intermediary was unfamiliar and, by offering users more direct, but still effective, access to online systems, could have the overall effect of increasing online usage. As a by-product, the human intermediary's role as expert and referral source is likely to be strengthened.

Some research has been done on developing intelligent interfaces to information retrieval systems, but on the assumptions that users can accurately specify their information requirements, that there should be some sort of intermediary device, and that standard query languages and retrieval strategies should be employed. Several researchers (eg, Croft 1985, Croft and Thompson 1985) have identified some disadvantages of the traditional, statistical approach to research in information retrieval systems. Although statistical techniques can be efficient and effective, they generally employ minimal knowledge of the users, their information requirements, and possible search strategies. All users and queries tend to be treated in the same manner, and there is usually very little assistance for the users as they construct the queries.

In recent years, there has been a growing consensus that the entire human-computer system should be viewed as an adaptive, cognitive system, in order for effective interfaces and even whole systems to be designed. The cognitive viewpoint implies that the processing of information is always mediated by some kind of model of the world (de Mey 1977), and a cognitive system, according to Hollnagel and Woods (1983), is an adaptive system which functions using knowledge about itself and the environment in the planning and modification of actions. They stress that the designer of a man-machine system should try to obtain a match between the machine's image and the user characteristics on a cognitive level, rather than just on the level of physical functions. Furthermore, the system should be dynamic:

users themselves may change over time, their tasks may vary, and there may be different populations of users. Ingwersen (1986) advocates a cognitive approach to the design of Information Retrieval systems, and points out that design approaches applied so far are, in general, system-driven, ie, rely on the command language software, database structures, indexing rules etc. He suggests that the design should rely far more on empirical analysis and characteristics of searchers.

An intelligent interface for an information system implies the following (Brooks Daniels & Belkin 1985):

1. it should mediate between the user and the knowledge resource
2. it consists of some computer hardware and software
3. it should simulate the functional behaviour of a good human intermediary.

Therefore, designing and building such an interface depends crucially upon identifying and specifying what a human intermediary in an information system does and needs to do. It appears that the model-building activity that takes place within the pre-online information interaction provides the basis for the development and selection of appropriate retrieval strategies and responses from the knowledge resource to the user. Therefore, if the types of models that are necessary to successful information system performance can be identified, the activities, or functions, which need to be performed in order to construct such models can be specified.

Research groups at The City University, London, and at the Free University, Berlin, have attempted to characterise the intermediary component of an IPM as a set of independent functions, or "experts", each responsible for completing a specific function or task, yet dependent upon one another for information necessary for accomplishing their individual tasks. These individual tasks are seen as components of the overall goal of providing an appropriate response to the user. The MONSTRAT model (Belkin Seeger & Wersig 1983) was derived from analyses of information interactions, and was intended to be a general model, applicable to different types of information systems. The model itself, which consists of a number of interacting experts, was empirically tested by means of a simulation, using humans to carry out the functions (Belkin Hennings & Seeger 1984). The results of the simulation supported the functional model, and showed the set of functions specified in the model to be necessary and sufficient for an intermediary mechanism.

Research groups at The City University (Brooks and Belkin 1983; Daniels, Brooks and Belkin 1985; Brooks, Daniels and Belkin 1985) have undertaken a number of research projects, one goal of which is to design an automated intelligent system interface, which could accomplish the functions that good human intermediaries perform in a similar situation. An important assumption informing this research is that the model-building undertaken by human intermediaries is a skilled and complex activity, and that a great deal of time and energy needs to be spent on developing a realistic model and a deep understanding of the situation before attempting implementation. This work has also resulted in a set of interacting functions or "experts", which are responsible for carrying out the same tasks as the functions in the MONSTRAT model.

Several of the 10 functions that have been identified by the above research groups are concerned with building appropriate models of various aspects of the user, the user's situation and the user's problem. For example,

PROBLEM DESCRIPTION generate a description of the user's problem

RESPONSE GENERATOR determine structure of response required by user which is appropriate to the situation.

USER MODEL generate a description of the user type, goals, knowledge, etc.

This thesis has resulted from work carried out at The City University as part of the research projects briefly described above. The eventual aim of the research is to develop an intelligent interface for document retrieval systems, which can simulate the functional behaviour of a competent human intermediary. It is envisaged that the interface will consist of ten, independent functions or "experts", each responsible for completing a particular function (eg, user modelling, building a Problem Description) yet dependent upon each other for accomplishing their tasks. The interface and user should be able to co-operate in a mixed-initiative, natural language dialogue whose purpose would be to develop conceptual models of the user and his/her problem, which in turn could be used to construct a search formulation which would effectively retrieve appropriate information.

The thesis will concentrate on just one of the ten functions of the intelligent interface: the USER MODELLING function. The following aspects will be investigated in detail:

- how the human intermediary carries out the function of user modelling
- the subfunctions, ie, the components of the User Model
- the Knowledge Resources required by the User Model
- the interaction of the User Model with the other functions
- a formalism in which the User Model might be implemented in an operational interface
- some approaches for verifying the adequacy of the User Model, its subfunctions, knowledge resources and interactions.

The goal of this thesis is **not** to implement an operational interface, but rather to suggest how the user modelling component should be designed, what it should encompass, how it should be represented for the purposes of implementation, and its relationship with the other experts. The following aspects of the system are beyond the scope of this thesis, and will therefore not be discussed here: Natural Language, Dialogue Capabilities, Explanation, Hardware and the structure of the back-end of the system.

The first part of Chapter 2 discusses user modelling, both within intelligent interfaces and in other types of system, and the second section briefly discusses knowledge elicitation techniques, and explains why the particular methodology used in this research was chosen. Chapter 3 describes the precursors to the present study, and the methodology employed for eliciting the User Model and the other functions is discussed in Chapter 4. The results of that analysis are described in Chapter 5 and a possible formalism for representing the User Model is presented in Chapter 6. Chapter 7 describes an attempt to demonstrate that the User Model is valid, ie, that it is independent of the data on which it is based. Chapter 8 discusses the findings of, and the issues raised by, this thesis, and presents some conclusions.

Chapter 2

USER MODELLING AND INTELLIGENT INTERFACES FOR INFORMATION RETRIEVAL SYSTEMS

Part I
Review of Theory

2.1 Introduction

Much of the literature which is relevant to this chapter has already been adequately reviewed by Brooks (1986), Daniels (1986), and Belkin and Vickery (1985). Therefore, the aim of this section of the thesis is not to provide a comprehensive and exhaustive review, but rather to focus attention on, and summarize, the most important concepts and issues in this field.

2.2 Intelligent Interfaces and Information Retrieval

In general, an intelligent interface to a computer system can be defined as one which embodies aspects of a human being's intelligence and knowledge, and which can adapt to, and cooperate with, the user. It could be argued that an interface which is truly intelligent should ideally be capable of sustaining an active, mixed-iniative dialogue in natural language with the user.

An intelligent interface for a document retrieval system should attempt to incorporate and use the expert knowledge and search skills of an experienced, competent human intermediary. Brooks (1986) has comprehensively reviewed recent work concerned with developing both non-expert, and expert (or intelligent), interfaces to document retrieval systems, and concludes that most interfaces proposed or developed to date are 'non-expert' in that they do not attempt to incorporate and use the human intermediary's skills and knowledge. Rather, they aim to assist the user with tasks such as connecting with the remote computer, request entry, query language commands, and so forth. Recent developments in AI have inspired some researchers to attempt to apply expert system techniques to develop intelligent interfaces, and several 'expert' intelligent interfaces for document retrieval systems are now being developed, for example, Expert (Marcus 1981, Yip 1981, Marcus 1985), IR-NLI (Guida and Tasso 1983), An Expert System for IR (Shoval 1985), CANSEARCH (Pollitt 1981, 1984, 1985), and PLEXUS (Vickery et al 1986), all of which are reviewed in Brooks (1986). However, there are various difficulties in applying expert system techniques to interfaces for document retrieval systems, one of the most important being that the problem domain itself, from pre-search interview to search strategy formulation, is not well understood. Very little of the current expert-system interface work is explicitly based on a preliminary study of either the problem domain or the human intermediary and her/his knowledge and skills (Fidel 1986).

To a lesser extent, some research has also been carried out on the design of systems which are complete information systems in their own right, for example, THOMAS (Oddy 1977), Grundy (Rich 1979a, 1979b) and ASK (Belkin 1977, 1980): these have also been reviewed by Brooks, and are discussed below, in Section 2.4, and in Chapter 3.

Whatever the problem domain of an intelligent interface, it is desirable that it should include at least some user modelling capabilities, so that it can adapt to the user's experience and characteristics, and provide responses and information appropriate to a user's individual needs. There is considerable debate regarding the utility and feasibility of user modelling, and there appears to be no consensus as to what an adequate user model should actually contain.

2.3 User Modelling

A more comprehensive survey may be found in Daniels (1986), who has reviewed recent and current work in user modelling, focussing on the utility of various types and aspects of user models to the IR situation.

2.3.1 What are User Models?

Broadly speaking, cognitive models are images that the components of a system, whether the components be people or machines, have of themselves, of each other, and of the world. For communication to take place between two parties, it is necessary for each party to have a model of the other, which almost certainly does not correspond to the other's model of himself. In the man-machine interface context, Borgman (1984) has identified a number of types of cognitive models, one of which is the **user model**, which refers to the computer's model of the user and should probably also include the computer's model of the user's model of the system. User models fall into two broad classes: Empirical Quantitative Models, which aim to correlate a person's external performance with given design parameters, and Analytical Cognitive Models, which attempt to represent a user's internal reasoning, beliefs or knowledge, ie, her/his **cognitive** processes. In general, this latter type of user model can enable systems to present users with interfaces appropriate to their needs and sensitive to their characteristics, and can provide systems with the means for inferring useful information about the user based on a small number of explicitly stated facts. It is with this second class of user models, ie, analytical cognitive models, that this thesis is concerned.

2.3.2 The Classification of User Models

Although user models traditionally have been classified according to three principal dimensions (Rich 1983), it may be more accurate and appropriate to refer to classifications of **data** within each user model, as there is no reason why a user model should not contain more than one type of data at any time.

Canonical vs. Individual Models

Systems which assume a standard, canonical user include Gershman's Yellow Pages Assistant (1981), and Wilensky's Unix Consultant (Wilensky, Arens and Chin 1984). In such systems, a standard user model is assumed at the beginning of each session, rather than being built from scratch as the session progresses. In many computer assisted instruction (CAI) systems, users are modelled as a subset or overlay of a wider class of knowledge, skills, etc., (Clowes et al 1985) or else the system may search for differences between observed user behaviour and some wider set of skills or knowledge, in which case the model is a differential one. Examples of systems using some form of subset user model include GUIDON (Clancey 1982), UMFE (Sleeman 1985), DEBUGGY (Burton 1982) and Grundy (Rich 1979a, 1979b; see also Chapter 6 and Section 2.4, this chapter).

Explicit vs. Inferred Models

Explicit user models are those which are constructed by the user or specified by the system designer; inferred models are those which are inferred or abstracted by the computer on the basis of the user's behaviour. According to Rich (1983), one approach to modelling is to let the users provide their own model, or explicitly create their own environments, for example, in electronic mail systems or 'personalised database systems' (Mittman and Borman 1975). However, this approach is not suitable for naive or casual users, since some expertise is required to know what, and how, something is to be specified. On the other hand, the system may be provided with enough information about users, or the means for acquiring that information, so that it can itself have the responsibility for constructing, or inferring, a model of its user. This can be done in a trivial sense, eg, where a program asks the user whether or not she has any system experience, and proceeds to use that reply to determine the level of complexity of error messages.

A more sophisticated approach to user modelling is required by many systems in order to deal with both the need for more information about each user, and with the problem that users cannot always specify what they need to know, or what they do not know, a central concern of the ASK hypothesis (Belkin 1977, 1980) and of CAI systems. It would appear that the best approach to implicit modelling is to allow the system to form a reasonable initial user model, perhaps based on the user's answer to just a few preliminary questions, or on known characteristics of the overall user community, ie, on a limited amount of whatever information is available to, or elicited by, the system. As a user proceeds to interact with the system, further information will become available, and the system can begin to update its user model, revise its hypotheses, alter default values, etc. An example of a system which constructs its user model in this manner is Grundy (Rich 1979a, 1979b) (see Section 2.4, this Chapter, and Chapter 6).

Perhaps the most important implication of choosing to let the system build its own model of the user, is that much of the information contained in the model will be speculative. The system will therefore require a means for ascertaining facts, resolving conflicts, assigning confidence ratings to the hypotheses, and updating and modifying the model.

Long-term vs. Short-term models

Long-term models refer to enduring user characteristics, such as areas of interest and expertise, whereas short-term models may include the subject of the last sentence typed, or the specific problem that the user is currently tackling. Although short-term models tend to be of limited usefulness, as they only refer to current context-specific goals and information, they are essential in assisting natural language understanding systems to cope with phenomena such as anaphora, ambiguity, and pronominal reference. Rich (1983) suggests that many systems could usefully employ a large amount of much more stable, long-term knowledge about their users, eg, user's experience with computer systems in general, user's familiarity with underlying task domain, etc. However, it seems possible that user characteristics such as system familiarity and even domain knowledge, should not necessarily be regarded as long-term characteristics, as the user's experience, familiarity etc. will almost certainly change during interaction with the system. In such cases, it may be more useful to describe these types of models as **dynamic** or **static.**

Dynamic vs. Static models

Dynamic models, according to Sparck-Jones (1984), are context-dependent changing models from the system's point of view, and refer to changes of user state which are dependent upon interaction with the system. Static models represent permanent features of users which are independent of the behaviour of the system, and are consistent over the session. Perhaps the most important types of dynamic models are those of the user's goals, plans and beliefs in the specific context of the system's operations.

Dynamic models are particularly useful in systems which display strongly interactive, negotiating behaviour, where the user has an active role, for example, CAI systems, query and advice systems, information retrieval systems. In CAI systems, dynamic models are used for discovering the conceptualisation underlying the user's actions and responses and, in the query and advice context, they may be employed to discover the real, underlying reasons for the user's query, which often are not apparent at the beginning of the session (Pollack et al 1982). In such cases, the maintenance of dynamic user models can enable the system to generate appropriate additional information and individually tailored explanations. In the IRS situation, for example, it is particularly vital that the system can detect and correct, using tailored explanations, misconceptions and inaccuracies in the user's model of the system. Otherwise, the user is liable to be disappointed by the output or frustrated by the system's apparent inability to retrieve the appropriate references.

2.3.3 Aspects of User Modelling

User modelling is not a straightforward task, and there is still no agreement as to what types of information an adequate user model should contain, how that information should be used by the rest of the system, and even whether a user model is required at all. Most recent user models that have been implemented tend to be rather simplistic and contain a limited amount of knowledge, but they have functioned adequately because researchers have been working in well-defined, limited domains and have proceeded initially on the basis of strong simplifying assumptions. However, more open-ended tasks will require much richer, more comprehensive user modelling systems, for example, see KNOESPHERE (Lenat et al 1983), a system which will eventually require an immensely detailed, wide-ranging user model. Any systems which have to interact with a wide range of users and having diverse problems or queries, such as information bureaux, social security offices, document retrieval systems, are confronted with a complex, open-ended task requiring correspondingly comprehensive user models. Such models should include more general traits and cognitive behaviour, for example, preferred learning style, background knowledge, level of motivation, etc.

Boguraev (1985) suggests that perhaps that there is no such thing as a separate, discrete component called 'user model', but rather, that systems should be capable of various activities which will depend on distributed knowledge about aspects of the user, eg, understanding of the problem, model of the system, beliefs, plans, goals, strategies, etc. Techniques from related areas in AI which are also concerned with modelling or, at the least, taking into consideration, various aspects of the user include:

Beliefs and Belief Systems

Recent work has concentrated on organisational problems, representational problems, and conceptual problems such as embedded beliefs and reflexive beliefs (eg, see Maida 1983, Konolige 1983). It is important to attempt to represent a user's beliefs about a system, and to reconstruct the user's reasoning with these beliefs, in order to identify the causes of user actions, errors and misconceptions.

Goal Analysis and Planning

Goal analysis and planning are essentially interrelated, and Genesereth (1982) defines a plan as " a program of action, or sequence of steps, designed to achieve a desired goal". CAI systems may attempt to recognize a user's plans so that tailored advice may be offered (eg, see Genesereth 1982, Woolf and McDonald 1983), and planning is also essential if a system is to have a natural language interface (eg, Carberry 1983, Sidner and Israel 1981). In general, it seems unlikely that a *user's* plans should be incorporated into the user models in the information retrieval situation, as the user probably does not have a plan as such in the pre-online interaction with the intermediary. It is more likely that the intermediary has some kind of plan, perhaps in the form of a search strategy, whereas the user has the higher-level *goal* of retrieving some information which will later assist him in problem management.

Discourse Modelling and Dialogue Capabilities

Recent approaches have stressed the importance of taking into account aspects of the user, the user's beliefs and viewpoints when, for example, expanding elliptical input or identifying the source of anaphoric reference. Peoples' goals affect what they say, and how they say it, and a listener interprets what a speaker says according to her view, or model, of the speaker's goals. People also rely upon a shared view of what the point of a conversation is at any given moment, the **focus of attention**, in order to economise substantially on what is being said, by using devices such as anaphora and ellipsis. The notion of **focus** is used by Grosz (1977), by Carberry (1983) in his TRACK system, and by Jerrams Smith (1985) in her system, SUSI.

The unifying theme underlying the above paragraphs is the importance of shared implicit knowledge to effective communication. This mutual knowledge may refer to each participant's model, or view, of certain aspects of the other, eg, the other's plans, goals, beliefs, etc., and also encompasses discourse-specific phenomena, such as the current focus, how to draw inferences from speech acts, etc. It is not sufficient to rely upon the surface features alone of a dialogue as a means for interpretation, generation and disambiguation, as such features may elicit incomplete, inaccurate and even misleading hypotheses and models; rather, the participants in an exchange also need recourse to internal, implicit, shared knowledge of each other and of current discourse features and processes, in order to facilitate and structure the conversation.

2.3.4 Examples of User Models in Specific Systems

A comprehensive review of user models which have been implemented in, or proposed for, specific systems, may be found in Daniels (1986). Specific examples of particular relevance to Expert Systems, CAI systems, Intelligent Front Ends, and Information Retrieval Systems, are discussed below.

User Models in Expert Systems

Sparck Jones (1984) has emphasized the necessity for user models if an expert system is to have a natural language interface, and suggests that user models in expert systems can take various forms and can be used for various purposes. She takes as an example a hypothetical social security expert system, and considers the contribution various types of user model could make to such a sytem. For example, *patient* models can be distinguished from *agent* models: patient models refer to the human subject of the expert system's decisions (eg, a client in a DHSS office) whereas agent models refer to the person conducting the interaction (eg, clerk in a DHSS office).

HAM-ANS (Hoeppner et al 1983) is a large natural language dialogue system which has been devised to serve as an interface to various background systems, one of which is an expert system for making hotel reservations. One of the principal concerns in this system is to examine explicit partner (user) models. The system adopts the role of hotel manager, and attempts to recognise the user's specific desires concerning the room as they are revealed - usually indirectly - in his utterances, and to make use of the various devices available in natural language that allow that particular room to be presented in an appealing manner. It should be emphasised that although the issue of need identification is addressed here, the system operates in a very narrow domain, where users approach the system with only one goal needing to be fulfilled.

KNOESPHERE (Lenat et al 1983) is a project concerned with building an encyclopaedic expert system, eventually encompassing the entire contents of the *Encyclopaedia Britannica*. A central issue in this work is how to aid users in searching complex bodies of knowledge, and the authors envisage constructing an extremely comprehensive user model which (if ever implemented) would appear capable of accounting for every user type, preference, goal, belief, etc, that could ever be encountered. The user model would be capable of both subset and individual user modelling, and each model would include physical appearance, goals, state of knowledge, memory capabilities, inferencing capabilities, interests and biases, preferences for different teaching methods, and other miscellaneous attributes. The user model will also be able to make inferences about the user's current disposition from dynamic cues, such as fidgeting at the terminal is likely to indicate boredom!

User Models in CAI Systems

It is most unlikely that a CAI system could perform effectively unless it included a user model in some form. As the user model (usually referred to in the CAI literature as *student* model) is usually constructed as a consequence of communication between the student and the system, it is clear that dynamic models will be most appropriate in the CAI situation, where the system functions only in relation to the user's individual behaviour (Sparck Jones 1984). Dynamic models are used for discovering the user's state of knowledge, and are necessary for generating tailored explanations. The system must be capable of updating and revising its user model as the student's knowledge increases and as more concepts or facts are learned. Implicit user models are required, as people are seldom able to identify what they do not know or what they know incorrectly; the system must be able to infer this.

Most CAI systems incorporate some kind of 'subset' user model, for example, WUSOR II and III (Goldstein 1982),and the West System (Burton and Brown 1982). GUIDON (Clancey 1982) is a case method tutor developed to explore the problem of transferring the expertise of a MYCIN-like expert system to a student. It presents the student with information about an imaginary patient and the student has to predict from the symptoms the correct diagnosis and therapy.

GUIDON employs an overlay student model, in which the student's domain knowledge is viewed as a subset of that of the expert program's. It also makes use of a focus record to keep track of factors in which the user has recently displayed interest, in order to maintain continuity during the dialogue. The overlay user model generates hypotheses about what the student knows and if the hypotheses are subsequently confirmed, the user model is updated. After updating, the tutor needs to deal with discrepancies between the student's hypothesis, and what the expert program knows.

The system operates on the basis of several fairly important assumptions:

(a) the student wishes to learn to solve MYCIN-type problems; and

(b) the student has a suitable background for solving the case, and can understand the terminology.

Essentially, the program is designed for well-motivated students capable of engaging in serious, mixed-initiative dialogue.

The problem with the type of user models used in CAI systems essentially stems from the fact that most of these models attempt to represent the user as a subset, overlay, etc. of an ideal or expert user. Canonical and subset modelling is likely to present insurmountable problems in the IR domain, as it is not possible to identify an expert or ideal user, whose knowledge or skill could provide the means for subset or overlay modelling. A standard, typical or ideal user probably does not exist in the IR situation: even within limited IR systems, for example academic online information retrieval services, users come to the system from a variety of backgrounds, with varying levels of experience and states of knowledge.

In CAI systems, there is little or no attempt to represent the user as possessing *individual* goals, motives, traits and so forth: it is assumed that all users approach the system in order to learn to solve a problem, execute a particular task, etc. The IR domain cannot be regarded as a simple task-execution domain where the user is solving a problem or learning how to do something online: in an IRS, the user is collecting information for later problem management, and his goals are not always apparent at the beginning of a session. Also, users in an IRS tend to have more than one goal; Daniels (1985) for example, has identified four levels of user goals in an IRS.

Apart from assuming a standard goal for all users, CAI systems also frequently assume 'good' levels of motivation and appropriate backgrounds. Neither of these assumptions can be made in the IR domain, where users have varying states of knowledge and academic backgrounds, and are not necessarily similarly motivated.

In conclusion, user models in CAI systems could probably not be applied in the IR situation for the following reasons: they generally attempt to represent the user as a subset or overlay of an *expert* or *ideal* user; they do not usually attempt to represent, or adapt to, *individual* user characteristics; simplifying assumptions are made, and the systems operate in limited, task-execution domains where such assumptions are generally appropriate.

User Models in Intelligent/Cooperative Front Ends

It is generally believed (eg, Boguraev 1985, Sparck Jones 1985) that if an intelligent interface is to be truly cooperative, and capable of sustaining an active, mixed-initiative dialogue with a user, then it will almost certainly require full natural language capabilities. In order to sustain such a dialogue, an interface will require the ability to construct and utilise appropriate user models. Furthermore, natural language can provide the *only* expressive and rich enough source of modelling information for systems which intend to construct and implement dynamic, implicit user models.

UMFE (Sleeman 1985) is a data-driven user modelling front end subsystem, which aims to provide a tailored explanation of the NEOMYCIN system and infers *overlay* user models. UMFE determines the user's level of sophistication by asking as few questions as possible and then presents a response in terms of concepts which it believes the user will understand. The system uses investigator-defined inference rules to suggest additional concepts the user may or may not know, given the concepts the user indicated he/she knew in earlier questioning.

The system consists of a hierarchy of concepts, each of which has associated with it a difficulty and an importance rating. The inference rules essentially exploit the hierarchical nature of the database, and assume that if a concept is known/not known, then so will its parent concepts and all of its siblings which have comparable and higher importance ratings.

It has been suggested that future work on UMFE may concentrate on an attempt to initialise the user model as the result of asking a few preliminary questions (eg, Rich 1979a, 1979b), or by telling the system that the user is a member of a certain category of users, and using an inference rule for each attribute. It is acknowledged that it would be important, however, to give low credibility to inferences produced by very general rules, and higher credibility to more specific inference rules.

2.3.5 User Models in the Information Retrieval Situation

An important feature of the information dialogue between user and intermediary is that ~~that~~ the user may be able to recognise that his/her own knowledge with respect to some problem is anomalous, and may be able to describe the problem itself, but is usually unable to say precisely what is required to remove the anomaly. In fact, the IR situation would appear to be a good example of the type of cooperative dialogue described by Pollack et al (1982): using an example of a naturally occurring expert system consultation, she makes the point that users are very active agents in the reasoning process of a cooperative dialogue, and that their goals, plans and beliefs undergo constant reassessment. Therefore, the expert, whether a person or a computer, needs to continually maintain and update its user model, and the only method of obtaining information about the user's perception of the problem, her goals, plans etc., is through active participation in natural language dialogue.

A user model should enable an intelligent interface for IR to engage in natural language dialogue with the user, by facilitating both the interpretation of linguistic inputs, and the generation of linguistic outputs. By identifying the user's model of the system, which is often inaccurate and incomplete, it could generate acceptable, tailored explanations concerning system operations which are understandable to the user. When the user has difficulty in specifying her/his problem and information requirements, the user model could assist the system in identifying likely areas of interest and potential sources of information by using hypotheses connected with the user's goals, beliefs, background, state of knowledge, and so on. A dynamic, individual user model could enable the system to adapt to, and cooperate with, the user by tracking his changing goals, plans, beliefs, model of the system and perception of the problem.

However, the difficulties involved in user modelling should not be underestimated. The problems involved in natural language interpretation and generation will eventually need to be confronted, and some of these issues are still not well understood. Inference is also a problem: information will often be conveyed indirectly, inaccurately or incorrectly and the system will ideally need a method to resolve conflicts between competing inferences. Another problem is that of property recognition (Sparck Jones 1984), which depends on how familiar the system is with the property concepts involved. In order to recognise a wide range of user properties, goals, attitudes, etc., the system will need to deploy extensive world, domain and user knowledge.

A further problem remains, one that is perhaps particularly relevant in the IR situation: what types of knowledge, facts, concepts, etc. should a user model contain, and how should it be used? It is fairly clear what sort of knowledge should be included in user models in well-defined domains such as CAI and Help systems: the user is pursuing a clear-cut task with specific pre-defined goals. However, the situation is quite different in contexts such as human advice-giving, where a number of goals at different levels may be pursued throughout a complex, dynamic, highly interactive dialogue. We do not yet fully understand how *humans* manage such problems, the types of knowledge they deploy, the models they build and the decision processes they engage in. Boguraev (1985) believes that it should be useful to analyse a range of real interactions between people, where language is the common medium for communication in the process of performing a certain task and achieving one or several goals. The pre-online information retrieval interaction appears to be one situation where this type of analysis can be successfully applied.

2.4 Information Retrieval Research incorporating User Models

2.4.1 THOMAS

THOMAS (Oddy 1977) is a computer program which offers no facilities for query formulation in the usual sense, and proceeds on the basis of the user's reaction to the references and document descriptions that he is shown. This work was intended to take account of the nonspecifiable nature of people's information needs, ie, users often have ill-defined needs which they cannot specify, and they often cope with this problem by browsing among the library shelves. Oddy suggested that users might learn more about what would be useful to them through learning about what was available, and therefore constructed a system designed for browsing through the database.

An important assumption is that the user and the intermediary each has his own world model, and that each must try to construct a model of the other's interest, in relation to his own view of the world. THOMAS forms an image of the user's interest, derived from its world model which consists solely of knowledge about the organisation of literature, and chooses references for display according to the state of the image. This image is modified continuously in the light of the user's reactions to the displays.

The system is mixed-initiative, and the user does not have to learn any commands. The important issue here is that THOMAS does manage to incorporate some aspects of the computer's model of the user, although the main component of this user model is the context of the topic of enquiry, and THOMAS adapts to its changing model of the user. Furthermore, the system was designed around the concept of the non-specifiability of users' information requirements.

2.4.2 Grundy

Grundy (Rich 1979a, 1979b; also see Chapter 6) is an interactive system that acts as librarian and recommends novels to library users. The user has a predefined goal of finding a novel, the task of the overall system is to retrieve such a book, and the computer's specific role is to construct a user model that can provide details of the user's likely interests. Rich proposes *stereotypes* as a useful mechanism for building models of individual users on the basis of limited but explicit information about them.

Grundy begins a consultation by taking just a few words of self description from the user (eg, serious, intellectual, feminist) which are then used to activate a number of stereotypes. These stereotypes enter values into the user model, and any conflicts between different stereotypes are then resolved. Depending on how confident Grundy is in the user model, the system will either describe and recommend a novel to the user, or ask the user further questions. If the user does not like the sound of Grundy's suggestion, Grundy attempts to find out why and modifies its suggestions accordingly.

The user model in Grundy is exploited to provide guidance to the underlying system in the performance in three ways:

1. it enables the system to focus on a small set of potentially appropriate novels;
2. it provides a way of evaluating each of these books in detail in order to select the best one to recommend; and

3. it provides a way of selecting which of the chosen books' many features would be of interest to a particular user.

In addition to demonstrating the usefulness of stereotypes for implicit user modelling, Grundy also illustrates how feedback may be used to update the user model continuously, and how confidence ratings and conflict resolution can help deal with uncertainty. However, the set of stereotypes, their facets and triggers do **not** appear to have been empirically elicited, rather they were intuited by Rich. Furthermore, although some valuable principles are illustrated by Grundy, as Clowes et al (1985) point out, the users have no explicit goal or set of tasks that they are trying to perform, so the system provides no insight into this area.

2.4.3 An Expert Assistant for IR

Croft (1985; Croft and Thompson 1985) is building an Expert Assistant for a document retrieval system, which controls the entire user/system interaction by using knowledge about different types of users, methods of formulating queries, different retrieval strategies, and how information should be displayed. The aim of the project is to improve retrieval effectiveness through a detailed understanding of the user's information need and the flexible use of search strategies, rather than by attempting to produce complex document representations with natural language analysis and detailed domain knowledge.

The Expert Assistant is claimed to have many similarities to a human intermediary who, although frequently not familiar with the subject area of each user that comes to her, knows a great deal about the system. By interviewing each user, the intermediary is able to form a model of the information need and decide how best to use the system to satisy that need. It is intended that the Expert Assistant system will be guided by plans derived from typical user/intermediary interactions, although it does not appear that any plans have yet been identified or represented.

The main components of the Expert Assistant are the query formulation controller, the search controller and the interface manager. These components interact within a blackboard style architecture similar to that employed by Hearsay-II (Erman and Lesser 1979), and the blackboard is equivalent to the system's STM (Short Term Memory). The query formulation controller is responsible for controlling most of the interaction with the user, and its task is to construct the user models and request models in the short-term database. It accesses the long-term database for information about the document collection, user stereotypes, and previously defined user models. It can help the user to browse the database when appropriate, and tells the interface manager which information to display, and which information is required.

The user models in this system are intended to represent specific characteristics of individual users which are independent of particular queries. They include domain knowledge provided by the user during query formulation, and information derived from user stereotypes, ie, descriptions of typical user types, for example, expert/novice user of the system. Other classes of users may be identified by the type of search required, eg, high recall, high precision, current awareness, etc. It may also be useful to identify users as being expert in a particular field, in which case that information could be used to determine how user-supplied domain knowledge would be treated.

At the present time, Croft's Expert Assistant has yet to be fully implemented and it is therefore difficult to comment on the adequacy of the user models in the system. However, they appear to contain very little detailed knowledge and seem to be confined to rather limited user descriptions, eg, expert/novice.

2.4.4 Plexus - An expert system for referral

The task of Plexus (Vickery et al 1986) is to assist users who have some problem in the domain of gardening/horticulture to identify information sources which potentially might be of use in helping to resolve that problem. Plexus therefore differs from the interfaces described in this section in that it does not interface with a knowledge resource which is a database of bibliographic references as such, but rather with a database of referral resources, ie, of information sources themselves.

The design of Plexus is based on a functional model comprising five functions, one of which is the GETUM module which is concerned with constructing a model of the user. The user characteristics represented in this model are:

(a) the extent of the user's knowledge of the Plexus system
(b) the relationship between the topic of the query and the user's job
(c) the extent of the user's practical experience in the field
(d) the user's familiarity with information resources in the domain
(e) what advice seeking activity the user has undertaken already
(f) the user's geographic location (in broad terms).

This user model is constructed at the beginning of the session by explicit questioning of the user: the user model is an explicit, static model. It appears that the user model is primarily used by the EVALUAT function, which uses knowledge about the user to determine the order in which items are displayed, for example, institutions which are located near to the user are presented before those which are located some distance away. In this prototype, a more or less linear sequencing of functions is assumed. Processing proceeds from the GETUM module to the other modules, but there is provision for some backtracking through the subtasks. The authors recognize that, for a more heterogeneous user community, Plexus will need sophisticated user modelling capabilities, an issue which will be investigated in Phase 2 of the project.

2.4.5 CODER

The CODER system (Fox and France 1986a, 1986b) has been principally designed to serve as a testbed for determining how useful various AI techniques are in improving the effectiveness, efficiency and usability of information storage systems, although the current aim is to apply the system to analysing and retrieving passages of messages in a collection of AI List Digests. CODER uses domain specific knowledge derived from the machine readable text of the Handbook of AI, and is supplemented by the Collins English dictionary, transformed into a Prolog fact base.

CODER is developed as a collection of experts communicating through central blackboards, and is a distributed system which can be divided up across several machines. Although it is said that the system has a User Model Builder, it far from clear exactly what this will contain. Rather than suggesting characteristics and information which should be modelled, the authors concentrate on what is **not** known about users, for example, whether incremental search is liked by a majority or a minority of users, which individuals benefit from being able to reason with a thesaurus or prefer to examine retrieved documents, what sort of information should be stored after a session concludes, and so forth. If a comprehensive user model is envisaged for CODER, it is clear that some research into user characteristics and attributes is required.

2.4.6 MONSTRAT and ASK: the work of Belkin and his colleagues

Belkin and his colleagues (eg, Daniels et al 1985, Brooks et al 1985) have been undertaking a series of research projects, one goal of which is to design an automated intelligent system interface, which could accomplish the functions that good human intermediaries perform in an online information retrieval situation. The basic approach has been to focus on the functions which need to be performed in the user-intermediary interface, and why they are performed. Some recent research has indicated that a number of these functions are explicitly concerned with building cognitive models, one of which is the User Model. This work is described in more detail in Chapter 3.

2.5 Conclusions

Most of the user models employed in the systems described in the preceding sections seem to work, or at least be capable of working, reasonably well, largely because they are operating in constrained domains where strong simplifying assumptions have been made. The user models are frequently generated from one knowledge source only, and contain limited types or 'dimensions' of knowledge. It does appear, however, that the issue of evaluation has been neglected, although various techniques could be applied for this purpose. User models also need to develop better capabilities for deciding between conflicting hypotheses or dealing with uncertainty in the data, a problem which arises especially in implicit modelling.

Although some work has been undertaken on cognitive modelling in IR, few researchers have yet attempted to identify, let alone implement, a specific, comprehensive user model and its subcomponents for an information retrieval system. However, it is possible to make a few suggestions as to what sort of user model would be most appropriate in the general IR situation:

(a) User models should be individual, rather than canonical, as IR has a wide domain of application where it is generally not possible to identify either an "ideal" user or a subset of "expert" knowledge or skill. However, intermediaries do, to some extent, make some stereotyped judgements about users, and such stereotypes depend largely on the type of system in question.

(b) Dynamic *and* static types of data are required in the IR situation. Dynamic models should account for the user's changing goals, view of the system, etc., whereas static models will include information such as the user's subject background, the institution to which she/he belongs, etc.

(c) Implicit user models are necessary in the IR context, mainly because users are frequently unable to specify precisely their information requirements and their anomalous states of knowledge (ASKS). However, there is no reason why some user modelling should not be carried out explicitly: intermediaries *do* ask some direct questions so that they can at least form a reasonable initial model of the user. A reasonable approach may be to allow explicit modelling to guide the initial interaction, whereas implicit modelling can account for individuality in users and allow the system to update and modify its models during the course of the interaction.

(d) It seems unlikely that a user's plans should be included as part of the user model in the general IR situation, as most users probably do not have a plan of how the interaction should proceed. On the other hand, the intermediary's search strategy can be regarded as a plan, and overall plans of some description will probably be required to drive the system.

(e) The issue of determining precisely what types of knowledge, and at which level of detail or granularity, should be represented in a user model, in whatever domain, is crucial. Knowledge elicitation perhaps presents fewer problems in simple task-execution domains, where the user has a well-defined goal and must learn a finite set of concepts or procedures. The IR domain is more complicated: neither the knowledge resource, nor the user's problem, is highly constrained or well-organised, and there may be interactions amongst a number of different cognitive models. Despite an increasing awareness of the importance of cognitive modelling in IR, few researchers have attempted to discover and identify, in detail, precisely what types of user models are constructed and how they are used. Several researchers (eg, Croft 1985) specify the user model as a separate and necessary subsystem, but apart from a few exceptions (eg, Brooks et al 1985, Daniels 1985) there has been little *empirical* investigation into user modelling activities in IR. Although the cognitive modelling is no doubt complex, it is possible to undertake some real-life analysis if we use, as our raw data, the pre-online intermediary-user dialogue.

Part II

Review of Methods

2.6 Introduction

A comprehensive review of knowledge acquisition methods is presented in Welbank (1983) and Brooks (1986). The aim of this section is to briefly summarise the main techniques available.

In order to find out what intermediaries do, what knowledge they need and how they apply that knowledge, it will be necessary to spend a considerable amount of time developing a realistic model and deep understanding of the situation before attempting implementation.

Buchanan (Buchanan et al 1983) defines knowledge acquisition as

> the transfer and transformation of problem solving expertise from some knowledge source to a program. Potential sources of knowledge include human experts, textbooks, databases and even one's own experience.

What would appear to be a relatively straightforward task turns out in practice to be extremely difficult, time consuming and complex (Buchanan 1982, Welbank 1983). This is because:

- human knowledge is complex, messy and ill-formulated
- experts find it difficult, even impossible, to articulate their knowledge
- the more expert the human expert, the more "unconscious" their problem solving
- current knowledge elicitation techniques are poorly understood and of limited applicability.

Within the field of AI itself, until recently (eg, see 1987 issue of International Journal of Man Machine Studies) there has been little in the way of methods or techniques that could be used to facilitate knowledge elicitation. Attention has been turned, therefore, to other disciplines where knowledge transfer is of interest, eg, cognitive psychology.

2.7 Techniques for Knowledge Elicitation

2.7.1 Interviewing the Expert

Interviewing the expert, either informally or by making use of structured interviewing methods, is a commonly used technique. It provides a relatively simple means of acquiring a substantial amount of basic knowledge about a problem domain, particularly concerning its concepts and terminology. Questioning techniques that have been made use of in acquiring expert knowledge include generalised checklists, critical incident reporting and autobiographical memory procedures (Gammack and Young 1985), and the use of Personal Construct Theory and the Repertory Grid (Kelly 1955) has also been investigated by Shaw (eg, 1984).

However, the technique of interviewing does have its disadvantages, and may not be appropriate in areas where experts have particular difficulty in articulating their knowledge (Welbank 1983).

2.7.2 Verbal Protocol Analysis

The use of verbal protocol analysis, ie, analysing recordings of experts thinking aloud as they carry out a task, is also common (eg, Kuipers and Kassirer 1983). A comprehensive survey of the use of verbal protocol analysis was carried out by Ericsson and Simon (1980). The survey is directed to research methods in cognitive psychology but the findings, and reflections on appropriate methodology, seem very relevant to knowledge elicitation for expert systems. Verbal protocol analysis has also been used to investigate various aspects of the functional behaviour of librarians (Ingwersen 1982). One disadvantage of protocol analysis is that having to verbalise may interfere with what the expert is doing (Berry and Broadbent 1984). Verbal protocol analysis is particularly inappropriate where the problem solving requires the expert to engage in cooperative dialogue with a client, for example, in the user-intermediary pre-online interaction, where it would be difficult, if not impossible, for the expert to carry out an additional verbal task.

2.7.3 Observational Studies

Observational studies involve observing and recording the behaviour of the expert as she/he works on a real problem, in his/her normal working environment, in as unobtrusive way as possible. This technique is extremely time-consuming and requires complex, indepth analysis (Welbank 1983). However, they do have the advantage that they can be used to discover what the expert actually does and are useful for extracting information about the role of the expert and the ordering of tasks. One example of the use of observational techniques in the expert system context is the investigation into student-advisor interactions carried out by Coombs and Alty, with a view to developing an expert system advisor (Alty and Coombs 1980; Coombs and Alty 1980; Coombs and Alty 1984).

Where a problem solving process involves a cooperative interaction between human expert and a client through the medium of a dialogue, observational methods may prove valuable and the most appropriate. Both parties must communicate their understanding of the problem being tackled, the constraints of the situation, what they themselves know about the problem domain and ways in which the problem could be solved. Usually it is also necessary for the expert to explain to the client what is being done, or could be done, and why. By recording such interactions, this information becomes available to the knowledge engineer.

2.7.4 Machine Induction

Machine induction is the generation of rules by a machine on the basis of a set of case studies presented to it. Several programs have been developed which automatically derive rules from case studies and examples (Quinlan 1979; Michalski, Carbonell and Mitchell 1983). Machine induction can produce a rule-base very quickly, and it always accounts completely for all the cases in the database. However, the effective use of such a technique requires a highly structured, well-defined problem domain and a large library of case studies.

2.8 Knowledge Elicitation for the IR domain

Current thinking tends towards the use of multiple techniques, with certain methods seen as more appropriate for particular types of knowledge (Gammack and Young 1985; Kidd 1985). In other contexts, this has been done for some time, eg, the detailed study of librarian-user interaction carried out by Ingwersen and Kaae (1980) made use of both verbal protocol analysis and interviews.

The methodology used in this research concentrates on observational techniques supported by interviews. The methods are defined as observational since they involve recording the expert at work on real problems in her normal environment. The subsequent analysis of the recordings needs to be conducted at the discourse level since by using discourse analysis techniques it is possible to go beyond an analysis of broad concepts and a simple functional analysis. Discourse analysis can provide a microlevel of representation which will allow complex, complete and accurate interface implementation. Supplementary interviews can be used to gain additional information, to clarify problem areas which may arise, and to allow the experts to comment on, and perhaps modify, the models which have been elicited through the discourse analysis. The methods used in this research are described in detail in Chapter 4.

Chapter 3

THE THEORETICAL CONTEXT: ASK AND THE MONSTRAT MODEL

3.1 Introduction

The ultimate goal of this research is to develop an intelligent interface for document retrieval systems, which is capable of simulating the functional behaviour of a competent human intermediary, ie, the interface should act intelligently in that it should be able to carry out the same tasks and functions as ~~of the~~ its human counterpart. The theoretical basis for this work derives from two earlier research projects, both of which are briefly described below.

3.2 The ASK project

The ASK project (Belkin, Oddy and Brooks 1982a, 1982b) was concerned with the issues of nonspecifiability of information requirements (see Chapter 1). The ASK hypothesis states that users engage in information seeking behaviour because they recognise that they have some gap or anomaly in their state of knowledge (an ASK). Since the user is not certain about something, she/he will probably be unable to state what sort of information is required to resolve the anomaly. However, most conventional information retrieval systems require users to precisely specify their requirements.

The major concern of the ASK project is to develop a framework in which the reasons for a person seeking information could be explicitly represented and used for information retrieval. Anomalies in a person's conceptual state of knowledge concerning a topic can be represented as a network of concepts and relations, with strength of association between concepts being indicated by distance from one another. These networks are elicited through the statistical analysis of texts of users' problem statements (Belkin, Oddy and Brooks 1982a, 1982b; Hapeshi and Belkin 1985), and are intended to be used for defining search strategies for texts which could respond directly to them.

The importance of the ASK project to this research is that it explores methods for representing user's *problems*, and reasons for seeking information, whereas most intelligent interface research in the IR domain has tended to concentrate on automating the processing of users' requests, whilst assuming the user is quite capable of actually formulating the request in the first place.

3.3 The MONSTRAT model and Distributed Expert Problem Treatment

Most intelligent interface design for IR systems has been based on two assumptions:

- that users can specify the information they require, and
- the tasks involved in transforming the user's request into a search strategy can be implemented as a linear sequence, with simple backtracking if necessary.

The fallacy of the first assumption has already been discussed , in this chapter, and in Chapters 1 and 2. Similarly, the linearity of event sequencing during human-human pre-search interviews has been refuted by various researchers, eg, Hitchingham 1979, Smith 1979, Cochrane 1981. Some tasks are more likely to be accomplished at certain stages in the dialogue, but it is usually hard to predict which tasks will precede or succeed certain other tasks. There are large amounts of backtracking, reiteration and side-sequencing.

The MONSTRAT model, which was developed by project INSTRAT (Belkin Seeger and Wersig 1983), attempted to identify what an intelligent interface should do, in order to understand the user and the user's situation and requirements, and thus provide her/him with a helpful response. The basic method used in that research was to audiotape interactions between intermediary/advisor and user/trainee in various kinds of information interactions, transcribe the data, and then code the utterances of both parties according to various criteria, including functions performed and knowledge needed to perform an utterance (Belkin, Seeger and Wersig 1983; Belkin 1984). The basic assumption behind this analysis was that any functional model should reflect the functional behaviour of a "good" human intermediary or advisor in an information provision mechanism.

The analysis succeeded in identifying a minimum number of functions which the interface must perform in order to understand enough about the user's problem and situation in order to make a reasonable response to it, and these functions are presented in Figure 3.1

Using the model, a number of information interactions were analysed and it was found that there was no simple sequencing of functions and that the functions themselves interacted in complex ways. It was therefore suggested that, in an intelligent interface, each function would be carried out by a separate "expert". All the experts are responsible for completing a specific function, yet are dependent on one another for information necessary for accomplishing their individual tasks. These individual tasks are seen as components of the overall single goal of providing an appropriate response to the user. This idea of distributed expert treatment builds on distributed problem solving approaches developed in AI research to handle complex problems.

A number of simulation experiments, using humans to carry out the tasks or functions, were carried out to investigate the general validity of the proposed model and to resolve the issue of architecture and control mechanisms (Belkin, Hennings and Seeger 1984). The experiments confirmed the validity of the overall functional model. The most effective control structure was the **blackboard** model, in which messages from experts are posted onto a "blackboard" where they can be read by other experts monitoring the blackboard. This type of control structure was used in the Hearsay II speech understanding system (Erman and Lesser 1979). The pattern of interactions observed in the simulation experiments closely followed the type of temporal patterning observed in the human-human information interactions. The sequencing of the functions was circumstance-driven and the activation of a function depended on output from several functions being available.

The power of the MONSTRAT model lies in its functional analysis of the behaviour of the information provision mechanism. Although an interface based on the model has yet to be implemented, some of the central ideas have been developed by Belkin and his colleagues (eg, Daniels, Brooks and Belkin 1985), and have been employed by Croft in his Expert Assistant (see Chapter 2).

FUNCTION NAME	DESCRIPTION
1. Problem State (PS)	Determine the position of the user in problem treatment process, eg, formulating the problem, problem well specified.
2. Problem Mode (PM)	Determine appropriate mechanism capability, eg, document retrieval.
3. User Model (UM)	Generate description of user type, goals, beliefs, knowledge, eg, graduate student, thesis.
4. Problem Description (PD)	Generate a description of the problem - the type, topic, structure, environment, etc.
5. Dialogue Mode (DM)	Determine the appropriate dialogue type and level for the situation, eg, menu, natural language.
6. Retrieval Strategy (RS)	Choose and apply appropriate retrieval strategies to knowledge resource.
7. Response Generator (RG)	Determine propositional structure of response to user which is appropriate to the situation.
8. Explanation (EX)	Describe mechanism of operation, restrictions etc. to the user as appropriate.
9. Input Analyst (IA)	Convert input from the user into structures usable by other functions.
10. Output Generator (OG)	Convert propositional response into a form appropriate to the user, situation and dialogue mode.

Figure 3.1: (After Belkin, Hennings and Seeger 1984). The functions of an intelligent interface for document retrieval systems.

3.4 User Modelling and Distributed Expert Problem Treatment

The theoretical basis for this study derives from the ASK and INSTRAT projects, and the ultimate goal of the overall research is to develop an intelligent interface for document retrieval systems which can simulate the functional behaviour of a human intermediary. In order to build an operational interface based on the model and assumptions described above and in the preceding chapters, a number of issues need to be tackled (Brooks, Daniels and Belkin 1985):

- Function specification: identification and detailed specification of the individual functions, at a level of detail which will allow implementation of that function.
- Identification and specification of the knowledge resources required for each function.
- Specification of the interactions among functions, again at a level of detail to allow construction of a communication and control architecture.
- Identification of the types of models necessary, especially the intermediary's models of various aspects of the user and her/his problem.
- Specification of the human-computer dialogue structure.

It is apparent that one of the functions that must be peformed by an information provision mechanism is that of user modelling (see Figure 3.1, this chapter). The aim of this thesis is to examine the user modelling function in detail, ie, to specify the components of the user model and the knowledge resources required by the intermediary in order to construct this model; to specify how the user model interacts with the other functions; to identify an appropriate formalism in which the user model may be represented in an eventual operational interface; and to undertake some form of validation of the adequacy of the user model. The investigation into the User Model will be carried out within the context of the distributed architecture proposed by the MONSTRAT model, ie, the User Model is just one of a number of distributed functions (or 'experts') which will probably be implemented within a blackboard system. The aim is **not** to implement the user model, let alone an entire intelligent interface: such a task is beyond the scope of this thesis. Instead, the emphasis will be on discovering what type of user model is required in the IR situation, and the methods necessary for accomplishing this.

Chapter 4

METHODS

4.1 Overview

The general method used in this research is the functional analysis of real-life dialogues between human experts (intermediaries) and clients (information system users) using detailed discourse analysis. This method provides a technique for the extremely detailed analysis of a very complex, naturally-occurring situation, in a way that is relatively unobtrusive and unconstrained by experimental bias and existing ideas. This method has been developed over a number of years by four team members, and aspects of this method have been described in various joint papers (Brooks and Belkin 1983, Daniels, Brooks and Belkin 1984, Belkin, Brooks and Daniels 1986), and in a recent thesis (Brooks 1986). The discourse analysis is supplemented, in this thesis, by interviews with the experts, ie, the intermediaries.

A number of audio-recordings were made of human intermediary-human user interactions at various search services at the University of London. The recorded dialogues were then transcribed onto paper, and the resulting transcriptions were partitioned into utterances and foci. Each utterance was categorised according to the task, or function, or goal that each party appeared to be performing. Also, an attempt was made to identify the knowledge resources being employed by the intermediary in carrying out his/her tasks.

The technique of discourse analysis provides an unobtrusive means of discovering what the participants are trying to achieve in the information interview, how they go about their tasks, and the resources they use. No third parties are involved at the time of recording nor is the intermediary required to do something other than what he/she normally does in the pre-online interaction with the user. That is, he/she is not forced to verbalise her/his thoughts, or stop to answer the knowledge engineer's questions.

4.2 The Research Team and its approach

The transcription, utterance and focus division, and coding were carried out by four team members: NJB, HMB, PJD and KS. One interview, Interview #4, was subjected to four separate analyses, one by each team member, each of whom attempted to identify and categorise the goal/s which were occurring in each utterance. This stabilised the inventory of goals.

On subsequent interviews, the work of each coder was independently checked by another member of the team. Coding was monitored so that a high degree of intercoder reliability could be maintained. Considerable care was taken with intercoder reliability and cross-checking. These are common procedures in psychological and sociological research, but there is little evidence of their use in the knowledge engineering literature.

4.3 Collecting the Recordings

4.3.1 The Search Centres

The collection of data took place at two academic online information retrieval services at the University of London with four different trained intermediaries, who carry out searches of bibliographic databases. The two search services were the library of the Institute of Education, and the Central Information Service (CIS), Senate House. These search centres are the longest established search services of the University, and CIS has been in existence since 1974.

The general procedure at both search centres is for a user to make a booking with the service in advance. At the time of the booking, the user is usually asked to give a brief description of his/her search problem, and this is either written on the booking form or related over the telephone. The search is usually carried out by the intermediary with the user present and participating, but occasionally searches are carried out on behalf of an absent user, with the assistance of an "intermediary" user, ie, someone who has agreed to attempt to relay to the intermediary details of the user, and the user's search topic, etc. However, intermediaries tend to discourage these types of "intermediary" searches, as a "second-hand" model of the user's requirements can result in an unsatisfactory search.

4.3.2 The Participants

The Users

In general, the users tend to be mainly postgraduate students, academic staff and research staff. At the Institute of Education, postgraduate students of all levels use the service. The Institute encourages students to utilise the service by its policy of subsidising student searches. At CIS, few users are students, as these would normally tend to make use of the services within their own institutions. CIS carries out searches for University of London users, whose college or institute library does not have a search service, or where the local service does not cover the subject area of the search. The service is also used by visiting academics and research workers outside the University, and is occasionally used by nonacademic users from industrial and commercial organisations.

The Intermediaries

The core data set consists of seven interviews, involving seven different users and four intermediaries. Two intermediaries are employed at CIS, and the other two work at the Institute of Education library. Of the four, two had more than five years experience at the time of recording, and one had more than nine years.

4.3.3 The Recordings

The intermediaries at both search services were asked for their co-operation, and the purpose of the research project was briefly explained to each user before obtaining their consent to record the interview. There were no refusals.

The users were asked for their permission immediately they entered the room where the search was to be conducted, and the microphones were attached to both user and intermediary as soon as both were ready to begin discussing the search problem. However, invariably it was impossible to capture on tape the first few minutes of casual conversation between the participants, which frequently included some discussion of the user, his/her current problem, etc. In five cases, recording ceased at the point when the intermediary logged onto the system, but in two of the interviews recording continued until the search was complete. A print-out of the online search process is available on file for each of the searches whose pre-search interviews were recorded at CIS.

Recording no.	Title of Search	Intermediary	Search Centre
#4	Community Education in Africa	A	I of E
#5	Japanese learning English	B	I of E
040684hba	Activated Carbon	C	CIS
120684hba	Forestry in Canada	C	CIS
190684hba	Greek-Turkish relations	D	CIS
260684ksa	Lung Cancer and Vitamin A	D	CIS
290684ksa	Delusional Thinking	D	CIS

Table 4.1: Interviews transcribed and analysed
Key:-
CIS - Central Information Service
I of E - Institute of Education

Table 4.1 lists the recording which were analysed and transcribed; recordings 260684ksa and 290684ksa, which were made by K. Stinton (see Stinton 1984), continue through until the end of the online search.

4.4 Transcription of the Recordings

The recordings of the pre-search interviews were transcribed from the tapes according to a specific format which included means for representing breath pauses, silences and their duration, failed utterances and extralinguistic phenomena such as laughter, coughs, etc. Each transcription was checked by at least one other member of the team, thus aiming towards accuracy and consistency.

4.4.1 The Protocol

All text is typed in lower case, except for proper nouns, which have an upper case initial letter.

The speech of each participant is typed on alternate lines, thus enabling consecutive, simultaneous and interrupted speech sequences to be represented as such. The participant is indicated at the beginning of each line: I represents the Intermediary, and U represents the User. Where there are two users participating in the same interaction, U1 denotes the "dominant" user (ie, the user who interacted more frequently with the intermediary) and U2 denotes the "secondary" user.

Normal punctuation conventions such as commas, full-stops etc. are not used to demarcate grammatical sentences or part-sentences. Question marks are used to indicate a query where rising intonation marks a functional rather than a grammatical question.

Colloquialisms are indicated wherever possible, and apostrophes are used to indicate contractions. For example,

'cause (because) yeah (yes) an' (and) etc.

True phonetic transcription is not used. However, attempts are made to record what is actually said rather than the correct orthographic form of the word. Broken-off words are indicated by using a hyphen.

```
----------------------------------------------------------------------
I what's the subject of y- your query                     right (..)
U                                       Greek Turkish relations
----------------------------------------------------------------------
----------------------------------------------------------------------
I any- anything particularly (,) specific
U                                          actually I'm interested in
----------------------------------------------------------------------
I                                                        right (,)
U their (,) in their disputes (,) other than Cyprus (....)
----------------------------------------------------------------------
I disputes (...) (cough.) (....) other than Cy- are there any (,)
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I any particular ones (...)  you know any
U                   um                   the Agean (,) dispute (.....)
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U and the: their disputes over the treatment of (....) the (,) Turkish
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                                         right (....)
U minority in Greece (....) and the Greek minority in Turkey
----------------------------------------------------------------------
```

Figure 4.1: An extract from the transcript of interview 190684hba

Intonational stress is indicated by underlining the word or words concerned. Emphasised words are recorded as such because they may indicate concepts or terms of particular importance.

The drawn-out enunciation of words is to be indicated by the use of a colon after the word, which represents a duration of 0.5 seconds.

Inaudible speech is to be marked by "inaud" in parentheses together with the length of the sequence in 0.5 second intervals, as denoted by full-stops. For example, (inaud ....) represents two seconds of inaudible speech.

Extra-linguistic phenomena such as laughter, coughs, etc. are indicated inside parentheses. Again, the duration is denoted by full-stops, each of which represents 0.5 seconds.

Breath pauses (ie, when a continuous speech sequence is briefly interrupted by an intake of breath) are indicated by a comma in parentheses.

True pauses are marked by full-stops in parentheses and again, each full-stop indicates a silence of 0.5 seconds. Silences longer than 5 seconds are indicated by a numerical value in parentheses, eg, (8 secs).

A sample of the transcript from interview 190684hba is given in figure 4.1.

4.4.2 Identifying the Utterances

Introduction

The interviews were divided up into utterances, rather than normal grammatical sentences, because discourse analysis usually tends to use the utterance as the unit for analysis. Broadly speaking, an utterance can be defined as a speech sequence by one participant during the exchange. It may or may not consist of complete grammatical entities, and is terminated by a contribution made by another participant.

Two methods for utterance identification have been used: Method U-1 and Method U-2. Method U-1 is described in Brooks and Belkin (1983) and elaborated on by Price (1983). Method U-2 was developed by K. Stinton (Stinton 1984), and both methods have been described, compared and contrasted in Brooks (1986). Interviews #4, #5 and 040684hba were analysed using Method U-1. The remaining interviews were analysed using Method U-2. Each interview was analysed, and the utterances identified, by at least two of the four team members.

Method U-1

An utterance is defined as any continuous speech sequence by one participant in the dialogue, which is terminated by a contribution made by the other participant. If the contribution of one participant **takes the conversational turn**, the previous speech sequence is regarded as a completed utterance. The end of an utterance is indicated by an utterance number which reflects the chronological ordering.

```
    eg, xxxxxxxxxxxxxxxxxxxxx / (1)
                xxxxxxxxxx / (2)    xxxxxxxxxxxxxxxxxxxxxx / (3)
```

An utterance is NOT completed by overlapping speech that does not take the conversational turn: the overlapping sequence is to be treated as an utterance in its own right. Eg, from interview #4

```
----------------------------------------------------------------------------
I it just means that there are other headings    that we might be
U                                              mmm /(73)
----------------------------------------------------------------------------
----------------------------------------------------------------------------
I interested in /(70)
U
----------------------------------------------------------------------------
```

Nor is an utterance completed by a speech sequence uttered by participant B which partially overlaps that of participant A and which interrupts the sequence but does so in a breath pause or short silence (1-2 seconds). The sequence of participant A is treated as one utterance. Eg, from interview #4

```
----------------------------------------------------------------------------
I once we get to the sect- subject headings (,) that the err systems
U                                              uhum /(45)
----------------------------------------------------------------------------
----------------------------------------------------------------------------
I use /(47)
U
----------------------------------------------------------------------------
```

Each utterance may be divided into one or more **part-utterances**. A part-utterance is a sub-division of an utterance which appears to introduce a new function or topic. This sub-division may or may not be a grammatical entity, and is often indicated by structural discourse features such as silences and hesitations. Part-utterances are signalled by the addition of a lower-case letter to the utterance number. Eg, from interview #5

```
---------------------------------------------------------------------------
I right (.) /(32a) fear (....) inhibition (....) /(32b) ok emm (...)
U
---------------------------------------------------------------------------
---------------------------------------------------------------------------
I /(32c) thats about all from (,) that /(32d)
U                                            umm (.) now what shall we look
---------------------------------------------------------------------------
```

Method U-2

This is a simplified version of U-1. The partitioning of long utterances into part-utterances gave rise to some problems: it involved some initial *functional* analysis, and the subdivision assigned by different coders did not always agree.

In Method U-2, an utterance is terminated by **any** interruption from the other participant/s, despite the fact that this interruption may not necessarily take the conversational turn. Overlapping speech sequences which run completely in parallel and do not interrupt the first speaker, are **not** regarded as completing the ongoing utterance. Eg, from interview 190684hba

```
---------------------------------------------------------------------------
I what about books? (...) which we haven't talked much (,) but its's
U                                 I /(70)          I am fully
---------------------------------------------------------------------------
---------------------------------------------------------------------------
I not (,) /(69) good s'right (,) i- its really /(72)
U conversant with /(71)      cos there's so few of them really that
---------------------------------------------------------------------------
---------------------------------------------------------------------------
I             yeah /(74)
U come up (,) /(73)
---------------------------------------------------------------------------
```

This method ensures that most utterances are shorter and thus ensures that each utterance mainly deals with only one topic or subtopic. One problem is that the "sense" of one particular utterance is not always contained within itself, as it would have been in Method U-1, but is continued along into the next few utterances. Eg, from interview 290684ksa

```
---------------------------------------------------------------------------
I they've got paranoia psychosis (..) /(120) and they've got paranoid
U                               yeah /(121)
---------------------------------------------------------------------------
---------------------------------------------------------------------------
I personality (.) /(122) paranoid schizophrenia /(124)
U           (inaud.) /(123)                 oh God all of those are
---------------------------------------------------------------------------
---------------------------------------------------------------------------
I             are they? /(126)
U relevant /(125)
---------------------------------------------------------------------------
```

In this extract, the sense of utterance (120) is continued into utterance (122) and on into utterance (124).

4.4.3 Identifying the Focus Shifts

The focus of a dialogue (Grosz 1978) can be said to highlight that part of the mutual knowledge of the participants relevant at any given point in a dialogue, by grouping together those concepts or themes that are in the focus of attention. Focus influences what gets discussed, how different concepts get introduced, and how concepts are referenced. Focus influences the choices among different senses of the word, the interpretation of noun phrases and actions, and the overall interpretation of the utterance (Grosz 1978). The current focus is likely to dictate the structure of the discourse, and the topics to which reference can be made at any given moment.

Shifts in focus occur within a task-oriented dialogue whenever the subtask being carried out changes (Grosz 1978). An attempt to shift focus by a participant may be regarded as "illegal" if she/he has not been granted "permission" to do so by the other participant. Focus shifts tend to be initiated by the participant with the higher status.

The importance of focus in the overall structuring of dialogue, particularly in task-oriented dialogues, and the correspondence that has been found between focus shift and the completion of goals or task, appeared to indicate that focus, and focus shift, could prove to be a useful higher-level unit for the analysis of the dialogues. It seemed probable that foci could be seen as concentrating the participants' attention on the accomplishment of a particular goal, with shifts of focus moving attention from goal to goal across the problem structure.

An initial analysis revealed that shifts of focus were generally indicated by the occurrence of "frame words" (Sinclair and Coulthard 1975), often accompanied by pauses of varying duration. Frame words can be characterised as words or phrases which indicate that some kind of boundary has been reached within the discourse, and that therefore a shift of focus is about to occur. Examples of frame words include:-

```
Yeah (.) ok (.) em (...)  (Interview #4)

Right ok (,) right (..) (Interview 190684hba)

now the next thing is to get (,) our strategy
      (Interview 190684hba)
```

The interviews were analysed, and partitioned, for focus and focus shift. Again, each transcript was analysed for focus by at least two team members working independently. Each focus was assigned a category on the basis of the functions or goals with which it appeared to be mainly concerned. The inventory of focus categories, and their distribution, is given in Daniels, Brooks and Belkin (1985).

4.5 Analysis of the Transcript

4.5.1 Introduction

The overall objectives of the analysis were to discover the goals, or tasks, or functions which take place during the information interaction, to identify the knowledge resources employed by the intermediary to accomplish these tasks, and to analyse how the User Model subfunctions interacted both with each other, and with the other interface functions.

4.5.2 Specifying the Functions

The overall goal of the analysis is to categorise the utterances according to both the functions and the subfunctions. An abstract analysis of the interviews at an early stage in the project led to a list of broad functions, and these functions are very similar to the functions identified in the MONSTRAT project. The discourse analysis was carried out in order to identify the detailed subfunctions: detailed analysis at this level was not carried out in the MONSTRAT project.

Each of the interviews was subjected to a detailed utterance-by-utterance analysis in order to identify and categorise the subfunctions or subgoals that each party appears to be trying to achieve, before the higher level goal of constructing an appropriate search formulation can be attained. Two interviews, 120684hba and #4, were subjected to separate, independent analysis by three team members (NJB, PJD and HMB). The results of these two analyses stabilised the inventory of subfunctions, each of the remaining five interviews was then analysed by at least two people independently. The level of agreement achieved was very high. For example, two analyses of interview 120684hba were compared in detail and the disagreements on subfunctions noted. Out of 108 utterances or part-utterances categorised, there was disagreement on 4 (3.7%). A complete agreed subfunction analysis for interview 120684hba is presented in Appendix B. The results were cumulated for all seven interactions. At this point, the final inventory of subfunctions was achieved.

It should be emphasized that the identification of the subfunctions through the technique of discourse analysis is **not** dependent on the functions of the MONSTRAT model. The technique was based initially on a set of functions for information interaction derived from an abstract analysis of the information situation, and from the research team's experience and observation of information interaction, but the functions and subfunctions were empirically confirmed through the discourse analysis itself. For these reasons, the inventory of functions identified differs somewhat from those listed in Figure 3.1.

A full inventory of functions and subfunctions together with the notation used, is given in Figure 4.5. The subfunctions include a category of metagoals which relate, not to an individual function, but to the participants' overall view of the interaction.

An extract from a transcript, the third focus of Interview #4, together with its subfunction analysis, are presented in figure 4.2 and table 4.2 respectively.

utterance 30 This frame word signals the end of the second focus.

utterance 31 The intermediary tries to establish what types of documents the user would like to retrieve from the search.

utterance 32 The user states that she would be satisfied with documents relating to other countries, as long as they are relevant to her search topic.

utterance 33 In this utterance, the intermediary refers to both the user's search topic, and the kind of documents that the user wishes to obtain.

utterance 34 Utterances such as these have been classified as **phatic communication**.

utterance 35 In the first part of this utterance, the user discusses an aspect of her research (she is looking at history as well), and then proceeds to define the background subject area to the search topic (ie, that community education was introduced to Africa during colonial times).

utterance 36 The intermediary proceeds to elicit further information about the general subject area of the search.

On the basis of this analysis, the following subfunctions for these utterances were assigned.

```
TOPIC ; DOCS    - two subfunctions are present in this
                  utterance, and the subfunctions
                  appear to be achieved simulaneously,
                  ie, there is no clear chronological
                  ordering of them.
```

```
--------------------------------------------------------------------------
I three countries /(29)    but, are , you prepared to read stuff about
U                     uhmhmm /(30)
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I I dunno ya know, anywhere else in the world, i::f /(31)
U                                                      yeah, if it's
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I                                      if it's yeah if it's about
U related to the questions I'm asking /(32)
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I community education but primarily in Africa (.) a::nd (.) and then
U                              mm/(34)
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I those three countries /(33)
U                           then I'm looking at (.) a history as well,
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I
U because these three (..) the reason I chose them is because they've
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I
U got a history of community education (,) which was introduced during
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I                         What (.) they actually called it community
U colonial times (...) /(35)
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I education (.) or /(36)
U
--------------------------------------------------------------------------
```

Figure 4.2: An extract from the transcript for interview #4

PARTICIPANT	UTTERANCE	SUBFUNCTION
U	30	Frame word
I	31	DOCS
U	32	DOCS
I	33	TOPIC;DOCS
U	34	ph
U	35	RES..SUBJ
I	36	SUBJ

Table 4.2: Subfunction analysis for the third focus of interview #4.

Utterance Number	Subfunction
3	USER
4	USER : CONFIRM
5	USER
6	USER
7	EXPLAIN
8	EXPLAIN : CONFIRM

Table 4.3: Subfunction analysis for the extract from interview 190684hba

```
RES .. SUBJ      - these two subfunctions are being
                   achieved by the utterance, and the
                   first one is accomplished before the
                   second.

MATCH : TERMS    - although a subfunction combination of
                   this type does not appear in the
                   above excerpt,  this notation
                   indicates that the second subfunction
                   qualifies the first.  For example,
                   the intermediary checks that the
                   search terms that she thinks
                   appropriate coincide with those that
                   the user considers to be
                   important.

ph               - phatic communication.
```

An example of a complete transcript, with utterance and focus divisions, can be seen in Appendix A. Foci and subfunction analysis for the transcript is presented in Appendix B.

4.5.3 The Interaction Maps

Most of the subfunctions appeared to contribute to a particular, higher-level function and could be grouped accordingly. The User Model, for example, appeared to consist of five subfunctions. The pattern of interaction between the subfunctions of the user modelling function, and the interaction between the user modelling subfunctions and the other functions, could be identified both through the discourse analysis of the transcripts and through the interviews with the intermediaries (see section 4.7, this chapter). These interactions can be diagrammed on a focus-by-focus basis, and are known as **interaction maps**.

An example of an interaction map for the User Model is presented in Appendix C: it is a representation of the performance of the user modelling subfunctions for all of interview 120684hba, divided by focus, and indicating to what functions or subfunctions they contributed information, and from what other functions and subfunctions they require information.

A short example from an interaction map for the User Model is presented below, encompassing focus 1 of interview 190684hba, in figure 4.4. The extract from this transcript is in figure 4.3, and table 4.3 presents the accompanying subfunction analysis.

4.5.4 Identifying the Knowledge Resources

Having identified the subfunctions associated with the higher level functions, it was necessary to identify the knowledge resources which the intermediary needs to access in order to achieve

```
---------------------------------------------------------------------
I now I gather you're (cough.) excuse me you're a visitor /(3)
U                                                       yes I am /(4)
---------------------------------------------------------------------
---------------------------------------------------------------------
I yes are you part of the university or /(5)
U                                      well I teach at a Canadian
---------------------------------------------------------------------
---------------------------------------------------------------------
I           ya (,) um I I jus' (,) we ask you this because i- its
U university /(6)
---------------------------------------------------------------------
---------------------------------------------------------------------
I awful to bring up charges straight away (laugh.) but just so that
U
---------------------------------------------------------------------
---------------------------------------------------------------------
I you know (,) you know that its a ten pound basic an its (inaud.....)
U                                                             yes /(8)
---------------------------------------------------------------------
---------------------------------------------------------------------
I /(7) right ok (,)
U
---------------------------------------------------------------------
```

Figure 4.3: An extract from the transcript for interview 190684hba

```
---------------------------------------------------------------------
USER     UGOAL     KNOW     IRS     BACK     I/O   OTHER
                                                      FUNCTIONS
---------------------------------------------------------------------
visitor -------------------------------------> EXPLAIN (charges are
                                               higher for visitors)

teaching
 -staff--------> greater subject
                 knowledge than
                 student

Canadian
 university----------------------------------> PREVREF (likely to have
                                               accessed dbs available in
                                               Canada)
---------------------------------------------------------------------
---------------------------------------------------------------------
```

Figure 4.4: Interaction Map for the User Modelling function for focus 1, interview 190684hba

FUNCTION	SUBFUNCTION	EXPLANATION OF SUBFUNCTION
USER MODEL	USER	Determine the status of the user
	UGOAL	Determine the user's goals
	KNOW	Determine the user's knowledge of the field
	IRS	Determine the user's familiarity with IR systems
	BACK	Describe the user's background
PROBLEM MODE	CAPAB	Explain the capabilities of the system to the user
PROBLEM STATE	PREV	Determine user's previous reference activities
	PREVNON	Determine user's previous non-reference activities
	PDIM	Determine the problem dimension
PROBLEM DESCRIPTION	SUBJ	Define the background subject area of the search
	RES	Specify the content of the user's research
	TOPIC	Specify the topic of the search
	DOCS	Determine the content or description of the documents the user would like to retrieve
	SLIT	Determine the formal characteristics of the subject literature
RETRIEVAL STRATEGIES	TERMS	Select terms for searching
	QUERY	Formulate the query
	STRAT	Evolve the search strategy
	DB	Select the database to be searched
RESPONSE GENERATOR	OUT	Select the output requirements
EXPLAIN	EXPL	Bring the user's knowledge up to minimum level necessary for user to cooperate usefully
	INFORM	Explain intermediary's intentions to user
	DISPLAY	Literal display of some aspect of the system
METAGOALS	MATCH	Compare models that participants hold
	CONFIRM	Agree models that participants hold
	PLAN	Specify the plan of the interview

Figure 4.5: The Functions and Subfunctions

the functions and subfunctions during the pre-online information interaction. Utterances relating to the user modelling subfunctions were analysed in detail, and inferences made about the types of knowledge which the intermediary appeared to be using. For example,

```
------------------------------------------------------------------------
I now I gather you're (cough.) excuse me you're a visitor /(3) (.)/(3)
U                                                                 yes I
------------------------------------------------------------------------
------------------------------------------------------------------------
I       yes are you part of the university or /(5)
U am /(4)
------------------------------------------------------------------------
```

(taken from 190684hba)

In the above example, the intermediary's question indicates that she possesses some internal knowledge regarding the potential range of categories according to which users can be classified. Here, the intermediary is concerned with establishing the user's status, ie, does the user belong to London university, or is she a visitor?

```
------------------------------------------------------------------------
I if you start to think of- (.) of sources really of information
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I there's the (.) Commonwealth Agricultural Bureaux (.) /(70) nd- are
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I you familiar with their (.) publications? (.) /(73)
U                                             no (.) /(74)
------------------------------------------------------------------------
```

(taken from 120684hba)

Here, the intermediary needs to access subject-specific knowledge concerning the use of secondary sources. As well as possessing knowledge concerning the levels of experience that various types of users may have, the intermediary must also be able to ascertain which types of sources a user is likely to have consulted, given that he/she is working in a particular subject/topic area. In this instance, the user is working on a thesis concerned with forestry, and the intermediary considers that CAB publications would be appropriate for this general subject area.

4.6 Additional data for the User Model: The Problem Statements

4.6.1 Introduction

As part of a separate, but related, project (Belkin and Kwasnik 1986), approximately fifty recordings of problem statements and pre-online information interactions were collected at one of the search services: the Institute of Education library. These were considered to be a useful supplementary data source for the user modelling function, and were analysed in order to discover whether the same user modelling subfunctions could be identified, and also whether any further subfunctions would be necessary to account for additional user modelling data.

Sentence number	User modelling subfunction
2	UGOAL
7	UGOAL
8	KNOW
9	UGOAL
12	USER . . KNOW
13	KNOW
16	UGOAL . . KNOW

Table 4.4: User modelling subfunctions for PS34

4.6.2 The Problem Statements

As the users entered the terminal room, they were asked if they would agree to participate in the study. Once their permission had been obtained, the research worker asked them to speak unprompted for a few minutes, in response to the following set of questions:

```
Tell me the general background to your research, the
specific problem that's led you to come here to have a
search carried out, and any further information that
might be specific to this search.
```

The problem statements were later transcribed, and an example of one problem statement, PS34, is presented in figure 4.6.

As these problem statements are monologues, rather than dialogues, they could not be formally be partitioned into utterances, and part-utterances. However, for the purposes of this thesis, each sentence was assigned a number, so that utterances concerned with user modelling could be assigned. The problem statements were analysed for user modelling subfunctions, and the results were cumulated with those from the discourse analysis. The user modelling subfunctions identified for PS34 are presented in table 4.4.

4.6.3 The Recordings of the Interactions

After the user's problem statement had been collected, the research worker went on to record the information interaction that then took place between the user and the intermediary. Although the intermediary had, in most cases, heard the user's problem statement, the ensuing dialogue did not appear to differ greatly from those interviews which were **not** preceded by a problem statement. Usually, the intermediary appeared to find it necessary to reiterate those questions which were used to elicit the problem statement, but in a different order and with rather different wording. This does seem to suggest that perhaps intermediaries have a general "plan" according to which they try to conduct the interviews, and that they prefer to elicit information in particular formats and at specific stages in the interview.

A further twenty-five recordings to accompany the problem statements were collected in this manner, but, unlike the core data set of seven transcripts, they were not transcribed. However, they were listened to and analysed in order to discover further user modelling categories, and to confirm the existing User Model subfunctions. Once again, the results were cumulated. The results of the analysis of the problem statements will not be reported in detail in the following chapter. Instead, a short description will be presented, indicating how the results of this analysis expanded and confirmed the user modelling subfunctions elicited by the discourse analysis.

The general field is cognitive development and the topic is metacognition /(1) And I need to know if there is any longitudinal study in this field, because I'm thinking of doing that: two years' longitudinal study. /(2) And as far as I know - all of the works I know - don't have anything till now, but I want to be sure about that./(3)

And the other thing I want to know about: the tests they use, especially Americans, because this is a special American study. /(4) And for that I need to know what they are doing now recently. /(5) I could find a lot of things in the library. /(6) I want to complement all of that; I have plenty of work.../(7)

This is the main thing: I don't have very clearly now my general hypothesis. /(8) I'm looking for that also and to do that I must know what they are doing, you know, very recently. /(9) And this is why I'm looking for this search information. /(10)

you couldn't tell us more about the background to the research?

No, because I'm just starting; I started on the ninth of January. /(11) I know because I did the Masters here that I'm looking ... that I know something about metacognition, metacomprehension, metalinguistics, all those metas, you know. /(12) But I don't have very clear the idea. /(13) I know what I want to do here, because I never saw anyone doing ... following a child during a time. /(14)

The other thing is I'm pretty sure that there are a lot of differences between children, and how they process information. /(15) And this is the other thing I want to look but I don't have very sure how I'm going to approach this; the general pattern I don't have very clear till now. /(16)

Figure 4.6: Transcript of Problem Statement 34

4.7 Interviewing the Informants

4.7.1 Introduction

Current thinking tends towards the use of multiple techniques in knowledge elicitation, with certain methods seen as more appropriate for particular types of knowledge (Gammack and Young 1985; Kidd 1985). In this project, the primary method of eliciting the required knowledge was through the functional discourse analysis of the recorded information interactions, but, in the case of the user modelling function, this method was supplemented by interviews with the experts, that is, the intermediaries.

4.7.2 The Questionnaire

The questionnaire (see Appendix E for complete set of questions) was designed for three broad purposes:

- to verify the subfunctions and knowledge resources associated with the User Model
- to obtain information and clarification on various problematic areas
- to obtain further information both on how the subfunctions of the User Model interact with each other, and also how the User Model interacts with the other interface functions, such as Problem Description, Response Generator, etc. How does the intermediary use information from the User Model? Conversely, can the other functions provide information that can be used by the User Model to formulate hypotheses and provide values?

4.7.3 The Informants

Interviews were carried out with three intermediaries, two of whom worked at the search services from which the data had been collected. The third intermediary to be interviewed worked in the Department of Information Science at the City University and was not involved in any way with this research project. The purpose of this third interview was to discover whether the user modelling function and its subfunctions are relevant and necessary in other academic online information retrieval services, ie, that the User Model is independent of the data on which it is based.

4.7.4 The Interviews

Each interview took between one and a half and two hours to complete. Each question to the intermediary was accompanied by a concrete example in the form of an excerpt from a transcript or from an audio-recording of a real-life information interaction between user and intermediary.

Some examples of the questions are listed below:

(1f) Is it useful to know whether a user is a Fellow, or holds a Research Fellowship?

Many problems concerned the level of detail necessary for complete specification of the subfunctions. For example, it was clear from the data that academic staff tend to fall into one of two categories, Teaching or Research (see Chapter 5), whereas Postgraduate Students are classified initially as either Taught Students or Research Students. Further analysis of more data revealed that several Academic users were research **Fellows**, and it was unclear whether a separate category should be created for such users.

(2c) Is it useful to know whether a user is carrying out the search in order to complete a **literature review**?

The above type of question was designed to elicit specific information regarding how user modelling subfunctions are used by the intermediary both to formulate hypotheses about the other functions and to provide information about the other parts of the User Model.

(3b) Do you tend to distinguish between a user's knowledge of the search **topic** and her/his knowledge of the general **subject** area?

The above distinction was apparent in the analysis of the transcripts. However, it was not clear to the research team why such a distinction should be useful or necessary, and the goal of question 3b was to discover whether the intermediary does make a conscious distinction and, if so, why.

Chapter 5

RESULTS

5.1 Introduction

5.1.1 The Transcripts of the user/intermediary interviews

The transcripts were the principal data source for this study, and were used to identify the following:

- the subfunctions of the User Model
- the knowledge resources of the User Model
- the interaction of the user modelling subfunctions with other functions and subfunctions

The interview transcripts are referred to throughout this work by their number, for example, Interview 120684hba, in order to distinguish them from the interviews with the intermediaries, (see section 5.1.3 below) which are referred to as simply "the intermediary interviews", or the "interviews with the intermediaries".

5.1.2 The Problem Statements

The problem statements, and their accompanying recorded interactions, were used as an additional data source to back up the results obtained from the transcripts. No further user modelling subfunctions were identified as a result of this analysis; rather, some further detailed classifications **within** subfunctions were noted, and these results have been incorporated into the overall results presented below.

5.1.3 The Interviews with the intermediaries

Once again, the interviews with the intermediaries were undertaken to provide an additional data source, and the results were cumulated with those obtained from the analysis of the transcripts and the problem statements. No further user modelling subfunctions were revealed and in fact, some detailed classifications of information within subfunctions were removed when it was found that the intermediaries themselves did not make such detailed distinctions. For example, the intermediaries stated that they do not make a distinction between the user's knowledge on her **topic** and her knowledge on the general **subject** area. Furthermore, the interviews provided some particularly valuable, detailed information concerning the interaction of the User Model with the other functions and subfunctions, and these results have been incorporated into Section 5.5, this chapter.

The third interview was conducted with an intermediary from a different search service, in the Department of Information Science at The City University, London, ie, a search service which had **not** participated at all in this research project. The purpose of this interview was to attempt to demonstrate that the User Model is not an artefact of the transcript analysis and the other two interviews. This aim was fulfilled: no new subfunctions were uncovered, and the interactions identified by the City University intermediary were in broad agreement with those identified both in the transcripts, and by the other two intermediaries.

5.2 Function Specification

The user modelling function appears to be a complex, multi-component function whose task is to construct models of relevant aspects of the user's personal background, goals, experience and knowledge. User modelling knowledge is both elicited by the intermediary through direct and indirect questioning of the user, and is also volunteered by the user, often in response to questions which appear to concern another function. Broadly speaking, user modelling generally occurs at the beginning and the end of the dialogues, with perhaps a smaller amount occurring at other stages of the dialogue. User modelling occurs much less frequently than Problem Description (see Brooks 1986), and possible reasons for this finding are discussed in Section 5.6, this chapter.

Intermediaries frequently start the information interview with a very general, open-ended question and the users may choose to reply to this with a statement concerned with user modelling, for example,

(Taken from Interview 120684hba)

```
-------------------------------------------------------------------------
I what's the problem? /(1)
U                    ok I'm just beginning a research project (.) /(2)
-------------------------------------------------------------------------
I
U err (.) I'm a research student at LSE (....)/(4a)
-------------------------------------------------------------------------
```

On the other hand, an intermediary may begin the interaction with an explicit question aimed at the construction of a user model, possibly in order to generate an appropriate explanation, or perhaps because so much of the ensuing dialogue is dependent on an implicit user model. For example,

(Taken from Interview 190684hba)

```
-------------------------------------------------------------------------
I now I gather you're (cough.) excuse me you're a visitor (.) /(3)
U
-------------------------------------------------------------------------
I yes are you part of the University or /(5)
U
-------------------------------------------------------------------------
```

Users often volunteer, unprompted, user modelling information, perhaps because they perceive this to be important to the overall goal of the interaction, and possibly because other aspects of the information problem can be more easily placed in some sort of context if some reference is made to the user's own personal traits, characteristics and experience. For example,

(Taken from Interview 190684hba)

```
-------------------------------------------------------------------------
I
U I might add that I've done a similar search down in Canada (.) at
-------------------------------------------------------------------------
I
```

```
U my university /(25)
--------------------------------------------------------------------------
```

(Taken from Interview #5)

```
--------------------------------------------------------------------------
I
U I have done a research project /(2b)
--------------------------------------------------------------------------
```

Analysis of the data revealed that different utterances seemed to be concerned with different aspects of the user, that is, the User Model does not appear to be a unitary, unidimensional model. The goal of some utterances is to establish the **status** of the user, for example, whether the user is a student or a member of staff:

(Taken from Interview 120684hba)

```
--------------------------------------------------------------------------
I
U I'm a research student at LSE /(4a)
--------------------------------------------------------------------------
```

(Taken from Interview 190684hba)

```
--------------------------------------------------------------------------
I
U well I teach at a Canadian university /(6)
--------------------------------------------------------------------------
```

Other aspects of a user's status may also include information relating to her/his source of funding, for example:

(Taken from PS 41)

```
I'm a Commonwealth Relations Trust Fellow funded by the Governments
of Australia and the United Kingdom...
```

It is often useful to know the position of the user, ie, is the user an end user, an intermediary user searching on behalf of someone else, or a team member, for example:

(Taken from Interview 040684hba)

```
--------------------------------------------------------------------------
I
U I'm not actually writing the report myself you see.../(72c)
--------------------------------------------------------------------------
```

Information referring to a dimension of **time** may also be relevant to a user's status: an academic user, for example, may be a former, a current or a prospective MA student:

(Taken from 190684hba)

```
--------------------------------------------------------------------------
I
U I used to be a student here (,) some years back /(336)
--------------------------------------------------------------------------
```

The overall goal of the information provision mechanism is to provide the user with some information which will be helpful in managing her/his problem and in the case of a document retrieval system, both the intermediary and the user co-operate in order to retrieve a set of useful bibliographic references. However, users come to the system with a wide range of individual goals, and a crucial task of the intermediary is to discover the user's goals and objectives, so that these may be considered when formulating the search strategy, etc. Users themselves realise the importance of their goals to the overall success of the interaction, and will often volunteer information of their own accord, for example:

(Taken from PS 22)

```
I would like to make, if I can, any contribution, any proposals from
the solutions furnished towards absolute solution of the problem (ie,
educated unemployment)
```

The above user has an ambitious, long-term goal! Other users have more specific intentions:

(Taken from PS 64)

```
I want to prepare lectures for the students..
```

(Taken from Interview 040684hba)

```
------------------------------------------------------------------------
I
U we're just starting to write a report../(28a)
------------------------------------------------------------------------
```

(Taken from Interview #4)

```
------------------------------------------------------------------------
I it is your dissertation /(3b)
U                                yeah MA dissertation /(4)
------------------------------------------------------------------------
```

Specific intentions, such as the submission of a thesis, give rise to other goals which lead to the search being carried out, for example:

(Taken from PS 34)

```
I'm thinking of doing ... two years' longitudinal study..
```

(Taken from Interview 260684ksa)

```
------------------------------------------------------------------------
I and you want to do a review of the literature.../(5)
U
------------------------------------------------------------------------
```

(Taken from Interview #5)

```
------------------------------------------------------------------------
I
U Now what I want to do is a pre and post questionnaire /(4b)
------------------------------------------------------------------------
```

(Taken from Interview 040684hba)

```
------------------------------------------------------------------------
I
U it's mainly a search to back up our own information.../(2b)
------------------------------------------------------------------------
```

At the time of the search itself, users invariably have very specific, short-term goals relating to the current search, eg, a particular type of output may be required:

(Taken from Interview 260684ksa)

```
--------------------------------------------------------------------------
I
U I want to pick up (,) any work that's been done with (,) relatively
--------------------------------------------------------------------------
I
U low vitamin A levels rather than (.) just frank deficiency /(26)
--------------------------------------------------------------------------
```

(Taken from Interview 190684hba)

```
--------------------------------------------------------------------------
I
U I mentioned the possibility of limiting the search to British (,)
--------------------------------------------------------------------------
I
U journals and magazines /(48)
--------------------------------------------------------------------------
```

Another goal relating to the current search may be simply to "get into" the subject area and/or the literature, in other words, to do a preliminary search before serious work on the project gets under way:
(Taken from PS 65)

```
this is sort of a stab at what is actually around..
```

Another group of utterances which would appear to be concerned with user modelling, are those relating to the user's knowledge of both the topic and/or subject matter of the search, and the information sources in which appropriate references are likely to be found. Intermediaries rarely ask **direct** questions relating to a user's knowledge of her topic (possibly because this could be rather embarrassing to a user who is rather ignorant), but nevertheless manage to elicit the information they need by other means, for example,
(Taken from Interview #4)

```
--------------------------------------------------------------------------
I can you give me a definition of community education just so I'm
U
--------------------------------------------------------------------------
I clear /(16)
U       I can't that's the thing /(17a)
--------------------------------------------------------------------------
```

(Taken from Interview 120684hba)

```
--------------------------------------------------------------------------
I but you- you- you've got an idea of what sort of things go on
U
--------------------------------------------------------------------------
I (laugh.) in the forest as it were (.) /(32)
U
--------------------------------------------------------------------------
```

Although users generally appear to be reticent about their knowledge on the topic/subject area, there are exceptions, for example:
(Taken from PS 67)

```
what I know is very bitty you know..
```

(Taken from Interview 040684hba)

```
------------------------------------------------------------------------
I
U I have noth- no idea what that is at all /(17b)
                                       (evaporative emission control)
------------------------------------------------------------------------
```

When users have a **good** knowledge of their topic, however, they are usually happy to volunteer this information! For example,

(Taken from Interview 120684hba)

```
------------------------------------------------------------------------
I
U so I (.) have a pretty good handle on what kinds of recreation /(22)
------------------------------------------------------------------------
```

Intermediaries consider that it is important to find out whether a user has a good knowledge of the appropriate information sources:

(Taken from Interview 120684hba)

```
------------------------------------------------------------------------
I there's the (.) Commonwealth Agricultural Bureaux (.) /(70) -nd are
U
------------------------------------------------------------------------
I you familiar with their (.) publications ? /(73)
U
------------------------------------------------------------------------
```

(Taken from Interview 190684hba)

```
------------------------------------------------------------------------
I
U I am fully conversant with (books) /(71)
------------------------------------------------------------------------
```

Intermediaries (unless they have carried out previous searches for the user) invariably inquire about a user's familiarity with information retrieval systems. The intermediaries who were interviewed stated that most students do not have any online experience, with the possible exception of PhD students in the final stages of their research. Utterances concerned with eliciting a model of the user's IRS experience tend to occur during the later stages of the interview, probably so that the subsequent explanation of procedures etc. is still fresh in the user's memory when they log in and go online. For example,

(Taken from Interview #4)

```
------------------------------------------------------------------------
I have you done one of these things before? /(128d)
U                                          computer search? /(129)
------------------------------------------------------------------------
I computer search /(130)     no? /132)
U           computer search /(131)   no /(133)
------------------------------------------------------------------------
```

(Taken from Interview 290684ksa)

```
------------------------------------------------------------------------
I have you done one of these searches before at all? /(3)
U                             never /(4)          no /(5)
------------------------------------------------------------------------
```

If a user has had some degree of IRS experience, either at this service or at another service, the intermediary will usually try to elicit further details about this experience, for example, can the user remember the procedures:

(Taken from Interview #5)

```
----------------------------------------------------------------------
I now you- you have done a search before haven't you /(356a) so you
U
----------------------------------------------------------------------
I know the proc- the general procedure /(356b)
U                                       yeh I did /(357) I think
----------------------------------------------------------------------
I
U so...../(358)
----------------------------------------------------------------------
```

(Taken from Interview 190684hba)

```
----------------------------------------------------------------------
I were you present at the searches (,) when you did them in (,)
U
----------------------------------------------------------------------
I Canada? (,) um can you remember (inaud.) ok so you know what goes
U                  I was actually /(290)
----------------------------------------------------------------------
I on /(289)
U
----------------------------------------------------------------------
```

When a user appears to recall details of her/his previous search or searches, it is likely that the intermediary will ask which systems and/or databases were used, providing that this current search concerns a similar subject field. For example,

(Taken from Interview 190684hba)

```
----------------------------------------------------------------------
I
U I might also add that I've done a similar search down in Canada (,)
----------------------------------------------------------------------
I                     ah (,) I wondered about that when you said you
U at my university /(25)
----------------------------------------------------------------------
I were from Canada (,) /(27) which't was using the Dialog system was
U                    yes /(29)
----------------------------------------------------------------------
I it (....) ca- can you remember what um databases you searched? /(30)
U
----------------------------------------------------------------------
I                 yeah /(32)        sounds like Dialog (,) /(33) yeah
U was an American based database /(31)
----------------------------------------------------------------------
I um can you remember which particular files (,) you searched.../(35)
U
----------------------------------------------------------------------
```

More experienced users may volunteer this type of information, as they are already aware that by doing this, they can assist the intermediary in choosing appropriate retrieval strategies. For example,

(Taken from Interview 040684hba)

I

U Prompt /(44a) I've tr- I have used before an it's quite good /(44b)

Although intermediaries rarely ask for details about various aspects of the user's background (with one exception: see below), users do occasionally refer to what they would appear to consider relevant aspects, apparently in order to explain or justify either other aspects of their user model or other functions and subfunctions (see section 5.5.3, this chapter). During the interviews, the intermediaries commented that although they tend not to ask for such details, unsolicited information about a user's background may be useful in a variey of ways.

Users may refer to their employment background:

(Taken from PS 64)

I'm an administrator

(Taken from Interview 120684hba)

I

U I spent four years with the Government as the co-ordinator of

---

I

U outdoor recreation.../(20)

The user's normal place of residence is another aspect of their background, for example:

(Taken from PS 30)

..I lived there (Java) for three years and I feel that I have a lot of first-hand experience of their position, and so I'm interested for that reason..

(Taken from Interview 120684hba)

I

U no it's just because I live there (Nova Scotia) /(58b)

In the above example (Interview 120684hba), utterance 58b was in response to the intermediary's question as to why the user is interested in Nova Scotia in particular. In both examples above, it is evident that the user is referring to his/her residence background in order to explain his/her interest in the topic, and to presumably justify a good state of knowledge on the subject area.

Users may refer to their academic backgrounds primarily for two reasons. In the first case, they may wish to explain that their academic training bears little, or no, relevance to the subject field of their current search topic, for example:

(Taken from PS 16)

I'm looking for um possibly accounts by medical officers and doctors that might have been published in obscure journals that I as a non medically trained person if you like wouldn't know about.

(Taken from PS 41) *The user's search topic is The use of microcomputers in classrooms*

```
My background is in humanities, so probably computer assisted
learning material that applies to geography and history would be of
more use to me personally.
```

In the second case, a user's academic training may have taken place many years ago which may imply that this user is unsure of recent developments in the subject area, or is perhaps having some difficulty in formulating the problem:

(Taken from PS 26)

```
Also my own particular problem, which is that I haven't been at
university for fifteen years and so I am still finding my way
around..
```

The above aspects of a user's background may be volunteered by the users, because they feel that they will facilitate the intermediary's understanding of the problem: they are rarely explicitly elicited by the intermediary. However, intermediaries **do** often ask about certain "temporary" features of the user's background, such as the user's departure date from the UK (in the case of an overseas student) and the maximum budget that she/he has available to spend on the search, for example:

(Taken from Interview 190684hba)

```
--------------------------------------------------------------------------
I yeah the thing is how long are you staying now? /(301)
U                                                    I'm leaving on
--------------------------------------------------------------------------
I
U Friday /(302)
--------------------------------------------------------------------------
```

(Taken from Interview 290684ksa)

```
--------------------------------------------------------------------------
I have you got any sort of financial (..) budget here you know you
U
--------------------------------------------------------------------------
I want to stick within......? /(164)
U
--------------------------------------------------------------------------
```

It seems then that the user modelling function, whose overall goal is to build a model of the user, has five subgoals, each of which is associated with a particular subfunction, referred to by its acronym:

USER determine the status of the user

UGOAL determine the user's goals

KNOW determine the user's state of knowledge in the field

IRS determine the user's familiarity with information retrieval systems

BACK determine relevant features of the user's background

Interview Number of Dialogue	% Contribution of UM utterances to Dialogue	Number of UM Utterances in Dialogue	Total Number of Utterances
260684ksa	4.4	5	113
290684ksa	8.2	15	184
#5	1.4	6	419
#4	2.4	8	331
190684hba	6.6	27	408
040684hba	14.9	13	87
120684hba	13.9	15	108

Table 5.1: Percentage Contribution of the User Model to each Interview

Interview Number	USER %	n	UGOAL %	n	KNOW %	n	IRS %	n	BACK %	n	Total Number of Utterances
260684ksa	0	0	40	2	20	1	40	2	0	0	113
290684ksa	0	0	20	3	7	1	40	6	33	5	184
#5	0	0	17	1	0	0	83	5	0	0	419
#4	0	0	25	2	12	1	63	5	0	0	331
190684hba	18	5	33	9	19	5	15	4	15	4	408
040684hba	15	2	31	4	23	3	31	4	0	0	87
120684hba	20	3	20	3	53	8	0	0	7	1	108

Table 5.2: Percentage Contribution of the User Modelling Subfunctions to the overall User Model in each Interview (n = number of utterances relating to user modelling subfunction).

5.3 The Frequency of the User Modelling Function and its Subfunctions

The extent to which user modelling occurs in each dialogue varies between interviews. Table 5.1 presents the percentage contribution of the User Model to each interaction. Some dialogues were partitioned into part-utterances as well as complete utterances (see Chapter 4): as each part-utterance is usually assigned a different subfunction in the transcript analysis, part-utterances has been counted as individual utterances for the purposes of the tables presented below. User modelling appears to be achieved by carrying out some, or all, of the User Model's five subfunctions, but the balance between the different elements, the extent to which particular subfunctions occur, varies considerably between interactions. A comparison between the occurrence of the user modelling subfunctions in the seven interviews is detailed in Table 5.2.

It can be seen from Table 5.1 that the User Model appears relatively infrequently in each interview, and possible explanations for this are discussed below in Section 5.6. Although the extent to which the individual subfunctions occur varies between the interviews, it is apparent that the BACK subfunction occurs least frequently, probably because intermediaries rarely ask for this type of information, although users may occasionally volunteer it. The IRS subfunction appears to be dominant: it is clearly important to ensure that the user understands the system capabilities and procedures. The UGOAL subfunction alone appears in all of the interactions, reflecting the importance of this subfunction to both user and intermediary, and to the overall success of the dialogue.

5.4 The Knowledge Resources of the User Model

Unlike the Problem Description function (eg, Brooks 1986), which employs both *internal* knowledge resources (ie, the intermediary's own internalised knowledge) and *external* resources (ie, information on pre-search booking forms, thesauri and manuals), the User Model relies primarily on detailed *internal* knowledge which tends to be mainly categorical and qualitative.

In order to fulfill the USER subfunction, the intermediary must be aware of the potential range of categories according to which users can be classified, and this will vary according to the particular information service in which the intermediary is operating. Within the information services used in this research project, it seems that users tend to be classified along three broad dimensions. Firstly, the **type** of user is determined, ie, is the user an academic or a non-academic, secondly, the **position** of the user is determined, ie, is the user an end user, an intermediary user or a team member, and thirdly information may be elicited or volunteered regarding the user's **status**. An academic user, for example, may be a former, a current or a prospective Master's Degree student. In other words, **status** refers to a dimension of time. Users are of course classified according to many more specific subdivisions, for example, an academic user is attached to a particular **institution** and **department**, and is either a **student** or a member of **staff**, and the student category may be further subdivided into **undergraduate**, **postgraduate**, and so forth. The intermediary will also require some knowledge relating to institutions, departments, etc., which will probably vary according to where the search service in question is located. In this research project, for example, all of the search services which participated in the study were located in London and therefore the majority of users consulting those services were attached to London colleges. It was evident that the intermediaries had very good knowledge concerning the subjects offered by the various colleges, the courses run by particular departments, etc. It would have been immediately apparent, for instance, to the CIS intermediaries that the LSE does **not** run courses in zoology, and that students who are attached to Wye college are likely to be searching on an agriculture-related topic.

To accomplish the subfunction of UGOAL, intermediaries must access knowledge regarding the likely range of goals which particular categories of users have, and this knowledge is likely to use information provided by the USER subfunction. A user who is classified according to the USER subfunction as a **research student**, for example, is very likely to have a goal of completing a thesis or dissertation. Four categories of user goals have been identified, which relate both to the user's current information problem and to his/her personal background and more long-term plans and objectives. The first type of goal is called **current search goal** which refers to goals such as obtaining a specific type of output from the search, for example

(Taken from Interview 190684hba)

```
-----------------------------------------------------------------------
I
U my primary interest is is to get (,) references that are (,) /(331)
-----------------------------------------------------------------------
I
U that are British (,) that I am (,) not very likely to be able to get
-----------------------------------------------------------------------
I
U (,) while I'm in Canada (.) /(334)
-----------------------------------------------------------------------
```

The second type of goal is known as **goals leading to search**, ie, the goals that made the user realise that she/he needed a search carried out. This category includes goals such as **using a particular methodology**, eg,

(Taken from PS 51)

```
We've been asked by the World Bank to write a state-of-the-art
paper..
```

and reviewing the literature, eg, (Taken from Interview 260684ksa)

```
-----------------------------------------------------------------------
```

```
I and you want to do a review of the literature..../(5)
U
-------------------------------------------------------------------------
```

The third sub-division of goal is **specific intention**, which gives rise to the above category of goal. For example, a user may have the specific intention of completing a **thesis** or **writing a report**, for which a **current search goal** of obtaining a specific type of output is required, eg,

(Taken from Interview 290684ksa)

```
-------------------------------------------------------------------------
I
U ..I'm doing a PhD on paranoia (.) and delusional thinking../(12)
-------------------------------------------------------------------------
```

(Taken from Interview 120684hba)

```
-------------------------------------------------------------------------
I
U I'm err (.) /(4b) doing a (.) thesis (.) beginning a thesis../(4c)
-------------------------------------------------------------------------
```

The final category of goal has been termed **general goal**, which includes long-term goals such as **personal advancement goals**, eg, obtaining a certain type of job, getting on a particular course, and **idealistic goals**, such as offering general advice and potential solutions to wide-ranging problems, for example:

(Taken from PS 22)

```
to make .. any proposals ..towards absolute solution of the problem
(ie, unemployment in Cyprus)
```

(Taken from PS 64)

```
I'm looking into this with the aim perhaps of advising policy makers
on the possibilities and perhaps ways of accomplishing this task
```

The knowledge resources necessary for the accomplishment of the KNOW subfunction need to contain information regarding the **qualitative** assessment of a user's state of knowledge and also some knowledge related to the Problem Description function. This is necessary because the intermediary must be capable of decomposing a particular topic into its component parts, so that the depth or quality of the user's knowledge on each of these parts can be ascertained. These two aspects, or dimensions, of the user's knowledge have been termed **know-type** and **evalue**. Know-type refers to that part of the user's state of knowledge which is to be evaluated, and encompasses a range of subdivisions such as **part-topic**, **entire-topic**, **part-information-source**, and **all-information-sources**. Then, any one of these aspects may be evaluated according to the Evalue dimension, which includes a range of descriptive values, such as **none**, **good**, **complete** and **sufficient**.

For example, the user in Interview 190684hba has **no** knowledge of one **part-information-source**, ie, British magazines:

```
-------------------------------------------------------------------------
I
U I have no idea what British journal or or magazine anyway /(55)
-------------------------------------------------------------------------
```

On the other hand, he has a **complete** knowledge of another **part-information-source**, newspapers:

```
----------------------------------------------------------------------------
I
U I've covered the newspapers /(23)
----------------------------------------------------------------------------
```

The same user has a **very good** knowledge of, or at least appears to have a great deal of material on, one **part-topic**, Cyprus:

```
----------------------------------------------------------------------------
I
U I've tons of material on Cyprus.../(80)
----------------------------------------------------------------------------
```

The user in Interview #4 appears to have a very **poor** knowledge on the **entire-topic**, Community Education in Africa:

```
----------------------------------------------------------------------------
I can you give me a definition of community ed- just so I'm clear/(16)
U
----------------------------------------------------------------------------
I
U I can't that's the thing /(17a)
----------------------------------------------------------------------------
```

In order to fulfill the goal of the IRS subfunction, the intermediary needs to possess knowledge which concerns both the use of online resources, and which is also subject specific. In other words, the intermediary must be able to ascertain which types of sources a user is likely to have consulted, given that a user is working in a particular subject/topic area and also, in many cases, comes from another country. For example,

(Taken from Interview 190684hba)

```
----------------------------------------------------------------------------
I ah (,) I wondered about that when you said you were from Canada (,)
U                                                                  yes /(29)
----------------------------------------------------------------------------
I which't was using the Dialog system was it ... /(30)
U
----------------------------------------------------------------------------
```

Intermediaries also need to access knowledge concerning the levels of experience which various types of users may have, and will probably use information provided by the USER subfunction. For example, in the interviews the intermediaries commented that **students** rarely have any IRS experience at all, with the exception of finishing PhD students. Furthermore, some academic institutions offer IRS training and the knowledge resources need to contain information listing which institutions provide this type of training. Such information could possibly reside in the USER subfunction, where information is held concerning an academic user's **institution**.

It seems likely that the intermediary will initially try to assign one of two "values" to the user: either she/he has had **no** experience of Information Retrieval Systems, or else he/she has had **some** experience. If the user has had no experience at all, the intermediary will invariably explain IRS procedures etc. If the user has had some experience, the intermediary may proceed to elicit further details, for example, was the user present at the search (in which case she will probably remember the procedures and will require a much briefer explanation), which system and/or databases were used (and are they compatible!), was the user satisfied with the search, and so forth. In some cases, the user is already known to the intermediary, having had previous searches conducted and in that case, much of the explanation can be curtailed. For example,

(Taken from Interview 040684hba)

```
------------------------------------------------------------------------
I
U it's better than the ones I've brought (laughter) to you before
------------------------------------------------------------------------
I
U certainly /(23b)
------------------------------------------------------------------------
```

Knowledge concerning user background (BACK) will necessarily be extensive, although the potential range of user backgrounds will probably depend largely on the type of users consulting that particular information service. That is, the relevant aspects of the backgrounds of users who consult academic online information services are likely to differ from those users consulting industrial information systems. The analysis of the transcripts and the problem statements reveals that there are four aspects, or subdivisions, of a user's background which may be relevant in the information interaction. The first of these, **employ**, is concerned with the user's current, prospective or previous, employment, and the intermediary needs to make a judgment concerning the relevance of this employment experience to the search topic. For example, the user in Interview 120684hba (whose search topic is Recreation and Forestry in Canada) mentions that he worked for the Government in Nova Scotia as the co-ordinator of outdoor recreation, and the intermediary concludes:

```
------------------------------------------------------------------------
I you've done y- you've done the err um the practical experience/(60)
U
------------------------------------------------------------------------
```

The second aspect, entitled **residence**, establishes details about the user's usual place of residence, which again may be relevant to the current search topic. The user in PS 30, for example, is interested in the position of women in Java because she actually lived there for three years.

The third aspect of the user's background which may be of relevance is her/his **academic** background. Important features include the relevance of the user's academic training to the current search topic and/or subject field, and how long ago this training took place. For example, the user in PS 26 had not been at university for fifteen years, and appeared to be having some difficulty finding his "way around".

The fourth subdivision of the user's background is **temporary features**, and the intermediaries do sometimes ask about this aspect, especially if the user is visiting the UK from abroad. This aspect is concerned with establishing situation-specific details such as the user's departure date from the UK, for example,

(Taken from PS 64)

```
 I'm a foreign student who's only here for about two months..
```

(Taken from Interview 190684hba)

```
------------------------------------------------------------------------
I yeah the thing is how long are you staying now? /(301)
U                                                        I'm leaving on
------------------------------------------------------------------------
I
U Friday /(302)
------------------------------------------------------------------------
```

It may also be useful to know about the facilities to which the user has access, such as translation services, photocopying machines and other libraries, for example:

(Taken from Interview 190684hba)

```
------------------------------------------------------------------------
I
U I have access to the (,) LSE library and ... other libraries /(336)
------------------------------------------------------------------------
```

Finally, it is generally necessary to find out about **financial** features, such as to whom the invoice should be sent, and the maximum budget available to spend on the search.

5.5 Interaction of the User Model with the other functions and subfunctions

5.5.1 Introduction

The interaction between the User Model and the other functions can be analysed at several levels. One level is the pattern of focus shift, ie, identifying the types of foci which preceed and follow foci which are largely concerned with user modelling. At a more detailed level it is possible to examine the interaction which appears to occur between the individual subfunctions associated with particular functions, both within individual foci and throughout the dialogue as a whole.

5.5.2 The User Model and Focus Shift

One level of interaction between the User Model and the other interface functions is the pattern of focus shifts: it is possible to identify the types of foci which preceed and succeed those foci which are largely concerned with user modelling. Brooks (1986) examined the pattern change across foci which were concerned with Problem Description and Retrieval Strategies, and found that there were always at least two, and frequently many more, foci concerned with modelling a description of the user's problem in each of the seven transcripts. For example, Interview 190684hba has six foci concerned with Problem Description, and just one focus concerned with user modelling. It was therefore possible to carry out a meaningful analysis of the Problem Description function's interaction with the other functions on the focus level.

However, there are relatively few foci in the seven transcripts which can be classified as user modelling foci. Although user modelling certainly does occur (see the Tables in Section 5.3), it tends to be carried out by individual utterances which are distributed across the foci, rather than by whole foci. There are some foci concerned wholly with user modelling, and these mainly occur towards the beginning and the end of the interactions. At this focus level, it is not really possible to identify how the User Model interacts with the other functions, as it would seem meaningless to generalise on the basis of a very small number of occurrences of user modelling foci. However, it is possible to identify a few patterns of interaction in some of the dialogues.

In Interview 190684hba, focus 17 has been classified as a user modelling focus which is concerned with the subfunction UGOAL: the user has the goal of retrieving British references from the search, which cannot be obtained in Canada, his home country. The following focus, focus 18, is concerned with Retrieval Strategies: in particular, the intermediary attempts to formulate a query which will accommodate the user's goal.

In Interview 290684ksa, the first focus is one which is concerned with user modelling, in particular with the IRS subfunction. Although the second focus (PLAN) does not appear to be influenced by the IRS subfunction, the sixth focus is classified as EXPLAIN:IRS - it appears that it is not necessary for foci to be strictly consecutive in order for interaction between functions to take place.

The analysis of the transcripts and the interviews with the three intermediaries revealed that although the User Model does not interact, on the focus level, with the other functions to the same extent as, for example, Problem Description and Retrieval Strategies, the individual subfunctions of the User Model do interact with the other functions and subfunctions (including those of the User Model itself), and these types of interactions are discussed in the following sections below.

5.5.3 The Interaction of the subfunctions of the User Model with the other functions and their subfunctions

To investigate what sort of information is passed to the User Model from the other functions and subfunctions, and vice versa, and to see how the user modelling process is affected both by the state of the other functions and their models and by the user modelling subfunctions themselves, it is necessary to have a micro-level of analysis. The transcripts were analysed in detail, and interaction maps were produced for each interview, which represent the interactions between the user modelling subfunctions, and between the User Model and the other functions and subfunctions. On some occasions, these interactions were quite explicit, in that the intermediary and/or user made direct reference to them; on other occasions, the members of the research team hypothesised that such interactions were occurring, on the basis of cues implicit in the dialogue and their own experience of acting as intermediaries. In addition, the recordings accompanying the Problem Statements were listened to, although these were not transcribed onto paper. In some cases, only one occurrence of a certain interaction was identified in the transcripts and recordings, and therefore it does not seem appropriate to quantify these interactions statistically. Although care must be exercised in generalising from this type of data, the significance of even single occurrences of interactions should not be overlooked.

The three intermediary interviews provided particularly valuable and detailed information about the interactions between the user modelling subfunction and the other functions and subfunctions, and frequently served to confirm the interactions observed in the data. The intermediaries were able to specify how information from the user modelling subfunctions is used, frequently in a rule-based way, to formulate hypotheses or provide values both to the other user modelling subfunctions and the other interface functions such as Problem Description, Retrieval Strategies and Response Generator. The complete set of interview questions is presented in Appendix E, and some examples of questions which were designed to obtain information on the interaction of the user modelling subfunctions with the other functions and subfunctions, are presented below:

```
I. The USER subfunction
   (1a) Why is it useful to know which type of degree the user is
        doing?
   (1g) Why do you ask about the user's source of funding?
   (1i) Is it helpful to know from which institution the user comes?

II. The UGOAL subfunction
   (2c) Why is it useful to know that someone is doing a literature
        review?
   (2g) How do you use the information that someone is just wanting
        to do a ''preliminary search'' to see what is around?

III.The KNOW subfunction
   (3a) When a user apparently has very little knowledge about her
        search topic and/or subject area, what does this imply to
```

```
       you?
  (3c) Do you form hypotheses about a user's knowledge from
       information that you have received about her background?

IV. The IRS subfunction
  (4b) What are the reasons for finding out about a user's previous
       online searches?

V.  The BACK subfunction
  (5c) Is it useful to know about someone's current employment?
  (5f) What do you make of information about a user's place/country
       of residence?
  (5i) If a user's academic training took place a long time ago, do
       you form any particular hypotheses about their states of
       knowledge, their IRS experience, etc?
```

Input to the User Model from the other functions and subfunctions

It is interesting that there is relatively little input to the User Model from the other functions: information for the User Model rarely appears to be donated by the other functions, such as Problem Description, Response Generator, etc., but is primarily obtained through other means, for example, explicit questioning of the user, information volunteered by the user, nonverbal information gained throughout the course of the dialogue, and in a few cases, by the intermediary accessing the model that she had built for this user on previous occasions.

However, the analysis and the interviews revealed that the User Model does in fact use information generated by one other function, the Problem State function, which comprises three subfunctions:

PREVREF determine the user's previous reference activities

PREVNON determine the user's previous non reference activities

PDIM determine the problem dimension (ie, just **beginning** the work, **finishing** a thesis, etc)

Each of the above subfunctions appears to contribute to the user modelling subfunction, KNOW, and, to a lesser extent, to the IRS subfunction.

PREVREF --> KNOW

The intermediaries remarked that if a user has done some reference work before approaching the search service, then she/he will certainly have some knowledge of the topic and of the possible information sources. For example,

(Taken from Interview 190684hba)

```
----------------------------------------------------------------------
I
U I I've covered the newspapers /(23)
----------------------------------------------------------------------
```

In Interview #4, although the user claims that she has looked at British Education Index in the library, persistent and searching questioning by the intermediary revealed that she had only done so in a very cursory manner. This served to confirm her poor state of knowledge on her search topic, which was already apparent at the beginning of the interview in focus 1, utterance 17a.

PREVNON --> KNOW

Similarly, if a user has engaged in some kind of non-reference activities, for example, attending a course, carrying out a project, her/his knowledge of the topic is likely to be average or above. For example,

(Taken from Interview 120684hba)

```
------------------------------------------------------------------------
I
U I just had additional course in Middle Eastern politics (,) when
------------------------------------------------------------------------
I
U people say Middle East they generally exclude (,) Cyprus /(162)
------------------------------------------------------------------------
```

(Taken from Interview #5)

```
------------------------------------------------------------------------
I
U I have done a research project which suggests that ... /(2b)
------------------------------------------------------------------------
```

PDIM --> KNOW

The interviews with the intermediaries revealed that the user's knowledge of the topic increases as his/her work enters more advanced stages. Intermediaries tend to make assumptions about the user's states of knowledge from utterances concerned with PDIM, eg,

(Taken from Interview 040684hba)

```
------------------------------------------------------------------------
I
U we're just starting to write a report /(28a)
------------------------------------------------------------------------
```

This poor state of knowledge is in fact confirmed earlier on in the dialogue:

```
------------------------------------------------------------------------
I
U it's mainly a search to back up our own information which isn't very
------------------------------------------------------------------------
I
U great /(2b)
------------------------------------------------------------------------
```

The Interaction of the User Modelling subfunctions with each other

Not only does the User Model contribute information to, and use hypotheses generated by, the other functions, its own subfunctions (USER, UGOAL, KNOW, IRS and BACK) interact with each other and depend on information generated by one another in order to build up a complex, comprehensive model of the user.

USER --> KNOW

The intermediaries who were interviewed said that the user's status affected their perception of her/his state of knowledge, for example, a member of staff is likely to have a good knowledge of the subject area, whereas a beginning Master's degree student is likely to have a relatively poor knowledge of the subject area.

Furthermore, it was clear both from the analysis of the transcripts, and from the intermediary interviews, that a user's knowledge of the subject area varies according to whether he/she is actually an end user, an intermediary user carrying out a search for someone else, or is a team member. An intermediary user often has only a vague idea of the subject area, and the intermediaries interviewed revealed that they regarded such a user with misgivings, because he/she is often incapable of contributing usefully to the dialogue. For example,

(Taken from Interview 040684hba)

```
-------------------------------------------------------------------------
I
U I'm not altogether sure on that one /(72b) I'm not actually writing
-------------------------------------------------------------------------
I
U the report myself you see /(72c)
-------------------------------------------------------------------------
```

USER --> IRS

The intermediaries frequently use information from the USER subfunction to make hypotheses about a user's IRS experience. For example, they usually operate on the default assumption that a student, especially a Master's degree or Diploma/Certificate student, does not have any IRS experience. They also find it useful to know about a user's academic institution, because: (a) some institutions have their own IR service, and (b) some institutions offer IRS training to their students. Intermediaries also use their knowledge about the types of systems available in foreign countries to make deductions about the characteristics of a user's previous searches:

(Taken from Interview 190684hba)

```
-------------------------------------------------------------------------
I
U I might also add that I've done a similar search down in Canada (.)
-------------------------------------------------------------------------
I                  ah (.) I wondered about that when you said you said
U at my university /(25)                         yes /(28)
-------------------------------------------------------------------------
I you were from Canada /(27) which't was using the Dialog system was
U                          yes /(29)
-------------------------------------------------------------------------
I it ... /(30)
U
-------------------------------------------------------------------------
```

In the above example, the intermediary is accessing knowledge residing in the USER subfunction that this user's academic **institution** is in **Canada**, to form an IRS hypothesis that this user has used the Dialog system.

USER --> BACK

Information about the user's status may be used as input to the BACK subfunction, for example, a user who used to be a student at a certain institution probably has access to that institution's facilities:

(Taken from Interview 190684hba)

```
-------------------------------------------------------------------------
I
U I have access to the (.) LSE library and and you know (,) other
-------------------------------------------------------------------------
I
U libraries because I used to be a a student here (,) some years
-------------------------------------------------------------------------
I
U back /(336)
-------------------------------------------------------------------------
```

UGOAL --> USER

Information about a user's goals can lead to deductions about the user's status, for example:

(Taken from Interview #4)

```
-------------------------------------------------------------------------
I you're doing your dissertation /(3)
U                                yeah MA dissertation /(4)
-------------------------------------------------------------------------
```

In the above example, it is straightforward to deduce that, since the user has the **specific intention** of completing a dissertation, she is therefore an academic, a postgraduate student, etc. A further example can be found in Interview 290684ksa:

```
-------------------------------------------------------------------------
I
U yeah (.) well err I'm doing a PhD on paranoia (.) /(10)
-------------------------------------------------------------------------
```

KNOW --> UGOAL

In the following example, the information which the user has is inadequate and therefore one goal, which can be classified as a **goal leading to search**, is to fill out the gaps in this knowledge:

(Taken from Interview 040684hba)

```
-------------------------------------------------------------------------
I
U it's mainly a search to back up our own information which isn't
-------------------------------------------------------------------------
I
U very great (.) /(2b)
-------------------------------------------------------------------------
```

BACK --> KNOW

In the interviews conducted with the intermediaries, it was apparent that information relating to a user's background may be used to formulate hypotheses about her/his state of knowledge. For example, if a user's formal academic training took place many years ago (perhaps more than ten years, for example) then an intermediary is likely to conclude that he/she has a rather poor state of knowledge concerning current developments in the subject area. This was evident in the recording of the interaction which accompanied PS 41: the user was researching into the use of microcomputers in schools, and explained that since he did his teacher training in the 1950s and has recently been working in an administrative capacity, he has not had the opportunity to discover the revolution which has taken place in the field of microcomputers.

Similarly, if a user has an academic background which appears to be unrelated/irrelevant to her/his search topic, then her/his state of knowledge is unlikely to be as complete or as good as someone who does have a relevant academic training. For example,

(From PS 16)

```
I'm looking for um possibly accounts by medical officers and doctors
that might have been published in obscure journals that I as a
non-medically trained person if you like wouldn't know about
```

It is also apparent that users who have been working in a field which is directly related to their present search topic, tend to have a good knowledge of their subject area, for example,

(Taken from Interview 120684hba)

```
-----------------------------------------------------------------------
I
U I spent four years with the Government as the coordinator of outdoor
-----------------------------------------------------------------------
I
U recreation so I (.) /(20) have a pretty good handle on what kinds of
-----------------------------------------------------------------------
I
U recreation /(22)
-----------------------------------------------------------------------
```

This good state of knowledge is confirmed by the intermediary later in the same focus:

```
-----------------------------------------------------------------------
I but you- you- you've got an idea of what sort of things go on
U
-----------------------------------------------------------------------
I (laugh) in the forest as it were (.) /(32)
U                          yeah (.) /(33) yeah /(34)
-----------------------------------------------------------------------
```

BACK --> IRS

Once again, the interviews with the intermediaries revealed that if a user's academic training took place many years ago, then he/she is unlikely to have had any IRS experience, and will therefore need an explanation.

Output from the User Model to the other functions and subfunctions

It was evident, from the transcript analyses, the intermediary interviews and the recordings accompanying the problem statements, that the User Model is heavily utilized by all of the functions and their subfunctions, to varying extents. It should again be noted, however, that although some recurring patterns of interaction can be observed, some interactions may occur perhaps on one or two occasions only, and caution should be exercised in generalising. It does not seem appropriate to undertake statistical tests of significance on such a small data set, and this procedure would be complicated by the fact that there are three different sources of data: the transcripts themselves, the interviews with the experts, and the problem statements with their accompanying recordings.

UM --> PROBLEM STATE

Although no instances of this type of interaction appeared to be evident in the transcripts and recordings, the intermediaries reported that when a user has a poor state of knowledge (KNOW) on his/her subject matter, then it generally assumed that he/she is at the beginning (PDIM) of the problem formulation process.

UM --> PROBLEM MODE

The task of the Problem Mode function is to determine the appropriate mechanism capability, eg, document retrieval. There are few instances of this function in the interactions, probably because it has already been established that the document retrieval system is, in general, capable of producing an appropriate response for the user's problem. However, it appears that the document retrieval system cannot always cope with the user's problem, and the User Model can be a valuable source of information for the Problem Mode function. For example, the intermediaries agreed that if a user has **no** knowledge at all of the search topic and/or subject field, then an online search should not be undertaken: the user should be persuaded to go away and do some more work, perhaps returning to the seach service at a later date when she/he has a clearer idea of the problem area. Furthermore, some users have erroneous or unrealistic expectations of the capabilities of the system, which may be expressed in their current search goals, for example:

(Taken from Interview 190684hba)

```
------------------------------------------------------------------------
I
U I mentioned ... the possibility of limiting the search to British
------------------------------------------------------------------------
I                    that's not as easy as it seems /(49)
U (,) journals and magazines /(48)
------------------------------------------------------------------------
```

UM --> PROBLEM DESCRIPTION

In general, information from the User Model is used as input to the SUBJ subfunction (determine the subject area of the user's problem) and the RES subfunction (determine the background of the user's research) of the Problem Description function. For example, the intermediaries commented that information about the user's institution, which resides in the USER subfunction, may be used to form a clearer impression of the general subject area behind the search topic, and one intermediary cited the case of the student from Wye college (a London college specializing in agriculture) who was searching on a chemistry topic. The intermediary therefore deduced that the search should probably have an agricultural "bias", a consideration which would not have been apparent from information relating only to the search topic and subject area. It seems likely that this type of interaction, USER --> PD(SUBJ) was taking place in Interview 120684hba when the user, whose search topic is Forestry and Recreation, offered the information that he is attached to the Geography Department at the LSE. No doubt a user who attends the London School of Economics has a rather different perception of Forestry than a user who studies, for example, in a botany department.

Similarly, information relating to the user's country of residence (the BACK subfunction) may be used, according to the intermediaries interviewed, to influence their general perception of the subject matter (SUBJ) or perhaps to explain the background of the user's research (RES). An example of the BACK --> PD(SUBJ/RES) interaction may be seen in Interview 120684hba where, in utterances 49, 52 and 57 the intermediary asks the user why he is concentrating his research on Nova Scotia: is it, for example, because Nova Scotia has a long experience of providing recreational facilities? The user replies by referring to his residence background:

```
------------------------------------------------------------------------
I
U No /(58a) it's just because I live there /(58b)
------------------------------------------------------------------------
```

Users frequently have the current search goal of obtaining a specific type of output from the search, and this may be used by the DOCS subfunction (ie, determine the types of document which the user wishes to retrieve) of Problem Description, eg,

(Taken from Interview 190684hba)

```
------------------------------------------------------------------------
I
U I'm dying to find some harmonious aspects of their relationships
------------------------------------------------------------------------
I
U (inaud.) I mean if there are articles that deal with other (,)
------------------------------------------------------------------------
I
U with a relationship that don't warrant disputes (,) that's fine
------------------------------------------------------------------------
I
U by me too /(130)
------------------------------------------------------------------------
```

UM --> EXPLAIN

In six of the seven transcripts, the User Model explicitly provided input to the EXPLAIN function, an interaction which was strongly emphasized by the intermediaries as being of central importance to the interaction. The intermediaries use information residing in the USER subfunction to generate appropriate explanations regarding system charges, etc., and an example of this type of interaction may be seen in Interview 190684hba, where the intermediary asks whether the user is a visiting academic, or from the University of London, and proceeds with the following explanation:

```
------------------------------------------------------------------------
I we ask you this because i- it's awful to bring up  charges straight
U
------------------------------------------------------------------------
I away (laugh.) but just so that you know (,) you know that it's a
U
------------------------------------------------------------------------
I ten pound basic .../(7)
U
------------------------------------------------------------------------
```

In the above example, the user is in fact a visiting academic from abroad, and users who are **not** part of the University of London are required to pay increased charges for an online search.

In Interviews 120684hba, 190684hba, #4, #5, 260684ksa and 290684ksa, the IRS subfunction interacts with the EXPLAIN function: the intermediary determines how much IRS experience the user has, and then proceeds to deliver a tailored explanation, for example:

(Taken from #4)

```
------------------------------------------------------------------------
I have you done one of these things before? /(128d)
U                                         computer search? /129
------------------------------------------------------------------------
I computer search /(130)          no /(132)  the way it works is that
U            computer search /(131)  no /(133)
------------------------------------------------------------------------
I we ...(EXPLANATION) /(134)
U
------------------------------------------------------------------------
```

(Taken from Interview 260684ksa)

```
------------------------------------------------------------------------
I right have you done one of these searches before? (.) at all? /(108)
U
------------------------------------------------------------------------
I                                         (EXPLANATION)
U um (,) a long time ago (laugh...) /(109)
------------------------------------------------------------------------
```

UM --> RESPONSE GENERATOR

The goal of the Response Generator function is to determine the type of response which will be appropriate to the user and his/her situation. It currently has just one subfunction, OUT, whose goal is to establish the type and format of output which the user requires. This function appears to require information from three subfunctions of the User Model: USER, UGOAL and BACK.

The intermediaries commented that a user's status (USER) is important when considering the **amount** of output that a user is likely to require, for example, a Master's degree student will require fewer references than a PhD student, or a member of staff. Also, a student who is studying full-time is likely to be able to cope with far more references than a user who is studying part-time whilst continuing with her/his normal employment:

(Taken from the recording accompanying PS 30)

```
------------------------------------------------------------------------
I you're not three years full time PhD   so that puts a limit on
U                                        no
------------------------------------------------------------------------
I what we'd want to overwhelm you with..
U
------------------------------------------------------------------------
```

The user's goals (UGOAL) are frequently used to help to determine the type of output that may be required, for example, the intermediaries said that if a user has a goal of reviewing the literature (a **goal leading to the search**), then **comprehensive** output will usually be required. When a user has the **current search goal** of doing a **preliminary** search, then references to current literature should be retrieved, which are limited by date rather than by subject.

Aspects of the user's background (BACK) may be used in two ways to influence the Response Generator function. In the interviews, the intermediaries were asked if they ever used information relating to a user's background to determine the type of response that is required, and they gave the example of where the user's **residence** background may be used as input to the Response Generator. For example, a user whose normal place of residence is a foreign country is likely to wish to retrieve material which relates to that country. The user in Interview 120684hba, for instance, lives in Nova Scotia and that is where he is concentrating his research.

The other aspect of the BACK subfunction which is used is **temporary features**, and intermediaries do sometimes question the user about this. For example, if a user has to leave the country soon, there is no point in printing the references offline, as they make take several days to arrive:

(Taken from Interview 190684hba)

```
--------------------------------------------------------------------------
I how long are you staying now? /(301)
U                                         I'm leaving on Friday /(302)
--------------------------------------------------------------------------
I ah right so there's not much point I was going to say.....we could
U
--------------------------------------------------------------------------
I print them offline but that's not going to be any good to you /(303)
U
--------------------------------------------------------------------------
```

Furthermore, the intermediary may ask about **financial** aspects, for example, the maximum budget that the user has available to spend on the search: the more references that are printed out, the more expensive the search. In Interview 290684ksa, the user explains that he has about $15 available to spend on the search, and the intermediary responds:

```
--------------------------------------------------------------------------
I right so we- we'll stick round about there /(170)
U
--------------------------------------------------------------------------
```

UM --> RETRIEVAL STRATEGIES

The goal of the Retrieval Strategies function is to choose and apply appropriate retrieval strategies to the knowledge resource. It comprises four subfunctions:

(a) QUERY: formulate the query
(b) STRAT: determine appropriate search strategy
(c) TERMS: choose appropriate search terms
(d) DB : select the databases to be searched.

All of the five user modelling subfunctions appear to be used as input to Retrieval Strategies. According to the intermediaries interviewed, aspects of the USER subfunction influence the Retrieval Strategies function. For example, knowledge concerning the user's **department** will affect the choice of **databases** to be accessed: if a user is attached to a Psychology department, then Psychological Abstracts may be an appropriate choice. Another user aspect is **source of funds** : if a user is funding her/himself, the intermediary will attempt to speed up the search in order to save the user money!

The UGOAL subfunction may be used as input to the RS function. In particular, a user's goals may be used to suggest additional **terms** for searching: the names of certain methodologies (**goals leading to search**) such as **case studies**, **longitudinal studies** can be used in the search, and **general goals** may possibly suggest some terms. With regard to the goal of using a particular methodology, the intermediaries reported that they tend to adopt particular retrieval strategies to accommodate such goals, for example, a goal of **reviewing the literature** implies an exhaustive search of many databases, whereas a **case study** will require a detailed, specific search. A **state of the art** paper probably needs a search which covers a wide range of databases, and which accesses very up-to-date sources.

A user's goal should be taken into account when devising a search **strategy**: the user in Interview 190684hba has a particular type of goal which is not easily accommodated:

```
---------------------------------------------------------------------------
I
U my primary interest is is to get (,) references that are (.)
---------------------------------------------------------------------------
I
U British (,) /(334)
---------------------------------------------------------------------------
```

This particular goal will be difficult, if not impossible, to accomplish, and the intermediary responds, several utterances later:

```
---------------------------------------------------------------------------
I so well the only thing I I can suggest is that we did something
U
---------------------------------------------------------------------------
I but tried to get it just to (,) where wh- where they indexed it
U
---------------------------------------------------------------------------
I as place of publication being British (,) /(347)
U
---------------------------------------------------------------------------
```

A user's knowledge (KNOW) can also be used in various ways by the Retrieval Strategies function. For example, the intermediaries agreed that if a user has a **good** knowledge of her/his subject, then he/she can take the initiative in suggesting **terms**, whereas if her/his knowledge is **poor**, or non-existent, perhaps in the case of an **intermediary** user, the intermediary will have to suggest the terms, and "push" the user for additional terms. For example, in Interview #4, focus 11, the intermediary appeared to be hoping that the user would have used British Education Index to a sufficient extent to enable her to suggest terms, or at least to comment usefully on the intermediary's suggestions. However, the user's knowledge on her topic, Community Education, was poor and the intermediary was forced to suggest terms himself. Furthermore, when the user has a relatively poor state of knowledge on his/her subject matter, the intermediary will tend to search a limited set of databases, and try to retrieve review articles.

A **good** state of knowledge on a **part-topic** can influence the choice of **terms**: in Interview 120684hba, for example, the user has a good knowledge of a part-topic, **recreation**, and therefore there is no need to attempt to retrieve many references concerning this aspect of the topic. A further example may be found in Interview 190684hba, where the user does **not** wish to retrieve material concerning Cyprus:

```
---------------------------------------------------------------------------
I it's really Greek Turkish relations (..) /(77)
U                                             other than Cyprus because
---------------------------------------------------------------------------
I  yes /(79)
U  (,) /(78) I've tons of material on Cyprus..../(82)
---------------------------------------------------------------------------
```

Further on in the dialogue, the user again stresses this requirement:

```
--------------------------------------------------------------------------
I
U I wanna cut all that out because I have (...) a lot of material
--------------------------------------------------------------------------
I
U on th- innumerable situations like this (,) /(112)
--------------------------------------------------------------------------
```

Furthermore, when a user has a **good** or **complete** knowledge of a **part information source**, there is no need to access references from these sources:
(Taken from Interview 190684hba)

```
--------------------------------------------------------------------------
I you say you've covered newspapers haven't you (inaud.) won't worry
U
--------------------------------------------------------------------------
I about that (,) ... /(175)
U
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I
U I am fully conversant with (books) /(71)
--------------------------------------------------------------------------
```

The user's **IRS** experience is used by the **db** subfunction of the Retrieval Strategies function. The intermediaries who were interviewed stated that if a user has had a search carried out on a similar or related topic, then they will try not to repeat that search, but may update it and use the same databases if the user was satisfied with the outcome. For example, the intermediary in the recording which accompanied PS 23 attempts to find out which databases were used in a previous, related search, so that the same ones can be accessed again. In Interview 040684hba, the user takes the initiative in recommending a database which he has used before:

```
--------------------------------------------------------------------------
I
U Prompt (.)/(44a) I've tr- I have used before and its quite good/(44)
--------------------------------------------------------------------------
```

The intermediary in Interview 190684hba asks about previous database use so that she can either avoid them, or update:

```
--------------------------------------------------------------------------
I yeah um can you remember which particular files (,) you searched ..
U
--------------------------------------------------------------------------
I because i- if you wanted we could try and avoid those (,) or was it
U
--------------------------------------------------------------------------
I that you wanted an update? /(35)
U
--------------------------------------------------------------------------
```

Finally, the BACK subfunction may provide input to the **db** subfunction of Retrieval Strategies. The relevant aspect of BACK is **temporary features**: if an overseas user is due to return home shortly, then databases which can provide references easily obtained by the user must be accessed.

5.6 Discussion

Users approach an Information Provision Mechanism because they have an ASK (Anomalous State of Knowledge), ie, some kind of problem which may be more efficiently managed by obtaining appropriate information from the system. As the user cannot precisely specify what she/he needs to know in order to resolve the problem, the intermediary and user engage in a co-operative discourse, during which various goals are pursued, such as Describing the User's Problem (the goal of the Problem Description Function), or Formulating the Retrieval Strategies (the goal of the Retrieval Strategies Function). One of these goals is to build a User Model, and the goal of the User Model is to

> specify and identify relevant aspects of the user's personal background, goals, experience and knowledge.

User modelling would appear to be an important function of an intelligent interface for document retrieval systems: once the intermediary, whether human or automatic, begins to construct a User Model, it can begin to distinguish between its own beliefs and what it believes the user knows, and to model the user's situation in terms that might be useful in constructing the query and searching the database. A User Model is also essential for generating tailored explanations, for estimating the type of output that may be required, and for guiding the nature of the interaction between user and intermediary.

Most designers of intelligent interfaces for document retrieval systems have either omitted to include specific user modelling capabilities, or else have specified a restricted user model, containing limited types of information, for example, experience of information retrieval systems. More importantly, those researchers who have proposed user modelling capabilities for their system (eg, Fox and France 1986a, 1986b, Croft 1985) appear to have undertaken very little empirical investigation into real-life user modelling in the pre-online interaction.

The analysis of the interactions, supplemented by the interviews with the intermediaries, carried out in the research described in this thesis, reveals that intermediaries do engage in user modelling and that since users also volunteer information about themselves, they too must somehow sense that these kinds of models are important to the success of the dialogue. Furthermore, the User Model is a complex, multi-component model, which aims to model aspects of the user's status, goals, knowledge, IRS experience and background. As the interactions are goal-directed dialogues, it seems unlikely that intermediaries would elicit, or users offer, information which is not considered useful to the current task. Therefore it can be concluded that this information is necessary for a comprehensive user model, and contributes towards the overall success of the interaction.

The User Model in an academic search service environment comprises five subfunctions:

USER : determine the status of the user

UGOAL : determine the user's goals

KNOW : assess the user's state of knowledge

IRS : determine the user's familiarity with information retrieval systems

BACK : describe relevant aspects of the user's background

The extent to which utterances relevant to user modelling occur varies between interviews, with Interview 040684hba displaying the greatest amount of user modelling: 14.9% of the dialogue. It is interesting to note that the user modelling function does not appear to play a large role in any of the interviews, and various explanations may be suggested for this.

In at least one case, Interview 040684hba, the user is already known to the intermediary as she has carried out several searches for him on previous occasions. On this occasion, it is probably unnecessary to pursue the subfunctions of BACK and USER, as these, on the whole, refer to relatively long-term, enduring aspects of the user. However, the user's **goals** and **knowledge** will vary according to the current topic and tasks, and it is apparent that both of these subfunctions do appear in Interview 040684hba.

In another case, Interview #5; only 1.4% of the dialogue is concerned with user modelling, yet there is no apparent evidence that the intermediary knows, or remembers, this user. However, it is clear that the intermediary is building an **implicit** user model, ie, is **inferring** a user model, on the basis of information from other sources. It is clear, throughout the dialogue, that this user has experience of online searching, and has a very clear idea of her subject/topic area and of the information that may be required to resolve the problem. Four reasons for these deductions can be proposed:

- the user initiates many focus changes, indicating confidence and experience
- she uses specialized IR terms, such as "related terms" (utterance 205) with ease
- she herself suggests which databases should be accessed, eg, in utterance 10b
- she takes the initiative in suggesting terms herself, eg, utterance 13, "shall we start with the anxiety"

It can be seen, from the above example, that even though **explicit** user modelling may play only a limited part in the dialogue, intermediaries can **infer** a good deal of detailed user modelling information from other sources. This, of course, gives rise to the question: can an intelligent interface, ie, a machine component, be "taught" to recognize or infer this type of implicit modelling information from sources such as those described above? Moreover, will the effort expended in gaining such information be justified by the utility of the resulting user model?

A further reason why there is relatively little **explicit** user modelling in the interactions, is that intermediaries appear to be making use of **stereotypical** information in the dialogues (eg, see Rich 1979a, 1979b). In other words, once one or two salient facts have been elicited, many others can be deduced on the basis of those facts. For example, once it is known that a user is a **PhD student**, it can be inferred that she is also an **academic**, a **postgraduate**, has less **knowledge** than a staff member, has the **goal** of submitting of a thesis, and so forth.

Despite the fact that user modelling does not appear to a great extent in each interview, it should be stressed that it does nonetheless appear in every dialogue and the interviews with the intermediaries confirmed that some user modelling information is vital to the overall success of the interaction. However, the extent to which the subfunctions appear in the interviews varies considerably, and some subfunctions do not appear at all in some interactions. Although the small number of interviews in the data set means that statistical analysis of these types of phenomena is not possible, nevertheless the observed differences are important and may indicate the perceived importance of the subfunctions to the dialogue by both user and intermediary.

The IRS subfunction appeared in the majority of the interactions, and the intermediaries interviewed confirmed that it is important that the user has a clear idea of the capabilities of the system and the procedures involved in online searching. The UGOAL subfunction is the only one to appear in all seven transcripts, and appears to be vital to the success of the interaction. In Interview 190684hba, for example, the user's goals account for 33% of the user modelling utterances, and occur with regularity throughout the dialogue. The interaction is unsuccessful and the user is disappointed with the outcome, largely because the intermediary did not seem to pay sufficient attention to the user's goals, which would have been difficult to fulfill in any event. It seems obvious that every user will approach an information system with some sort of goal, irrespective of her/his other attributes and the type of system being consulted, and that it is crucial that the intermediary should elicit such goals in order to avoid disappointment and to assess the likelihood that these goals can be fulfilled by the system.

The KNOW subfunction appears in all dialogues, with the exception of Interview #5, where it can be clearly inferred (see above) that the user has a **good** state of knowledge on her topic/subject. Intermediaries rarely **explicitly** elicit information relating to a user's knowledge of her subject/topic, but may deduce the user's state of knowledge from models created by the other subfunctions, or may indirectly question the user (eg, see Interview #4).

The subfunctions BACK and USER appear relatively infrequently throughout the dialogue. The intermediaries interviewed confirmed the necessity of the USER subfunction but nevertheless, in four of the seven transcripts analysed, it did not appear at all. In Interview #4, there were no USER utterances, but the user's status could be deduced, in some detail, from the UGOAL of completing a MA dissertation. The USER subfunction contains a great deal of **hierarchically** organised information, and many deductions can be made by proceeding either up or down the "tree" or hierarchy. A similar situation prevails in Interview 290684ksa: there are no USER utterances, but USER information can be deduced from the user's **goal** of completing a PhD. It is possible that USER information may only be necessary when information concerning the user's goals or state of knowledge is not so readily apparent: USER information provides a focus for further questioning by the intermediary, and can be a valuable means of "getting into" the dialogue.

The BACK subfunction occurs least frequently and it has already been noted that, with the exception of information relating to **temporary features**, intermediaries rarely ask for these background details, although users may volunteer them in order to explain and justify their knowledge on the topic, for example, or their interest in the subject area. This finding was confirmed by the intermediary interviews: the intermediaries stated that they seldom ask for background information, but often find it to be useful if it is volunteered by a user.

To summarize, user modelling does not occur according to any pre-defined strategy or sequence, with the possible exception of the IRS subfunction, which is often to be found either at the beginning or the end of a dialogue. Highly individual user models are built for each user, with the proportion of subfunctions varying between users. It seems, therefore, that **dynamic** user modelling capabilities are required, which are sensitive to **individual** user needs and traits.

The interactions analysed, and the interviews conducted, in this research project all took place in similar environments (university search services), concerned similar types of problem, and dealt with largely homogeneous sets of users. Although the interview with the third intermediary confirmed, to some extent, that the results for user modelling appeared to be valid for **another**, ie, a different, academic search service, it may not be possible to generalise the detailed results to information interactions in **other** kinds of settings dealing with different users and different types of problem. However, although very little comparable work has been carried out on user modelling in other types of information systems, it seems reasonable to suggest that the five basic user modelling subfunctions may still be applicable in other systems, with perhaps adjustments being made to the individual categories **within** each subfunction, for example, the characteristic **academic background** within the BACK subfunction may be replaced by another, more appropriate characteristic for users in an industrial IRS. However, it seems likely that a model of a user's goals, in whatever context, will be required, and also a representation of the user's beliefs about, and experience with, the type of system with which he/she is interacting. It maybe be assumed that typically, the goal of every user approaching an information system is to retrieve sufficient information to enable him/her to manage that problem more efficiently, in the same way that it is assumed that a user in a CAI system wishes to learn something. Such an assumption is undoubtedly true, but detailed analysis of the data used in this research reveals that users frequently have several different types, or levels, of goals when interacting with the system.

In a different type of information system, it may be necessary to represent the KNOW subfunction in a rather broader manner, perhaps replacing it by "experience with topic", or "background in topic". Plexus (Vickery et al. 1986), for example, attempts to assess both the relationship between the topic of the query and the user's job, and the extent of the user's practical experience in the field: these two factors combined amount to the user's knowledge of the subject field. Furthermore, it is sometimes difficult to **directly** elicit information concerning a user's knowledge: it may be preferable to **infer** such information providing this can be done with reasonable accuracy, by accessing information residing in other subfunctions, such as BACK, and in other functions, such as Problem State.

Information relating to user status (USER) will probably be required in many types of information systems, enabling the system to form some initial stereotyped judgements based on a small set of facts. For example, a useful starting point for a medical information system may be to establish whether the user is a Consultant, a Registrar or a Student, and to determine the type of hospital or practice within which he/she is working.

Besides the specification of the User Model function itself, the knowledge resources it needs to employ have also been identified, both through the analysis of the transcripts and the interviews with the three intermediaries. Unlike the Problem Description function (Brooks 1986), the knowledge resources necessary for the accomplishment of the User Model are primarily internal rather than external, ie, the intermediary relies on her own internalised knowledge about users and their characteristics. It should be noted that the knowledge resources identified in Section 5.4, this chapter, are not exhaustive, and the resources for some subfunctions have been specified more completely than others. It is unlikely, for example, that any further subdivisions for the USER subfunction will be identified: the types of users who approach the information services studied in this research are finite, and the intermediaries interviewed considered that the level of detail in this subfunction is adequate and covers most eventualities. However, the subfunctions of UGOAL and BACK have proved rather more problematic. The difficulty is that these particular subfunctions employ more diverse knowledge resources and embody less restricted and categorical information which cannot be so easily categorised and organised. At present, it should be possible to claim only that the intermediary has a kind of empty structure or framework serving as the knowledge resources of the UGOAL subfunction, ie, she knows that users' goals can be classified according to four dimensions, but that each dimension may contain an unlimited number of possibilities, and that the user has therefore to provide the detailed input to this subfunction.

It should be noted that the user's contribution is essential; the intermediary is incapable of constructing an adequate user model without input, either direct or indirect, from the user. Admittedly, the intermediary may not always pay sufficient attention to some of the user's input, as in Interview 190684hba for example, when an important user goal is disregarded, but this is likely to lead to an unsatisfactory outcome to the search. The intermediaries seem to use their knowledge to organise the information supplied by the user and to help resolve ambiguities and to fill gaps, either by using "default" values (ie, values assumed to be true for certain categories of users) or by prompting the user for further information.

The User Model interacts with the other functions, and its subfunctions also interact with each other. This interaction was evident from both the analysis of the transcripts and the interviews with the intermediaries. An interesting result is that only one function, **Problem State**, donates information **to** the User Model, principally to the KNOW subfunction. It seems then that the User Model is **not** heavily dependent on the other functions for its own processing: it obtains the information it needs through other means, eg, explicit questioning of the user, information volunteered by the user, nonverbal communication, stored user models constructed on previous occasions, default values, etc.

The User Model is heavily used by all of the other functions, however, and it is possible to claim that this justifies the existence of the user modelling function. Although some interactions were found to occur on single occasions only, nevertheless the User Model was found to be used by Retrieval Strategies, Response Generator, Explain, Problem State, Problem Mode and Problem Description. The principal functions which interact with Retrieval Strategies are UGOAL and KNOW, and the Explain function is largely dependant on the IRS subfunction.

It is possible that further interactions may exist, but may be difficult to identify: much of the intermediary's knowledge may be unconscious and hard to articulate, a recurring problem for knowledge engineers.

It is evident that the User Model is accomplished by the user and intermediary **co-operating** together in a mixed-initiative, goal-directed dialogue: as much user modelling information is volunteered by the user as is elicited by the intermediary. Both parties contribute to the model-building process, and the user seems aware of the relevance and importance of her/his contribution.

The complex, multi-component nature of the user modelling function should also be stressed. The User Model comprises five subfunctions, each of which is concerned with accomplishing its own individual goal of describing the user's status, the user's goals, the user's knowledge, the user's IRS experience and the user's background. The User Model is **not** a single, unidimensional entity aimed at producing a single user description, and any design for a truly intelligent, co-operative interface for document retrieval systems should account for this. Furthermore, user modelling does not occur in a pre-determined, linear sequence; the way in which it is accomplished varies considerably between interviews, and the models constructed during the dialogue engage in complex patterns of interaction with each other.

Chapter 6

A FORMALISM FOR THE USER MODELLING FUNCTION

6.1 Introduction

> A central theme of research in AI is how to represent knowledge in such a way that it can be used effectively to perform tasks that have been regarded in the past as requiring human knowledge.
>
> (Steels and Campbell 1985)

Implementation of the User Model requires not only that the knowledge resources it employs be identified and specified, but also that a formalism be selected or developed to represent these knowledge resources. This is not a straightforward task as a wide range of types of knowledge are used, both in the User Model and in the other interface functions such as Problem Description. The chosen formalism must also be capable of representing the user model of each individual user which the system needs to construct. The skeleton of this model consists of the framework derived from the intermediary's existing, internal knowledge resources, and is filled out as the interaction proceeds with information elicited from, or volunteered by, the user and from the processing of the other functions or experts. Furthermore, the user modelling subfunctions interact both with each other and with the functions, providing information necessary for the other functions to carry out their own processing - it is therefore important to identify formalisms which cater adequately for the necessary interactions.

It is unlikely that a single representational formalism will be appropriate for all the models and functions in the intelligent interface (Brooks, Daniels and Belkin 1985). The analysis revealed that the various functions are quite different entities, using a wide range of knowledge resources and constructing models which differ not only in content but also in manner and scope. For example, the specific knowledge used by the user modelling function embodies a finite, restricted range of possibilities and tends to be qualitative rather than quantitative. In contrast, the knowledge needed by the Problem Description function appears much more diverse, and to be closely connected with the subject domain of the search topic (Brooks 1985).

Fikes and Kehler (1985) suggest that there are three basic criteria for a knowledge representation language:

(a) Expressive power: can experts communicate their knowledge effectively to the system?
(b) Understandibility: can experts understand what the system knows?
(c) Accessibility: can the system use the information it has been given?

They warn that none of the major knowledge representation languages is by itself able to satisfy all of these criteria.

6.2 Knowledge Representation Techniques

> In AI, a representation of knowledge is a combination of data structures and interpretive procedures that, if used in the right way in a program, will lead to "knowledgeable" behaviour.
>
> (Barr and Feigenbaum 1981)

Two important issues to consider when attempting to identify an appropriate knowledge representation formalism are Scope: what portion of the external world can be represented in the system? and Granularity (grain size): in what level of detail are objects and events to be represented, and how much of this detail is actually needed by the reasoning mechanisms?

There are a number of representational schemes which can be used in AI programs, and the principle ones are briefly described below.

(a) **Formal Logic.** Barr and Feigenbaum (1981) consider that the most important feature of logic and similar systems is that deductions are guaranteed correct to an extent that other representation schemes cannot attain. Logic was one of the first representation schemes used in AI. Predicate calulus is used to represent statements about specific objects and to postulate relationships between these objects. The problem with logic as a means of knowledge representation, in addition to being allegedly difficult to use, arises from the separation of representation and processing: logic concentrates on the representational issues but does not address the problems of processing, ie, how to **use** the information once it is represented.

(b) **Procedural Representation.** In this technique, knowledge about the world is contained in procedures, ie, small programs that know how to do specific things, how to proceed in well-specified situations. The underlying knowledge is not stated explicitly, and so is not easily extractable in a form that humans can readily understand. However, in contrast to formal logic representations, the concern of procedural representation is how the knowledge represented is actually used, and this means for representing this type of knowledge can be a considerable advantage.

(c) **Semantic Nets.** The semantic net, developed by Quillian (Quillian 1968) and others, was originally developed as an explicitly psychological model of human associative memory. Semantic nets are a set of knowledge representation techniques based on a common notation of labelled nodes, representing objects, concepts and events, and links (or arcs) between the nodes, representing the relations between them. A significant feature of semantic net representations is that important associations can be made explicitly and concisely: relevant facts about an object, situation etc. can be inferred from the nodes to which they are directly linked, without a search through a large database (Barr and Feigenbaum 1981). Partitioned semantic networks (Hendrix 1979) have been used to represent the Problem Description function (Brooks 1986), where each element of the Problem Description, eg, Search Topic, Research, etc. is represented in a different plane.

(d) **Production Systems.** The basic idea of production systems, developed by Newell and Simon (1972), is that the database consists of rules, called production rules, in the form of condition-action pairs. An advantage of this technique is that the conditions in which the rule is applicable are made explicit and, theoretically, the interactions between rules are minimized (Barr and Feigenbaum 1981). A production system formalism provides an easy and natural means of expressing certain important types of knowledge. Human experts frequently explain how they go about their tasks using 'if-then' type of rules (see Chapter 5), and these types of statement are naturally encoded into production rules.

(e) **Frames.**

> A frame is a data structure containing slots for items of information which in principle belong together in some context.
>
> (Campbell 1986)

In general, a frame (Minsky 1975) is a datastructure providing a framework within which new data are interpreted in terms of concepts acquired through previous experience. Frames include declarative, and usually procedural, information in predefined internal relations, and a major advantage is their ability to facilitate expectation-driven processing, ie, searching for entities that are expected based on the context within which someone is processing. Although Minsky's main goal for introducing frames was to direct the reasoning of scene-analysis systems, Fikes and Kehler (1985) point out that most of the subsequent work on frame-based systems has focussed on knowledge representation issues rather on the control of reasoning, which has been left to other parts of the system. However, some recent experience with frame-based systems in complex application domains has shown that frames can in fact play an important role in the control of reasoning components, and can assist the system designer in determining strategies for controlling a system's reasoning. The concept of frames will be described in further detail in Section 6.3.2 below.

6.3 Choosing an appropriate formalism for the User Model

6.3.1 Overview

The user modelling function encompasses five separate subfunctions, each of which needs recourse to particular knowledge resources. The User Model has a unity in that all of the subfunctions are concerned with identifying personal characteristics of the user, and these characteristics are **qualitatively** represented. The user contributes much of the user modelling information explicitly, and the intermediary provides the structure and the context. A formalism is required which will allow the structure of the intermediary's knowledge to be represented, and the user modelling information relating to individual users to be incorporated. The analysis and interviews have demonstrated that there is a considerable amount of "default" information about typical users already represented in the intermediary's knowledge structure, ie, information which the intermediary assumes to be valid unless overridden by explicit information gained from an individual user. Information to fill out the User Model can also be generated by the processing of the individual user modelling subfunctions and, less frequently, by the processing of the other interface experts. Furthermore, the other experts or functions use information and hypotheses generated by the User Model for their own processing, and the formalism chosen to represent the User Model must be able to allow for this type of interaction.

The analysis indicates that the user modelling knowledge employed by the intermediary is categorical in nature, embodying a restricted, finite range of possibilities, and is qualitative rather than quantitative. It is therefore likely that a standard set of frames can be derived for a particular type of search service or institution, embodying the user knowledge utilised by human intermediaries in that service. It seems quite probable that the five basic user modelling subfunctions will be applicable in most academic search services - perhaps even in other types of search service - and that individual modifications could be made to the core set of frames according to each individual service's own characteristics.

An important advantage of using a frame representation for the User Model is that rules, or procedures, can be attached to the slots in each frame to drive the problem-solving behaviour of the system. For example, a rule may be used to compute a slot value, using a hypothesis or information generated either by another user modelling subfunction or by one of the other functions. Another type of procedure could be used to post information generated by the User Model onto the Blackboard to be used by the other functions in their own processing.

Fikes and Kehler (1985) mention a further advantage of frames: they capture the way that experts typically think about much of their knowledge, provide an accurate structural representation of useful relations, and support a concise definition-by-specialisation technique that is easy for most domain experts to use.

Frames can also play a useful role in a system's reasoning behaviour by providing a kind of "automatic" inference capability: information stored in a frame which is higher up in the frame hierarchy can be inherited by frames which are further down the tree. Furthermore, the inclusion of **default** information is a common feature of most frame representations.

I have chosen frames as the appropriate formalism for the User Model for the reasons outlined above. I do not wish to make any strong claims for the validity of that decision: semantic nets have many features in common with frames, and it is quite possible that the User Model could be adequately represented by another representation formalism. However, in order to make some concrete progress towards the eventual implementation of the User Model (and, indeed, of the other experts) it was necessary to opt for one particular formalism.

6.3.2 Frames and Frame-based Systems

> There is abundant psychological evidence that people use a large, well-coordinated body of knowledge from previous experiences to interpret new situations in their everyday cognitive activity.
>
> (Bartlett 1932)

Frames are a way of representing knowledge about the objects and events **typical** to specific situations, objects and concepts. They were originally proposed by Minsky (1975) as a basis for understanding visual perception, natural language dialogues and other complex behaviours. **Scripts**, developed by Schank and Abelson (1977), are frame-like structures specifically designed for representing sequences of events.

According to Minsky (1975) a frame is "a remembered framework to be adapted to fit reality by changing details as necessary". Frames provide a framework within which new data are interpreted in terms of concepts acquired through previous experience. Frames facilitate **expectation driven** processing, ie, looking for entities that are expected based on the context within which one is processing. The representational mechanism that makes this possible is the **SLOT**, the place where knowledge fits within the larger context created by the frame. By supplying a place for knowledge, and thus creating the possibility of missing or incompletely specified knowledge, the slot mechanism permits reasoning based on seeking confirmation of expectations - "filling in slots".

A frame usually consists of a NAME and an AKO (a kind of) slot,which allows information about the higher level frame to be inherited by lower-level frames, much like an **ISA** link in semantic net representations. In other words, the value in the AKO slot points to the more general frame of which the current frame is a modification. In this sense, a frame system is a tree structure, in which generic information is "bumped" up the tree, with particular frames specifying new distinguishing knowledge. Each frame will usually have a set of **slots**, and each slot generally consists of a **slot-name** and a **value**, which may simply point to a lower-level frame, or which may be selected from a range of **terminal values**, ie, a bottom-level object, or a datum, which need not be a frame and cannot be specified further. In many cases, the value of a slot must be restricted to one of a specified range of frames or terminal values, although sometimes there may be no particular restrictions on the value which a slot can take. It should be noted that the frame-names and the slot-names themselves are not actually included in any of the reasoning processes. This data independence is a useful feature of frame-based systems: **any** slot-name, frame-name etc. can be chosen without affecting the system's reasoning behaviour. Many slots have a **default** value attached, which suggests a value for that slot unless there is contradictory or overriding evidence, ie, explicit information elicited from, or volunteered by, the user.

Restrictions, defaults, placeholders for values, procedures, etc. attached to slots are specified by attaching **facets** to slots. For example, a slot called FUR-TYPE in the frame CAT may have the following facets (prefaced by $):

```
[CAT
  AKO       $value    mammal
  FUR-TYPE  $value    ;this empty facet is waiting for the value
            $require  (black-and-white tabby tortoiseshell)
            $default black-and-white]
```

Default and inherited values are relatively inexpensive methods for filling in slots: they do not require powerful reasoning processes. These methods account for a large part of the power of frames - any new frames interpreting the situation can make use of values determined by prior experience, without having to recompute them.

A frame **instance** denotes an individual. Given a frame representing a concept, we can generate an instance of the concept by filling in its slots - this is known as **instantiation**. For example, we may wish to generate a frame representing TOM, an actual real-life cat rather than a concept, and TOM is an INSTANCE of a CAT:

```
[TOM
  INSTANCE    $value    cat
  FUR-TYPE    $value    tabby]
```

If we had no information about Tom's fur-type, he would have inherited "black-and-white" from the $default facet of the **fur-type** slot of the CAT frame.

Underlying the declarative structure of frames is an important dynamic or procedural aspect of frame-based systems. In particular, procedures can be attached to slots to drive the reasoning or problem-solving behaviour of the system. For example, there may be a procedure in the **$if-needed** facet of a slot, which can be used to compute the slot's value if necessary. For example,

```
[CAT
  AKO       $value      mammal
  FUR-TYPE  $value      ;this empty facet is waiting for the value
            $require    (black-and-white tabby tortoiseshell)
            $default    black-and-white
  COST      $if-needed COST-DEMON (IF FUR-TYPE = 'tortoise-shell
                                   THEN 15-pounds
                                   ELSE 'free)]
```

An **if-added** procedure, or demon, specifies an action that must be taken, either if **any** value, or else a **specific** value, is used to fill the empty $value facet of a slot. For example,

```
[CAT
    AKO       $value    mammal
    FUR-TYPE  $value    ;this empty facet is waiting for the value
              $require  (black-and-white tabby tortoiseshell)
              $default  black-and-white
    COST      $if-needed COST-DEMON (IF FUR-TYPE = 'tortoise-shell
```

```
                                          THEN 15-pounds
                                          ELSE 'free)
    FUR-LENGTH $require   (long short)
               $if-added  FUR-LENGTH-DEMON (IF 'long THEN
                                                 brush-a-lot)]
```

A further example is the demon that resides in the **$if-removed** facet of a slot: this specifies the action to be taken if a value is removed from the slot. IF-ADDED, IF-REMOVED and similar procedures are often known as **TRIGGER** procedures, since they take control only when certain events or data occur: they implement event, or data, driven processing.

Once a particular frame has been selected to represent the current situation or object or whatever, the principle process in a frame-based system is often just filling in the details called for by its slots: a **matching** process tries to assign values to each frame's slots consistent with the slot restrictions, ie, the restrictions imposed by the $require facet. There are a variety of ways through which a slot may be filled: explicit information may be elicited or volunteered by the user of the system; the value may be directly inherited from a higher-level frame; default values may be used and attached if-needed procedures may compute the value in the absence of any explicit information.

Frame-based systems have been used in a variety of domains, and briefly described below are some frame systems which are primarily concerned with natural language understanding and discourse handling, and also two systems, in the librianship and reference domains, which have explicit frame-based user-models.

a. NUDGE. NUDGE, (Goldstein and Roberts 1979, 1980), is a system used to understand incomplete and possibly inconsistent management-scheduling requests and to provide a complete specification for a conventional scheduling algorithm. NUDGE accepts informal requests and produces a calendar containing possible conflicts and an associated set of strategies for resolving those strategies. It is implemented in FRL (Roberts and Goldstein 1977), which allows properties to be described by comments, abstractions, defaults, constraints and so forth. It uses a broad database of knowledge to supply missing details, resolve inconsistencies, determine available options, notice necessary prerequisites and plan expected outcomes. NUDGE contains a hierarchy for activities involving information transfer, for the people involved, for the plans etc. There about one hundred objects in the system, each of which is represented in a **frame**.

NUDGE has a language understanding component called PAL (Sidner 1979) which is designed to understand the English form of requests for arranging various events, and the design depends upon a theory about how to interpret a speaker's extended discourse. A frame-type representation is seen as useful in understanding how English speakers use referential terms, and purpose interpretation is viewed as a slot-filling operation. In PAL, association links between concepts can be regarded as slots associated with a certain concept - for example, a MEETING has built-in association links with a Time, a set of Participants, a topic of Discussion etc., but **no** explicit links with other concepts such as cost, colour and age. These association links between concepts are easily expressed in the frame structure of FRL. One important goal of a discourse with PAL is to fill these slots with values and required information. The values given to these slots are useful in interpreting co-reference and in understanding the purpose of a sentence of the discourse.

Each English sentence presented to PAL by a speaker is interpreted via a parser, case frame interpreter and representation mapping program into a series of FRL frames. For example, just two of the frames for the request "Schedule a meeting in my office" are presented below:

```
[FRAME       schedule-201
 AKO         schedule
 TYPE        ''imperative''
 ACTOR       PAL
```

```
 EVENT          meeting-203]

[FRAME          meeting-203
 AKO            meeting
 PLACE          office-207
 DETERMINER     ''a'']
```

The goal of the knowledge-based phase of NUDGE is to compute a FRAME GESTALT, which is a set of generic frames, instantiated appropriately for a particular scheduling request. The input to the gestalt formation process is a set of partially instantiated frames, representing the information actually present in the informal request. The formation of a frame gestalt occurs by expanding the frames extracted from the initial request in terms of the knowledge stored in the FRL database. Various representation techniques embedded in FRL contribute to this process, for example:

(a) Abstraction, ie, the inheritance of information between concepts

(b) Defaults, ie, the slots of a generic activity typically supply default answers to the common questions asked about such events

(c) Constraints, ie, FRL allows constraints to be attached to a slot by means of facets for requirements and preferences. Requirements are predicates which must be true of the values in a slot; preferences can be relaxed yet leave a valid slot. For example:

```
[MEETING
 AKO    $value          activity
 WHO    $require        (exists ?who (has-role 'chairman))
 WHEN   $prefer         (not (> (duration ?when)(hour 1.5))]
```

In the above frame, the **who** slot has a facet, $require, which constrains the value used to fill the slot, ie, the value must be someone who has the role of Chairman. The **when** slot has a $prefer facet, which indicates a preference, but not a rigid requirement, for a particular value: in this case, the meeting should not last for more than one and a half hours.

NUDGE translates an ill-defined, under-specified scheduling request into a complete specification, represented by the frame gestalt. This gestalt becomes the input to a scheduling program, BARGAIN, that seeks the best time for the requested activity if the desired time is unavailable. Goldstein and Roberts found that the frame gestalt derived from a frame structure for a concept supplies missing information similar to that generated by competent human schedulers to handle informal requests.

FRL (Roberts and Goldstein 1977), which was used to implement NUDGE, consists of a specialised datastructure (the Frame) and a collection of LISP functions for defining frames and storing and retrieving information. Apart from NUDGE and PAL, it has been used to represent the discourse structure of news articles (Rosenberg 1977), and for the implementation of COMEX (Stansfield 1977), a system for understanding discourse about the commodities market.

An FRL frame is implemented as nested association lists with at most five levels of embedding. The respective substructures of a frame are slot, facet, datum, comment and message. A slot is a generalisation of the attribute-value pair in the traditional LISP p-list representation. $value is the slot facet which indicates its values. Five other facets indicate other types of knowledge associated with the slot. Data in the $default facet supplies defaults. Data in the $if-added and $if-removed facets are procedures triggered whenever a slot value is added or removed. $if-needed data are procedures which may compute a slot value. The $require facet holds predicates which describe and restrict the value.

Two slots are recognised by FRL system functions: AKO (A Kind Of) and INSTANCE. These define a relation between frames along which data is inherited. FRL maintains AKO and INSTANCE as inverses. The AKO relation can be used to establish a conceptual hierarchy of frames in which general information stored higher in the hierarchy is inherited by more specialised concepts lower in the hierarchy.

b. GUS. GUS, a Genial Understander System (Bobrow et al 1977), attempts to model natural language and is restricted to the role of a travel agent in a conversation with a client who wants to make a single return trip to a city in California. The system attempts to demonstrate how various aspects of dialogue understanding - such as handling mixed-initiative dialogues, indirect answers and anaphoric references - could be facilitated by the ability to make defaults and expectations available with frames. It is similar to PAL in that it expects a discourse to provide information about the slots of a frame.

Frames are used to represent collections of information at many levels within the system. Some frames describe the sequence of a normal dialogue, others represent the attributes of a date, a trip plan, or a traveller. In GUS, every frame is an instance of some prototype, with the exception of the most abstract frames in the permanent database. Most instances are created during the process of reasoning, although some (for example, those representing individual cities) are in the initial database. A prototype serves as a template for its instances.

Example of **prototype** frame for DATE:

```
[DATE
        MONTH           Name
        DAY             (bounded integer 1 31)
        YEAR            integer
        WEEKDAY         (member (mon tues wed thu fri sat sun))]
```

Example of **instance** frame for 28th MAY:

```
[28-MAY
        IS-A            DATE
        MONTH           May
        DAY             28]
```

In GUS, frames are used at several levels to direct the course of a conversation. At the top-level, GUS assumes that the conversation will be of a known pattern for making trip arrangements. To conduct a dialogue, the system first creates an instance of the Dialogue frame, and then goes through the slots of this instance attempting to find fillers for them in accordance with the specifications given in the prototype. When a slot is filled by a new instance of a frame, the slots of that instance are filled in the same way. GUS follows this simple depth-first, recursive process, systematically completing work on a given slot before proceeding on to the next. This is how GUS attempts to retain the initiative in the dialogue.

c. GRUNDY. In Grundy (Rich 1979a, 1979b - also see Chapter 2, Section 2.4.2), a system which attempts to model the characteristics of fiction readers, stereotypes can be thought of as frames which contain a number of user traits that are expected to appear in that stereotype. These frames are ordered in hierarchies of stereotype generality, and information is inherited down the hierarchy when a lower level stereotype is activated. The most general node of any stereotype structure is the stereotype ANY-PERSON. It provides default values for all facets, but provides them with very low ratings. These values can be used to prevent the system from doing anything that might be offensive until it learns enough about the individual user to know his/her particular inclinations. This is essentially a kind of notion of a canonical or typical person with respect to whom individual characteristics can be compared.

The core of the user-modeller is its model of an individual user. This model, called the User Synopsis, or USS, is built by combining the direct information provided by the user, direct inferences from the user's actions, and predictions made by stereotypes that are deemed appropriate for this user. The information in the USS can be used to guide the rest of the system.

The stereotypes themselves actually consist of a set of triples: attribute, value and rating - the rating indicates how confident it is that that piece of information is correct, ie, confidence

FEMINIST STEREOTYPE

Activated-by	Feminist-word-trigger	
Generalisation	Any-person	
FACET	VALUE	RATING
Genres		
woman	3	700
Politics	Liberal	700
Sex-open	5	900
Piety	-5	900
..	..	..
..	..	..

Figure 6.1: A sample stereotype for Feminist

that any person who fits the stereotype will exhibit a particular characteristic. The USS needs to know **why** it believes the things it believes, so it consists of a set of quadruples - attribute, value, rating and justification. Here, the ratings indicate the confidence the system has in its belief that a particular fact is true of a particular user. Figure 6.1 represents a sample stereotype.

The values indicate whether the characteristic represented by the facet is positive or negative, eg, the facet Piety can take values ranging from -5 (not at all pious) to +5 (very pious). The confidence ratings (0 - 1000) indicate how certain Grundy is that the values assigned to the facets are true for a particular user. Therefore, in the example, a feminist is shown to have a very low piety value, ie, she is not at all pious, and the system is highly confident of this value, ie, there is a rating of 900/1000.

In addition to the representations of users, Grundy also contains descriptions of the novels which it knows about. Each description of each individual book comprises a set of facets with attached values. Once there is a sufficiently developed user model, Grundy uses the model to recommend books. A facet is selected from the USS which has a very high or very low value, and associated with this value a high confidence rating. This facet is then searched for in the inverted index to the database of documents.

A common feature of frame-based systems is that it is seldom necessary that **all** slots in **all** frames need have values assigned: frames and their corresponding slots are intended to serve as placeholders for any data that may be relevant, even though it may not always be present in every interaction with the system. Only those frames and slots which are required for the current reasoning process need be instantiated. Similarly, although Grundy's stereotypes can contain any number of characteristics, they will almost never contain values for **all** the characteristics that the system understands. But since usually more that one stereotype will be active for a particular user, it will be possible to build up a fairly complete representation of a user even though the individual stereotypes contain only partial information.

d. PLEXUS. Plexus (Vickery et al 1986), a reference system for the gardening domain, uses a variety of knowledge representation methods: production rules, semantic nets, and frames (see also Chapter 2, Section 2.4.4 for more detailed description). Every term in its dictionary is assigned to a category, and associated with each category/subcategory is a Frame Structure. The frames define what other concepts may be associated with the frame concept in a well-framed problem statement.

Plexus can be regarded as both data and expectation-driven. The information supplied by the user in response to the system questions determines the pattern and sequence of activation of the rules. The system can also be regarded as expectation-driven. The frames used to develop the problem model set up expectations of the data required by the system in order to formulate an effective search strategy.

For each of the terms in the input, a frame is instantiated. The structure of the frame depends on the semantic category of that term. Rules attempt to fill as many of the slots as possible from existing information. The set of instantiated frames constitutes the system's model of the user's problem.

6.3.3 The Frames for the User Model

The General Structure and Attributes of the User-Modelling Frames

A series of frames has been devised for each of the five user modelling subfunctions, and the User Model currently consists of approximately 160 frames in total. It should be emphasized that although these frames have been devised with a view to eventual implementation, the design that is presented here is intended to represent the general structure and attributes of the frames, and has been written independently of any particular language or implementation. The complete hierarchy of frames for each of the user modelling subfunctions is presented in Appendix F.

Each frame has a NAME, an AKO slot indicating the higher-level frame from which it inherits information, and a series of one or more SLOTS. Each slot has a SLOT-NAME and an empty VALUE facet, which will be created and filled at instantiation time. Every frame has a TYPE slot, which is the principal slot, serving as a placeholder for the most important distinguishing feature of that frame, and is the slot through which instantiation of most of the values in the frame dependents takes place. Each frame may also have some subsidiary slots, indicated by the suffix **-slot**. Furthermore, slots may also have the following facets: $require, $default, $if-needed, $if-added and $prefer. Facets are prefaced by a $ sign.

The $require facet indicates that the slot value must be restricted to one of a range of values, either a lower-level frame or a terminal value. Terminal values are also represented as frames, and an example, FORMER-ACADEMIC, which is **a kind of** USER-STATUS, is presented below.

```
[USER-STATUS
AKO             $value          user-model

type            $require        current-academic former-academic
                                prospective-academic
                                current-non-academic

                $default        current-academic

position-slot   $require        position]

[FORMER-ACADEMIC
AKO             $value          user-status

type            $value          terminal]
```

The $require facet in the USER-STATUS example indicates that the value for the **type** slot must be one of the following frames, eg, CURRENT-ACADEMIC, FORMER-ACADEMIC, etc. Some of these frames may have **type** slots indicating a range of lower-level frames to be further instantiated, others, eg, FORMER-ACADEMIC, may have the value of their **type** slot set to terminal, in which case no further instantiation takes place. The value is entered into the empty $value facet and, as every slot-filler is represented as a lower-level frame, that sub-frame is located and filled out.

In the case of some UGOAL and BACK frames, the **type** slot can occasionally take more than one value, for example, a user is quite likely to have more than one type of goal to fulfill when having a online search carried out. Insofar as the user modelling frames are currently formulated, it would be necessary to instantiate the parent frame more than once in order to allow a particular user to have, for example, more than one goal or several different background categories. No precedent for such a case has yet been identified in the frame system literature (although it seems likely that such a problem has been discussed elsewhere) and this appears to be an implementational problem for frames, as well as a conceptual problem.

In order to account for the possibility of allowing multiple values to be instantiated in some **type** slots, a number of stategies could be suggested for possible development in future work on the User Model:-

(a) a new facet could be added to those **type** slots where multiple values may be instantiated, for example, $any-of or $one-or-more-of.

(b) the contents of all $require (and $prefer) facets could be made into lists, and either the key-word 'any-of' or 'one-of' would appear as the first element of the list, according to whether single or multiple values are permitted at the time of instantiation.

In addition to the main slot, the **type** slot, the USER-STATUS frame also has a subsidiary slot, the **position-slot** slot. The $require facet attached to **position-slot** indicates that its value must be the POSITION frame, and therefore this frame is located and filled out. The POSITION frame is presented below:

```
[POSITION
AKO        $value       user-status-slot

type       $require     end intermediary team
           $default     end]
```

In the case of the above frame, POSITION, the value of its AKO slot is not the USER-STATUS frame itself, but a user-status-SLOT. This is because the user's position, ie, the fact that the user is an end user, a team member, or an intermediary user searching on behalf of someone else, is not the main concept represented in the USER-STATUS frame, rather it qualifies the main concepts represented by the **type** slot.

Frames which occur as a **slot** in other frames probably do not inherit in exactly the same way as frames which are linked to other frames via the **type** slot. In the above example, POSITION, which is a **user-status-slot**, will not inherit all of the attributes of USER-STATUS, whereas CURRENT-ACADEMIC, for example, does inherit all of USER-STATUS's attributes.

It appears that frames which appear as subsidiary slots (rather than **type** slots) in frames tend to denote concepts which can stand in their own right, and which could probably appear as slots to qualify new frames which may be defined in the future. For example, **institution** occurs as a slot in the CURRENT-ACADEMIC frame, but could occur just as appropriately in a frame representing the concept of, say, MEDICAL-DOCTOR. In fact, the **relevant-to-search-topic-slot** slot, in the BACK frames, already occurs as a slot in three different frames: EMPLOYER, RESIDENCE and ACADEMIC-BACK. It should therefore again be emphasized that a frame which is a slot will not inherit in the manner of a frame which is AKO frame, but at this stage of the research it is not yet clear exactly what type of inheritance is implied. It may be necessary to define a new type of relation, such as OA (Occurs As), in addition to AKO, to emphasize the difference in inheritance between frames which are ~~frames~~ types and frames which are **slots**. This issue will be examined in more detail in future work on the User Model.

It seems likely that, for implementation purposes, the **type** slot **must** be filled in any frame that is to be instantiated: although the subsidiary slots could possibly be filled by values without the TYPE slot itself having a value, such information would appear to be inadequate, even meaningless, on its own. In the transcripts analysed to date, no frames have been instantiated in which subsidiary slots have values without the **type** slot being filled.

The $default facet suggests a value for the slot to which it is attached, unless there is contradictory evidence, for example, explicit information elicited from, or volunteered by, the user. In the USER-STATUS frame presented above, the type slot has a $default value of **current-academic**.

The $if-needed facet of a slot contains a demon which is fired if the value cannot be inherited or obtained from the user. For example, the SPECIFIC-INTENTIONS frame (a frame from the UGOAL set of frames) has a **type** slot:

```
[SPECIFIC-INTENTIONS
AKO          $value        ugoal

type         $prefer       review-article thesis
                           dissertation report
                           paper teaching-materials
                           advise-others

             $if-needed    specific-intention-demon
                           (IF USER-STATUS Research
                           THEN thesis OR
                           dissertation]
```

In the above example, if a value cannot be found for **type**, then the **specific-intention-demon** in the $if-needed facet is invoked. This demon contains a function whose approximate natural language translation is as follows:-

If there is a user modelling frame in the USER-STATUS set called RESEARCH, then set the value of the **type** slot of SPECIFIC-INTENTIONS to Thesis or Dissertation. If there is no frame called RESEARCH, then do nothing.

The $prefer facet is similar to the $require facet, in that it places some sort of constraint on the $value facet, but the $prefer facet merely indicates a preference for the types of values and is not an inviolable requirement. There is an example of a $prefer facet in the above SPECIFIC-INTENTIONS frame.

The final facet is the $if-added facet which, like the $if-needed facet, contains a demon which can be invoked in specific circumstances. The most common use for this demon in the User Model is to post information generated by the user modelling subfunctions onto the Blackboard so that it can be used by the other experts in their own processing. For example, in the set of User Status frames, there is a frame called UK-UNIVERSITY:

```
[UK-UNIVERSITY
AKO          $value        institution

type         $require      a-uk-university

             $if-added     uk-uni-demon (TRIGGER (EXPLAIN))
                           ;charges higher for external users]
```

This **if-added** demon is invoked if the **type** slot is filled, whereupon it alerts (indicated by a **TRIGGER** keyword) the EXPLAIN function. The specific information directed to the function is included in the form of a comment, prefaced by a semi-colon. In this case, the EXPLAIN function or expert is alerted to the fact that charges for using the service are higher for those users who are **not** part of London University.

In addition to posting information onto the Blackboard for the benefit of the other functions, the IF-ADDED demons can also send information to fill out slots, or alert other frames, in the User Model itself. For example, there is an $if-added facet attached to the **type** slot in the terminal frame, UNDERGRAD:

```
[UNDERGRAD
AKO             $value          student

type            $value terminal
                $if-added       undergrad-demon
                                  (TRIGGER IRS Experience = None)]
```

In the above example, if the UNDERGRAD frame is instantiated, the user modelling frame **IRS** is alerted, and the value of its **Experience** slot is set to **None**.

When the intelligent interface is eventually fully implemented, it is anticipated that the user will interact with the system using natural language. In that case, it is likely that the system and user will engage in a mixed-initiative dialogue, and that it will not be necessary for the frames to be filled out in a rigid, hierarchical order, ie, higher-level frames being instantiated before lower-level ones. Instead, frames will be instantiated as and when they are identified during the course of the dialogue, and this implies both data and goal-driven processing. Normally, a dialogue would begin by a high-level frame being instantiated, followed by the lower-level frames which are usually the slot-values of that higher-level frame. In this way, a system usually operates in a top-down manner, working from the most general to the most specific, finishing up with frames whose slot values are terminal assignments. However, in the course of a dialogue with this intelligent interface, it should be possible for the user to volunteer terminal values early on in interaction, and for the system to engage in upward-chaining in order to arrive at the top-level frames. For example, a user may assert that she is a postgraduate student and the system should then be able to work **backwards** through the frame hierarchy in order to deduce that this user is therefore a **current academic**, a **student**, etc. The frame hierarchy for each subfunction is presented in Appendix F.

It is also necessary to consider the order in which new frames are to be instantiated. There appear to be two basic ways: when a slot is filled by a new instance of a frame, the slots of that frame instance are filled out in the same way, ie, a simple depth-first, recursive process which ensures that work is completed on a given slot before the next slot in that frame is dealt with. This is the method used by GUS (Bobrow et al 1977). On the other hand, the system could attempt to fill out **all** the slots in a particular frame, before going on to instantiate those slot values which are lower-level frames: a breadth-first process. The question of order of instantiation largely depends on how much initiative the system wishes to retain, and is also a matter of preferred dialogue and programming style.

At the present time, it is expected that the User Model will be implemented independently of the other interface functions as a small, experimental prototype, and that the mode of interaction with the user will differ considerably from those described above. Each of the five sets of user modelling frames has a "top-level" frame, ie, the highest frame in the frame hierarchy for each subfunction. The current set of user modelling frames exists merely as prototypes, ie, they are "empty" frames which do not yet hold values pertaining to any individual user. It is anticipated that the prototype system will work by first instantiating the five top-level user modelling frames, ie, by creating **instances** of these five frames, the slots of which will be filled by explicit values obtained from the user who is being modelled. Having obtained slot values for these top-level frames, it will then be possible to begin the modelling process, which could proceed in a variety of ways. For example, one way of proceeding would be to attempt to fill out as many as possible of the frames in **one** subfunction, say USER-STATUS, before proceeding onto the next subfunction. On the other hand, the system could begin to fill out the frames of one subfunction, and then find that frames in another user modelling subfunction are triggered by means of the IF-ADDED demons, in which case the system could then attempt to instantiate those frames.

```
--------------------------------------------------------------------
I now I gather you're (cough .) excuse me you're a visitor /(3)  yes
U                                                      yes I am /(4)
--------------------------------------------------------------------
--------------------------------------------------------------------
I yes are you part of the university or /(5)
U                                  well I teach at a Canadian univ-
--------------------------------------------------------------------
--------------------------------------------------------------------
I     ya (,) um I I jus' (,) we ask you this because i- its awful to
U ersity /(6)
--------------------------------------------------------------------
--------------------------------------------------------------------
I bring up charges straight away (laugh . ) but just so that you know
U
--------------------------------------------------------------------
--------------------------------------------------------------------
I (,) you know that its a ten pound basic an its (inaud . . . .) /(7)
U                                                          yes /(8)
--------------------------------------------------------------------
--------------------------------------------------------------------
I right ok (,)
U
--------------------------------------------------------------------
```

Figure 6.2: Focus 1 of Interview 190684hba

Applying the Formalism

As a demonstration of how this representation convention works overall, an example is presented for Focus 1 of Interview 190684hba. The subfunctions for each utterance in that focus are in Table 6.1, and the transcript extract is presented in Figure 6.2. An explanation of how values were assigned, demons fired, etc. follows the set of instantiated frames. A complete set of user modelling frames for all of Interview 120684hba is presented in Appendix D.

Figure 6.3 demonstrates how some of the user modelling frames for the USER subfunction were instantiated during this focus. The actual values which are assigned to the slots for this user during this interview appear in the $value facet in uppercase characters. Information to be posted onto the Blackboard resulting from demons firing, or values to be entered in the slots of other user modelling subfunctions' frames, appear in parentheses below the frames themselves.

Utterance Number	Subfunction
3	USER
4	USER: CONFIRM
5	USER
6	USER
7	EXPLAIN
8	EXPLAIN: CONFIRM

Table 6.1: Subfunction analysis for Focus 1 of Interview 190684hba

```
--------------------------------------------------------------------------------
[USER-STATUS
AKO             $value          user-model

type            $require        current-academic
                                prospective academic
                                former academic
                                current-non-academic
                $default        current-academic
                $value          CURRENT-ACADEMIC

position-slot   $require        position]
--------------------------------------------------------------------------------
[POSITION
AKO             $value          user-status-slot

type            $require        end intermediary team
                $default        end
                $value          END]
--------------------------------------------------------------------------------
[END
AKO             $value          position

type            $value          terminal]
--------------------------------------------------------------------------------
[CURRENT-ACADEMIC
AKO             $value          user-status

type            $require        staff student
                $default        student
                $value          STAFF

institution-slot
                $require        institution

funding-source-slot
                $require        funding-source

time-slot       $require        time]
--------------------------------------------------------------------------------
[STAFF
AKO             $value          current-academic

type            $require        teaching research-staff
                $default        research-staff
                $value          TEACHING]
--------------------------------------------------------------------------------
[TEACHING
```

```
AKO         $value       staff

type        $value       terminal]
--------------------------------------------------------------------------------
[INSTITUTION
AKO         $value       current-academic-slot

type        $require     London-uni UK-uni
                         foreign-uni
            $default     London-uni
            $value       FOREIGN-UNI
            $if-needed   institution-demon
                          (IF BACK Residence
                           THEN foreign-uni)

department-slot
            $require     department]
--------------------------------------------------------------------------------
[FOREIGN-UNI
AKO         $value       institution

type        $require     foreign-uni-name
            $value       CANADA-UNIVERSITY
            $if-added    foreign-uni-demon
                         (TRIGGER EXPLAIN; charges
                          increased AND TRIGGER
                          BACK Temp-features;
                          has user special
                          requirements?]

(DEMON FIRED:-
  Foreign-uni-demon: The EXPLAIN function is alerted to
explain to the user about higher charges for the use of
the system, and the TEMPORARY-FEATURES frame of the user
-modelling subfunction, BACK, is activated so that
special background features of users from foreign
universities may be elicited.)
--------------------------------------------------------------------------------
[CANADA-UNIVERSITY
AKO         $value       foreign-uni

type        $value       terminal]
--------------------------------------------------------------------------------
```

Figure 6.3: USER Frames instantiated during the first focus of Interview 190684hba

In the above example, the first frame to be instantiated was the top-level frame, USER-STATUS. The **type** slot was filled with the value **current-academic**, which is obtained from the statement in the present tense in utterance 6. The next frame to be filled out is the POSITION frame, one of the subsidiary slots in the USER-STATUS frame, and the value **end** for the **type** slot is found by accepting the default value, in the absence (so far) of any explicit information from the user. The terminal frame END is therefore instantiated.

The fourth frame to be filled out is the CURRENT-ACADEMIC frame. The **type** slot is filled out by the value STAFF, again supplied by information in utterance 6, and as there is some information relating to the user's institution, also in utterance 6, the INSTITUTION frame is filled out. The **type** slot in the STAFF frame takes the value **teaching**, and therefore the terminal frame, TEACHING, is instantiated.

The seventh frame, INSTITUTION, has its **type** slot filled by the value **foreign-uni**, and as a consequence the FOREIGN-UNI frame is instantiated with the value **Canada-university**, as the user has not specified a named university. A demon is fired in this frame, the **Foreign-uni-demon**, and the result can be seen in utterance 7, when the intermediary does begin to explain about charges. Also, this demon activates the TEMPORARY-FEATURES frame of the BACK subfunction, and the intermediary attempts to fill out this frame at a later stage in the dialogue. Finally, the terminal frame FOREIGN-UNI-NAME is instantiated, and this becomes CANADA-UNIVERSITY, which was the value assigned to the **type** slot of the FOREIGN-UNI frame.

6.4 Knowledge Resources for the User Model

Essentially, it appears that most of the **internal** knowledge of the intermediary is represented directly in the frame structure of the User Model, ie, the knowledge, information, concepts etc. to be elicited and hypothesized in order to conduct a successful interaction, are reflected by the frames that need to be instantiated, and the slots that require values.

The knowledge resources for the UGOAL frames are represented directly in the frames themselves. This knowledge concerns the likely range of goals which particular categories of users typically have, and this knowledge is generally informed by information generated by the USER subfunction. For example, a Research Student is likely to have a short-term goal of completing a Thesis or Dissertation, or a member of the Teaching Staff may wish to prepare Teaching Materials.

Knowledge relating to the USER subfunction is mainly categorical, in that each user can be classified according to one of a finite range of categories, ie, Academic vs. Non-academic, and category subdivisions, eg, Staff vs. Student. This information is contained mainly in the frames themselves. However, some information relating to user's institutions is **not** explicitly specified in the frame/slot structure, and there is a **require** constraint which should check that the institution or country which is entered as a value, is in fact a member of a list of valid institutions, foreign countries etc. This information could probably be represented in the form of a look-up table or some similar structure.

The BACK subfunction requires extensive knowledge resources, and the potential range of user backgrounds will probably depend largely on the type of users consulting the information service for which a particular set of frames have been devised. For example, it will probably be necessary for the system to have some knowledge regarding the typical ranges of employment in which its users are engaged. In the case of foreign users, the system needs to check that details of the user's residence background are valid, for example, that the name of the country or town is acceptable. Furthermore, it is sometimes useful to discover whether a user has access to another library (in order to retrieve secondary sources) and in this case, the system needs to access information relating to other UK libraries or information services. Users frequently have a maximum budget available to spend on the search: the system must be able to calculate how long the search may take, or how many references may be retrieved, without exceeding the user's budget, and to inform the user of the restrictions which this budgetary constraint will impose (a task of the EXPLAIN function).

The IRS subfunction needs access to two types of knowledge resources: system-specific information, such as known databases and hosts, and subject-specific information, ie, the types of secondary sources a user is likely to have consulted given that he or she is working in a particular subject area. Such information is not represented in the frame/slot structure, and needs to be represented elsewhere. It will probably be useful to indicate which information services access which hosts, and which universities or institutions have their own services, and also offer IRS training.

The knowledge resources of the KNOW subfunction represent information regarding the qualitative assessment of a user's state of knowledge, and this is represented explicitly by the **evalue** slot in the KNOW frame, which allows the user or the intermediary to assign qualitative values to aspects of the user's knowledge, eg, **good** knowledge on one part of the search topic, **no** knowledge of information sources, etc. The KNOW subfunction also needs to access knowledge relating to Problem Description: PD knowledge is necessary because the intermediary must be capable of decomposing a particular topic into its component parts, so that the depth or quality of the user's knowledge on each of these component parts can be ascertained. Problem Description information will be obtained, or volunteered, by the Problem Description function.

Chapter 7

A VALIDATION OF THE USER MODEL

7.1 Introduction

The goal of the validation is to demonstrate that the User Model and its representation is independent of the data on which it is based. The subfunctions and knowledge resources of the User Model were identified after detailed analysis of seven dialogues, further analysis of twenty-five transcribed problem statements and their accompanying recorded interactions, and interviews with three intermediaries, one of whom was not involved in this research project in any way. In order to discover whether the User Model and its formalism can account for utterances concerned with user modelling in **other** pre-online information interactions, it is necessary to apply the results of the User Model analysis to a dialogue that has **not** previously been used as data for analysis.

In this chapter, a new dialogue will be presented, together with its subfunction analysis. The appropriate frames for each utterance concerned with user modelling will be identified and discussed, in order to see how adequately the current subfunctions and frames can deal with those utterances.

7.2 The Transcript

The recording was made at CIS on 13th November, 1985, and the search was conducted by one of the intermediaries described in Chapter 4. The recording methodology, transcription and analysis were also described in that chapter. Method U-2 was used for utterance identification. However, Interview 131185pda differed slightly from the core set of transcripts in that it involved **two** users, one who appeared to be the "dominant" user (perhaps the research team leader) and a second, "subsidiary" user (perhaps the first user's assistant). As both the users are working on the same problem, the same project and search topic, they are treated for the purposes of this analysis as **one** user, ie, it is suggested that the intermediary builds just one User Model, which will probably contain salient characteristics of both users. In Interview 131185pda, it is quite clear that the users' user modelling attributes are complementary, and that no potentially conflicting characteristics are elicited. Over the next 10 pages, the transcript of Interview 131185pda is presented in full, and the utterance analysis follows the transcript. The text continues on page 112.

```text
---------------------------------------------------------------------
I  right (.) good (.) ok (...) right what's the subject I think we
U1
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I  just had down as phonetics in the diary /(1)
U1                                         yes yes that's the
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I                                                    mmm /(3)
U1 department that we're from at University College /(2) but er we've
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I
U1 got a very specific search through /(4a) we've actually made um one
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I                                             mm /(8)
U1 search is that /(4b)  one on this topic that's right at University
U2                 yeah /(5)    yeah /(7)        yeah /(9)
---------------------------------------------------------------------
---------------------------------------------------------------------
I          ah yes mmmm /(10)
U1 College /(6)        and we want to narrow it down even a bit
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I
U1 further but we're not sure what we'll find but we may we may catch
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I                                            mmmmm /(12)
U1 something that we haven't caught before /(11)    the very specific
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I
U1 topic of things that we're looking for um the actual narrow title
U2
---------------------------------------------------------------------
---------------------------------------------------------------------
I
U1 is intervocalic no laryngeal ok laryngeal co-articulation effects
U2                         laryngeal /(14)
---------------------------------------------------------------------
---------------------------------------------------------------------
I                   mmhmm /(15)    is that two ls or (laugh) one/(17)
U1 in intervocalic /(13)  consonants /(16)
```

```
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I           now could you explain a bit to me about what that is /(19)
U1 no one /(18)                               yes ok /(20)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 when you um an intervocalic consonant is just a consonant between
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I                  mmm /(22)
U1 two vowels um /(21)   and when you pronounce consonants um most of
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 the stuff most of the interesting stuff traditionally is concerned
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                 mm /(24)
U1 to be going on in the mouth /(23) but in fact the larynx is doing
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 all kinds of interesting things at the same time and so this is the
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                          uh huh /(26)
U1 those are what we would call laryngeal /(25) co-articulations /(27)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  yes /(28)                       yes yes /(30)
U1       cause they're the articulations that are being made in
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 co-ordination with the with the er the mouth articulations /(29)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  uh huh /(31)          yes yes that's interesting so that's that's
U1        so that's the /(32)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
```

```
I  what you really want to home in on on this particular search /(33)
U1
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 yeah probably what I think p'raps intervocalic is the least
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I                     mmmm /(35)                        what what d'you
U1 important label that really could have been found wouldn't you
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  want /(36)   yeah /(37)
U1 say /(34)        that if we if we can find anything something about
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I                          is it is it /(39)
U1 co-articul- larynx effects /(38)     we'll go along with consonants
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I    if we if we simply simplified to anything where larynx is
U1 anyway /(40)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  mentioned where consonants are also mentioned would that be over-
U1
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  simplification /(41)
U1             err I think we'll /(42)
U2                             no that'll probably no that'll be
----------------------------------------------------------------------
----------------------------------------------------------------------
I                     mmm or would it be too specific is a- would
U1
U2 that'll be useful /(43)
----------------------------------------------------------------------
----------------------------------------------------------------------
I  that be cutting it down too much /(44)
U1                                              I think larynx is /(46)
U2                    in relation to consonants /(45)
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1                                               yeah do you think that
U2 I think that'll be reasonably useful actually /(47)
```

```
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 there are any other any other words that would do double duty with
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                                 yes /(51)
U1 with laryngeal or larynx and /(48)   yes glottal /(50) you see
U2                              glottal /(49)
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U1 you see voicing won't do it that's just plus or minus voice isn't
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I      yes if we started to to put in voice or speech should that be
U1 it /(52)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  an' larynx that would be too much /(53)                uh huh /(55)
U1                  voice voice and speech is too general /(54)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I          right /(57)                          again what we'll find as we
U1 (inaud.) /(56)     but yeah glottal um /(58)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  go along is that it depends on the database erm perhaps in a
U1
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  linguistics database we might need to go into very specific ones
U1                              um /(60)
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  but say we're talking about a medical database then maybe simply
U1
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  speech or speaking would erm be as specific as you could get
U1
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I  there /(59a) but we'll what we'll do is draw up a strategy where
```

```
U1
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I  we've got the specific and the general so erm we can see the
U1
U2                                                     what about (....)
------------------------------------------------------------------------
------------------------------------------------------------------------
I results /(59b)                          mm hmm /(63)
U1               oh vocal folds yes that might be good vocal fold or
U2 vocal folds /(61)
------------------------------------------------------------------------
------------------------------------------------------------------------
I                   vocal chords again all all linked with well /(64)
U1 vocal chord (6 secs) /(62)
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I                               yes /(66)            er /(67)
U1 the vocal folds vocal chords and glottal are linked with
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I        yes yeah so really these are alternatives to larynx it's a
U1 larynx /(65)
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I  good thing (inaud..) putting it into groups  and consonant /(68)
U1                                        and what we're interested
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I                        yes /(70)  yes /(71)       yes mmhm and
U1 in is what goes on here when you make consonants /(69)
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I  you're not really interested in any other type of erm sound really
U1
U2
------------------------------------------------------------------------
------------------------------------------------------------------------
I  you mainly want consonants /(72)
U1
U2                         I think consonants are an essential /(73)
------------------------------------------------------------------------
------------------------------------------------------------------------
I
U1 you see the important thing is we're not interested in this in
U2
------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------
I           no no /(75)
U1 general its what goes on here when you're making an articulation
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I                           mm so articulation then er (...) mm if we
U1 somewhere else so /(74)
U2                yeah /(76)
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I  put speech or speaking it would be too broad would it I mean
U1
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I  obviously linking it with larynx with we we could put it in if we
U1
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I  get nowhere with those we could broaden it out (.....)/(77a) now
U1
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I which databases did you use before was it the Language and Language
U1
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I  Behaviour Abstracts /(77b)
U1                     no what's it called it's the physics one /(78)
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I  er Inspec /(79)          did you get much out of that one /(82)
U1         Inspec that's right /(80)
U2                Inspec /81)
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I
U1 well we got about twenty titles of something out of that one but
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I                   no /(84)  mm /(85)
U1 none of them were specifically (.) I mean they were more general
U2
--------------------------------------------------------------------------
--------------------------------------------------------------------------
I                                                              mm /(86)
U1 sort of things likely some of them were quite useful but /(83)
```

U2

I there's actually one called it it deals more
U1
U2 so there's one on /(87)

I with linguistics erm Language and Language Behaviour Abstracts and
U1 mm /(89)
U2 mm /(90)

I there's erm /(88) 'cause 'cause your
U1 definitely look at that one /(91)
U2

I subject really overlaps quite a wide area quite a lot of different
U1
U2

I fields /(92a) there's also erm that's that's I'd go for that
U1
U2

I mainly also the medical one could well have something
U1
U2

I Medline /(92b) it's a possibility anyway erm (....)
U1 yes it's (inaud.) /(93)
U2

I there's there's a multidisciplinary one you know the Current
U1
U2

I Contents database Sci-search and Social well probably more Science
U1
U2

I Citation Index /(94) that might be
U1 yes we don't know much about that one /(95)
U2

I worth just dipping into very quickly cos th- what we can do is we
U1
U2

I probably would select Language and Language Behaviour Abstracts to
U1
U2

I start the search with erm and then maybe that'll be the answer to
U1   mmm /(97)
U2

I your prayer (laugh) then if not then we can save the strategy that
U1   mm /(98)
U2

I we've typed in and get the system to run it automatically on these
U1
U2

I other databases erm /(96a) there's another one called Modern L-
U1
U2

I actually we're just compiling a list of databases covering
U1
U2

I linguistics (laugh) so you've come at the right time erm there's
U1   right /(99)
U2

I Psychological Abstracts but I think possibly only if we're really
U1
U2

I getting nowhere /(96b)   and another one is the education one and
U1   mmm /(100)
U2

I that does a lot with erm speech defects and things like that so
U1
U2

I that covers quite a bit of linguistics in that it (inaud.) so try
U1
U2

I this one first (.......) /(101a) erm those are the main ones
U1
U2

I (....) /101b) for interest we'll look at the headings some of the
U1
U2

I databases like the medical one have got a list of terms that they
U1
U2

I use in indexing (....) /(101c) if we have a look at the erm medical
U1
U2

I one and they've actually got laryngeal muscles and they've probably
U1
U2

I got larynx as well (.) and if we search on larynx or laryngeal or
U1
U2

I glottal or vocal fold vocal chords erm and then link with speech
U1
U2

I speaking articulation consonant (.) and see how we go from there
U1
U2

I (.....) /(101d) so p'raps if we start is there anything else you'd
U1
U2

I like to add from that I mean its a pretty specific search /(101e)

```
U1                                                        mmmm
U2
----------------------------------------------------------------------
----------------------------------------------------------------------
I            see how we go and do that /(103)
U1 yeah /(102)
U2
----------------------------------------------------------------------
```

The full utterance analysis of Interview 131185pda is listed below:

Focus	Utterance	Speaker	Subfunction
1	1	I	TOPIC
[USER MODEL]	2	U1	USER
	3	I	ph
	4a	U1	UGOAL
	4b	U1	IRS
	5	U2	ph
	6	U1	IRS
	7	U2	ph
	8	I	ph
	9	U2	ph
	10	I	ph
	11	U1	UGOAL
	12	I	ph
2	13	U1	TOPIC
[TOPIC]	14	U2	TOPIC
	15	I	ph
	16	U1	TOPIC
	17	I	TERMS:MATCH
	18	U1	TERMS:CONFIRM
3	19	I	SUBJ
[SUBJ]	20	U1	SUBJ
	21	U1	SUBJ
	22	I	ph
	23	U1	SUBJ
	24	I	ph
	25	U1	SUBJ
	26	I	ph
	27	U1	SUBJ
	28	I	ph
	29	U1	SUBJ
	30	I	ph
	31	I	ph
	32	U1	SUBJ
	33	I	UGOAL
	34	U1	TERMS
	35	I	ph
	36	I	failed
	37	I	ph

	38	U1	DOCS
	39	I	failed
	40	U1	TERMS
4	41	I	QUERY:MATCH
[QUERY]	42	U1	failed
	43	U2	QUERY:MATCH:CONFIRM
	44	I	QUERY:MATCH
	45	U2	QUERY
	46	U1	failed
	47	U2	QUERY:MATCH:CONFIRM
5	48	U1	TERMS
[TERMS]	49	U2	TERMS
	50	U1	TERMS
	51	I	ph
	52	U1	TERMS
	53	I	TERMS:MATCH
	54	U1	TERMS:MATCH:CONFIRM
	55	I	TERMS:MATCH:CONFIRM
	56	U1	failed
	57	I	ph
	58	U1	TERMS
6	59a	I	EXPLAIN:DB
[TERMS]	59b	I	STRAT
	60	U1	ph
	61	U2	TERMS
	62	U1	TERMS
	63	I	ph
7	64	I	QUERY
[QUERY]	65	U1	QUERY
	66	I	ph
	67	I	ph
	68	I	QUERY
8	69	U1	RES
[RES]	70	I	RES:CONFIRM
	71	I	RES:CONFIRM
	72	I	RES:MATCH
	73	U2	RES:CONFIRM
	74	U1	RES
	75	I	RES:MATCH:CONFIRM
	76	U2	RES
	77a	I	TERMS..QUERY
9	77b	I	IRS
[IRS]	78	U1	IRS
	79	I	IRS
	80	U1	IRS
	81	U2	IRS
	82	I	IRS

	83	U1	IRS
	84	I	ph
	85	I	ph
	86	I	ph
10	87	U2	DB
[DB]	88	I	DB
	89	U1	ph
	90	U2	ph
	91	U1	DB
	92a	I	SUBJ
	92b	I	DB
	93	U1	failed
	94	I	DB
	95	U1	IRS;DB
	96a	I	DB..STRAT
	96b	I	DB..STRAT
	97	U1	ph
	98	U1	ph
	99	U1	ph
	100	U1	ph
	101a	I	DB:EXPLAIN
	101b	I	DB
	101c	I	DB:EXPLAIN
	101d	I	DB..TERMS..QUERY
	101e	I	STRAT
	102	U1	ph
	103	I	STRAT

7.3 The User Modelling Frames for Interview 131185pda

The user modelling process begins in Focus 1, in utterance 2, when the users mention their department and their university. Therefore, the first frame to be instantiated is the frame **UNIVERSITY-COLLEGE**, and by upward-chaining from this terminal USER-STATUS frame, the following frames are also instantiated: LONDON-UNI, INSTITUTION, CURRENT-ACADEMIC, POSITION and USER-STATUS. Also in utterance 2, the DEPARTMENT frame is instantiated, which is a kind of **institution-slot**. The INSTITUTION frame could be instantiated in either one of two ways: by upward-chaining from the terminal frame **UNIVERSITY-COLLEGE**, or by direct instantiation as its subsidiary slot, **department-slot**, is filled out, thereby requiring the **type** slot of that frame to be filled. Once the INSTITUTION frame has been instantiated by either of the above two methods, the other USER-STATUS frames listed above can be instantiated. The frames instantiated in utterance 2 are listed below:-

```
------------------------------------------------------------------------
[UNIVERSITY-COLLEGE
AKO        $value         London-uni

type       $value         terminal
           $if-added      a-London-college-demon
                           (TRIGGER PD(Subj); deduce)]
```

```
------------------------------------------------------------------------
[DEPARTMENT
AKO         $value          institution-slot

type        $require        any-valid-subject
            $value          PHONETICS
            $if-needed      dept-demon
                             (TRIGGER PD(Subj) ;deduce)
            $if-added       dept-add-demon
                             (TRIGGER PD(Subj); subject
                              matter of search can be
                              deduced, AND TRIGGER RS(db);
                              select appropriate db)]

{The effect of the DEPT-ADD-DEMON firing can be seen
later on in the dialogue, in utterances 88 and 92a, when
the intermediary comments on the users' subject field,
and begins choosing appropriate databases.}
------------------------------------------------------------------------
[PHONETICS
AKO         $value          department

type        $value          terminal]
------------------------------------------------------------------------
[LONDON-UNI
AKO         $value          institution

type        $require        a-London-college
            $value          UNIVERSITY-COLLEGE]
------------------------------------------------------------------------
[INSTITUTION
AKO         $value          current-academic-slot

type        $require        London-uni UK-uni foreign-uni
            $default        London-uni
            $value          LONDON-UNI
            $if-needed      instit-demon
                             (IF BACK Residence THEN
                              Foreign-uni)

department-slot
            $require        department]
------------------------------------------------------------------------
[CURRENT-ACADEMIC
AKO         $value          user-status

type        $require        staff student
            $default        student
            $value          STUDENT

institution-slot
            $require        institution
```

```
funding-source-slot
           $require          funding-source

time-slot  $require          time]
------------------------------------------------------------------------
[POSITION
AKO        $value            user-status-slot

type       $require          end intermediary team
           $default          end
           $value            END]
------------------------------------------------------------------------
[END
AKO        $value            position

type       $value            terminal]
------------------------------------------------------------------------
[USER-STATUS
AKO        $value            user-model-slot

type       $require          current-academic
                             former-academic
                             prospective-academic
                             current-non-academic
           $default          current-academic
           $value            CURRENT-ACADEMIC

position   $require          position]
------------------------------------------------------------------------
```

Although the fact that the users are CURRENT-ACADEMICS can be directly inferred as explained above, there is no explicit information in the dialogue relating to the **type** slot in that frame, ie, whether the users are STAFF or STUDENTS. Ideally, the **type** slot **must** be filled in each instantiated frame (see Chapter 6), as this is the principle slot which holds the most important information. In this case, therefore, the default value is accepted, and the value of the **type** slot becomes STUDENT. However, it does seem unlikely that these two users are, in fact, students: they appear to have a great deal of technical knowledge about their subject, they are working together on a joint project, and they have done previous online searches. It is suggested that perhaps in the light of such contradictory evidence, the default value should be overridden in any similar cases in the future. For present purposes, however, the default value is accepted, but the STUDENT frame will not be instantiated unless confirmation should arise further on in the dialogue.

The POSITION frame is instantiated, as this is a required value of one of the subsidiary slots in the USER-STATUS frame. Once again, there is no explicit information in the dialogue as to whether the users are end users, intermediary users, or intermediary users searching on behalf of someone else, so the default value, END, is accepted. In this case, however, it seems likely that the default value is probably correct, and no contradictory evidence can be found in the remainder of the dialogue.

The CURRENT-ACADEMIC frame has three subsidiary slots, **institution-slot**, **funding-source-slot**, and **time-slot**. The **INSTITUTION** frame has already been instantiated, but the FUNDING-SOURCE frame and the TIME frame are not instantiated in this dialogue. This is because there is no explicit information in the dialogue relating to their **type** slots, and there are no default values attached to the **type** slots which could be accepted.

A UGOAL frame is instantiated by utterance 4a, Focus 1, when the users explain that they have a "very specific search" to carry out. The terminal UGOAL frame, SPECIFIC-OUTPUT, is therefore instantiated, and the frames CURRENT-SEARCH-GOAL and UGOAL are obtained by upward-chaining from that frame:-

```
--------------------------------------------------------------------------
[SPECIFIC-OUTPUT
AKO          $value            current-search-goal

type         $value            terminal
             $if-added         specific-output-demon
                               (TRIGGER RG(Out)]

{The effect of SPECIFIC-OUTPUT-DEMON is apparent throughout the
dialogue, and the intermediary explicitly refers to it in utterance
101e.}
--------------------------------------------------------------------------
[CURRENT-SEARCH-GOAL
AKO          $value            ugoal

type         $prefer           specific-output
                               preliminary-search
                               exhaustivity
             $value            SPECIFIC-OUTPUT]
--------------------------------------------------------------------------
[UGOAL
AKO          $value            user-model-slot

type         $require          current-search-goal
                               goals-leading-to-search
                               specific-intentions
                               general- goals
             $value             CURRENT-SEARCH-GOAL]
--------------------------------------------------------------------------
```

Utterances 4b and 6 in Focus 1 invoke the top-level IRS frame, and the **type** value is set to SOME, as the users state they have carried out a search previously. The **type** value of SOME would appear to be set to INTERMEDIATE, as (at this stage) only **one** search is mentioned, although this value could be adjusted should contradictory evidence appear further on in the dialogue. The MEMORY-FOR-PROCEDURES frame is a subsidiary slot of the SOME frame and should ideally be instantiated, but it is not possible to assign a value to the **type** slot at this stage in the dialogue, ie, it is not yet clear whether or not the users remember the procedures involved in online searching. It is suggested that perhaps this frame could be placed on a queue, and instantiated when an explicit value becomes available. The value of the **type** slot of the CHARACTERISTICS frame can be found: the search was carried out at UCL's own search service, which is A-SERVICE, rather than THIS-SERVICE or A-COUNTRY. The CHARACTERISTICS frame also has a number of subsidiary slots: **when-slot**, **presence-slot**, **host-slot**, **db-slot** and **satisfaction-slot**. With the exception of PRESENCE, which has a default value for its **type** slot, none of these frames can be instantiated yet : there are no default values available for the **type** slots and, at this early stage in the dialogue, there is no explicit information forthcoming from the users. Again, these frames should probably be activated and placed on a queue to await instantiation. The instantiated IRS frames for Focus 1 are presented below:-

----

```
[IRS
AKO         $value          user-model-slot
type        $require        none some
            $value          SOME
            $if-needed      IRS-demon
                              (IF USER-STATUS Student
                                THEN None AND IF BACK
                                 Acad Former THEN None)
-------------------------------------------------------------------
[SOME
AKO         $value          IRS
type        $require        intermediate considerable
            $value          INTERMEDIATE

memory-for-procedures-slot
            $require        memory-for-procedures

characteristics-slot
            $require        characteristics]
-------------------------------------------------------------------
[INTERMEDIATE
AKO         $value          some

type        $value          terminal]
-------------------------------------------------------------------
[CHARACTERISTICS
AKO         $value          some-slot
type        $require        a-country a-service
                            this-service
            $value          A-SERVICE

when-slot   $require        when

presence-slot
            $require        presence

host-slot   $require        host

db-slot     $require        db

satisfaction-slot
            $require        satisfaction]
-------------------------------------------------------------------
[A-SERVICE
AKO         $value          characteristics

type        $require        a-named-service
            $value          UCL-SERVICE]
-------------------------------------------------------------------
[UCL-SERVICE
```

```
AKO         $value            a-service

type        $value            terminal]
-------------------------------------------------------------------------
[PRESENCE
AKO         $value            characteristics-slot

type        $require          present absent
            $default          present
            $value            PRESENT]
-------------------------------------------------------------------------
[PRESENT
AKO         $value            presence

type        $value            terminal]
-------------------------------------------------------------------------
```

The first part of utterance 11 in Focus 1 confirms the SPECIFIC-OUTPUT frame of the UGOAL subfunction: the user reiterates that he wants to carry out a very narrow, ie, specific search. If confidence ratings were to be attached to the frames, the rating attached to SPECIFIC-OUTPUT would now be incremented. The second part of utterance 11 also seems to concern the users' goal:-

... we may catch something that we haven't caught before ...

It would seem appropriate to classify this part-utterance as a CURRENT-SEARCH-GOAL, and for present purposes the value of the **type** slot could be set to EXHAUSTIVITY, ie, carry out an exhaustive search, but it could be argued that such a categorisation is not wholly accurate and does not adequately capture the user's precise goal. In that case, it would become necessary to add a further value to the $require facet of the **type** slot of the CURRENT-SEARCH-GOAL, for example, BE-CERTAIN-OF-MISSING-NOTHING. The current UGOAL frames for this part-utterance are presented below:-

```
-------------------------------------------------------------------------
[EXHAUSTIVITY
AKO         $value              current-search-goal

type        $value              terminal
            $if-added           exhaustivity-demon
                                  (TRIGGER RG(Out);
                                       be exhaustive]

{The effect of EXHAUSTIVITY-DEMON firing can be seen
towards the end of the dialogue, when the intermediary
plans to apply the search strategy to as many different
databases as possible}
-------------------------------------------------------------------------
[CURRENT-SEARCH-GOAL
AKO         $value               ugoal

type        $require            specific-output
                                preliminary-search
                                exhaustivity
            $value              EXHAUSTIVITY]
-------------------------------------------------------------------------
```

```
[UGOAL
AKO          $value           user-model-slot

type         $require         current-search-goal
                              goals-leading-to-search
                              specific-intentions
                              general-goals
             $value           CURRENT-SEARCH-GOAL]
---------------------------------------------------------------------------
```

In the third focus, the Intermediary confirms the SPECIFIC-OUTPUT frame in utterance 33: if confidence ratings were attached, the rating attached to that frame would be incremented.

In Focus 9, further IRS frames are instantiated in utterances 77b, 78, 79, 80, 81 and 83. It was not possible to fill out all of the subframes of the CHARACTERISTICS frame in Focus 1: the Intermediary now returns to that frame and begins by attempting to instantiate the DB frame. It is also likely that the value of the **type** slot in the PRESENCE frame can be confirmed as PRESENT, although this is still not explicitly stated. The frames WHEN and HOST still cannot be instantiated, but the SATISFACTION frame can now be filled out by information forthcoming in utterance 83. Although the user does not actually state that he was dissatisfied with the search carried out at UCL, it does appear that the search was too general and should therefore not be updated or repeated.

At this point, it appears that an extra **characteristics** slot will be required: the YES and NO demons should only be invoked if the user is doing a search on a similar or related topic. The fact that the user is dissatisfied with the search on Inspec would become quite irrelevant to the current search situation if he had been searching on an entirely different topic or subject.

The current IRS frames for Focus 9 are presented below:-

```
---------------------------------------------------------------------------
[DB
AKO          $value           characteristics-slot

type         $require         a-named-db
             $value           INSPEC]
---------------------------------------------------------------------------
[INSPEC
AKO          $value           db

type         $value           terminal
             $if-added        named-db-demon
                                (TRIGGER consistency check
                                  with A-NAMED-HOST]
---------------------------------------------------------------------------
[SATISFACTION
AKO          $value           characteristics-slot

type         $require         yes no
             $value           NO]
---------------------------------------------------------------------------
[NO
AKO          $value           satisfaction

type         $value           terminal
```

```
     $if-added       no-demon
                      (TRIGGER RS(db); avoid
                       searching named dbs if
                       search topic is similar)]

{The NO-DEMON is fired: the Intermediary does not search Inspec}
-------------------------------------------------------------------
```

In utterance 94 in Focus 10, the Intermediary suggests some possible databases to the users:-

> ... there's there's a multidisciplinary one you know the Current Contents database Sci-search and Social well probably more Science Citation Index ...

In utterance 95 the user replies:-

> ... yes we don't know much about that one ...

This utterance is problematic: neither the KNOW frames nor the IRS frames can currently account adequately for this utterance, which concerns the user's knowledge of a particular database. The KNOW frames can cover knowledge of INFORMATION SOURCES, ie, books, journals, newspapers and magazines, but they do not include a category for **online** information sources. Similarly, the IRS frames can handle databases which have been searched before, but do not include any frames or slots for databases with which the user is **unfamiliar**. It may be necessary to create a further category for ignorance of, or lack of experience with, particular online information sources, and it seems likely that such a category could be included within the IRS subfunction.

7.4 Discussion

The aim of this analysis was to discover whether the user modelling frames, in their current state, can adequately account for all the utterances in a dialogue which had not been used in the original analysis. In general, the five broad user modelling subfunctions were able to account for all the utterances which were concerned with user modelling, and it has not been necessary to create any further subfunctions at this stage. However, it may be necessary to create some further slots, or slot values, within the existing frames, eg, within the UGOAL and IRS subfunctions as discussed in Section 7.3 above. This need not be regarded as a major shortcoming of the current formalism: the addition of new slots and slot values to existing frames is straightforward, and is inevitable if the current model is to be extended at some stage to apply to other types of information service, or even other categories of information interaction.

In one case, it was necessary to accept a default value for a Type slot which was probably inaccurate, ie, it is unlikely that the two users in Interview 131185pda were STUDENTS. In all other cases, the default values which were accepted would appear to be correct. In this real-life dialogue, it was probably obvious to the human intermediary that these users were not students, and therefore there was no need to attempt to confirm or disprove this hypothesis during the dialogue. An interesting consideration is whether the process by which the human intermediary made this deduction can be automated: it is most unlikely that the eventual system, however sophisticated, will be able to take account of evidence such as physical appearance, accent and tone, bearing and manner, etc (although it is predicted that the eventual version of KNOESPHERE (Lenat et al 1983) will be able to respond to precisely these sorts of cues !). To overcome this kind of problem, confidence ratings or certainty values could be attached to slot values, and these could be adjusted appropriately as the dialogue proceeds, in the light of confirmatory or contradictory evidence. In order to do this, further work will be required to identify types of evidence, and how and when the intermediary uses this evidence in preference to explicit questioning of the user. A useful additional exercise may be to interview both user and intermediary **after** the frame instantiation has taken place, in order to confirm default values, demons fired, etc.

In Interview 131185pda, some frames were instantiated (ie, their Type slots were filled) without all of their subsidiary slots being filled, eg, the CHARACTERISTICS frame in Focus 1. In such cases, it is likely that the subframes which constitute the slot-fillers could be placed on a queue to await instantiation, ie, when an explicit, or default, value is not available to fill the Type slot of a frame, that frame is nonetheless activated or triggered to search for, or elicit, values.

Chapter 8

DISCUSSION AND CONCLUSIONS

8.1 Discussion

8.1.1 Overview

One of the functions that must be performed by the intermediary component in an information provision mechanism is that of user modelling. The aim of the research undertaken in this thesis was to investigate the user modelling function within the context of an academic online information retrieval service. The goals of the investigation were to specify the components of the User Model and the knowledge resources required by the intermediary in order to construct the model; to identify how the User Model interacts with the other functions; to identify an appropriate formalism in which the user model may be implemented in an operational interface; and to undertake some form of validation of the adequacy of the User Model and its subfunctions.

The investigation was carried out by employing a variety of methods, the principal one of which was the detailed analysis of real-life human-human interactions, ie, the presearch dialogue between the user and the intermediary. This analysis enabled the user modelling function and its subfunctions to be specified, together with its knowledge resources and its interactions with the other functions. This analysis was supplemented by additional data in the form of transcribed problem statements and their accompanying recorded interactions (see Belkin and Kwasnik 1986). Interviews with three intermediaries were also arranged, one of whom worked in a search service which had **not** participated in the research project, in order to verify the User Model and to obtain further information on its interactions with the other functions. A formalism for representing the User Model was developed, and a validation was attempted, the goal of which was to demonstrate that the User Model and its representation is independent of the data on which it is based.

The ultimate aim of this research is to use the results to design and implement the user modelling function in an intelligent interface to document retrieval systems. The interface design is based on the MONSTRAT model (Belkin, Seeger and Wersig 1983), which comprises a number of functions or experts, each responsible for completing a specific function, yet dependent on each other for information necessary for accomplishing their tasks. The User Model is one of these experts, and its development and design depends on investigating how a human intermediary carries out this function.

The discourse analysis, the interviews and the validation have resulted in a detailed specification of the User Model, which can now be used in an implementation. Although a formalism has been proposed for the User Model, the specification of the model is essentially independent of any committment to system design or knowledge representation techniques, and its value is not restricted to its implementational utility. Very little is known about intermediaries' model building activities and although it is generally accepted that a User Model is probably necessary in an intelligent interface, most research into user modelling for operational interfaces has taken place in constrained domains, eg, computer assisted instruction, where strong simplifying assumptions can be made. Few researchers in the IR field have yet attempted to identify, let alone implement, a comprehensive User Model, and there is little consensus regarding the utility and feasibility of user modelling.

It is evident from the results of this research that intermediaries **do** construct User Models, that users also realise the importance of these models, and that the User Model provides vital input to the other system functions. It is therefore desirable that human intermediaries are capable of carrying out this kind of cognitive modelling, and that they are able to cooperate with the user in order to elicit the necessary information. It may be possible to use the transcripts for teaching materials for training intermediaries in interviewing techniques.

The analysis also provides an indication of where interviews "fail" and the possible reasons for such failures. For example, the intermediary in interview 190684hba failed to take account of an important user modelling subfunction, the user's goal, despite the fact that the UGOAL subfunction persistently recurred throughout the dialogue. It would be useful to ensure that the interface would be capable of identifying a "malfunctioning" dialogue, and its causes, so that it could take remedial action.

8.1.2 The User Modelling Function

It is apparent that human intermediaries and users do engage in user modelling, and that the User Model is a separate, necessary function whose aim is to model various aspects and characteristics of the user, rather than aspects of his/her **problem**. User modelling is a **negotiated** process: the intermediary is dependent upon the user's contribution for the accomplishment of this function. The users appear to have an awareness (possibly unconscious) of the significance of user modelling, as they frequently volunteer user modelling information themselves without needing to be prompted. The User Model is a multi-component function, ie, it comprises five elements or subfunctions, the goal of each being to describe or model a different aspect of the user. Although user modelling certainly took place in all seven transcripts, the extent to which each subfunction occurred varied between interviews, with only the UGOAL subfunction appearing in all interviews. This variation is probably due to differences in the perceived importance of the goals by both participants, other dialogue features including the predominance of other functions such as Problem Description, and also whether the intermediary already has a model of this user.

It was possible to identify the types of knowledge resources which the intermediary needs to access in order to accomplish the goal of user modelling. Unlike those of Problem Description, the knowledge resources of the User Model are mainly **internal**, ie, the intermediary employs her own internalised knowledge about users and their characteristics, rather than using external sources such as manuals and thesauri.

The interviews with the intermediaries confirmed that user modelling is important in the pre-online interaction, and clarified what is necessary for accomplishing this goal, in contrast to other goals which are accomplished during the dialogue, such as phatic communication and discourse goals, which are essential for the overall success of the dialogue but which are not **functionally** necessary for user modelling.

8.1.3 The Interaction of the User Model with the other functions

The analysis of the transcripts, and the interviews with the intermediaries, revealed that the User Model interacts with the other interface functions, and that the user modelling subfunctions interact with each other. The intermediary interviews provided some particularly valuable detailed information, and in many cases it was possible to specify which subfunctions, in addition to functions, were interacting with the user modelling subfunctions. Only one function was found to contribute information to the User Model, ie, the Problem State function, and it can be concluded that the User Model is rarely dependent on the other functions for information necessary for its own processing. It obtains the information it needs through other means, such as explicit questioning of the user. However, to be certain that the User Model does not require input from the other functions, further analysis needs to be undertaken on the other functions and their processing and behaviour.

The User Model is used by all of the other functions, to varying extents, and this type of interaction demonstrates that the existence of the User Model is justified, although it appears to constitute a relatively small proportion of each dialogue. It is interesting that the User Model is particularly heavily used by the Retrieval Strategies function: Brooks (1986) has pointed out that the Retrieval Strategies function requires very little input from the user as the expertise necessary to formulate a strategy lies largely with the intermediary. However, the intermediary is **indirectly** dependent on the user's input, as the User Model appears to provide a wealth of information to the Retrieval Strategies function.

The subfunctions of the User Model also interact with each other: the User Model is highly dynamic, and the hypotheses generated by one subfunction may need to be revised and updated several times during the course of the interaction as a result of information received from another subfunction.

The interaction of the User Model with the other functions and subfunctions is an issue which requires further investigation. Many detailed interactions have been identified as a result of the research described in this thesis, but a larger corpus of interview transcripts is required in order for **quantitative** analysis to be undertaken. Although many interactions have been confirmed by the intermediary interviews, some have been observed only once or twice, and a comparative analysis of the interaction patterns between dialogues is required.

8.1.4 The Methodology for Eliciting the User Model

> We need to be eclectic in acquiring knowledge engineering techniques from all possible sources ...
>
> (Gaines 1986)

The primary method used in this research for specifying the user modelling function, its knowledge resources and interactions, was functional discourse analysis of real life interactions. Boguraev (1985) has advocated analysing real interactions between people in order to understand how they build models, deploy knowledge, and manage problems, and discourse analysis is the ideal method. It has provided a considerable amount of detailed information regarding the model building activities of the intermediary, the knowledge resources employed and the types of interactions which occur. Furthermore, this technique enabled the dialogues to be partitioned into foci, a useful means of identifying current discourse goals and topics in the focus of attention, and drew attention to the process of negotiation between user and intermediary. Aspects such as these are of central importance to the design of an interactive interface, which should ideally be able to handle user-initiated inputs. The discourse analysis has revealed the cooperative, mixed-initiative nature of the dialogues, and has elicited a large amount of very detailed expert knowledge. This level of detail should hopefully allow complex and accurate interface implementation.

However, as Labov and Fanshel (1977) discovered in their analysis of a fifteen minute segment of a therapeutic interview, discourse analysis has its disadvantages, the principal one of which is its time-consuming complexity. Each of the seven interviews used in this research took several months to record, transcribe, partition into utterances and foci, and analyse, and a great deal of effort and time is required to compile a large corpus of analysed dialogues. It is therefore difficult, at present, to carry out any **quantitative**, statistical analysis on the data, as a much larger data set is required in order for such analysis to be meaningful.

The discourse analysis was supplemented by the informal analysis of the Problem Statements and their accompanying recorded interactions, by the interviews with the three intermediaries, and by a subsequent validation of the User Model. Current thinking in knowledge elicitation techniques supports the use of multiple techniques, and the supplementary methods employed in this research have produced a more detailed set of user modelling subfunctions, and have enabled the interaction of the User Model to be specified more precisely. The analysis of the problem statements and the recordings confirmed that no further subfunctions of the User Model were necessary, and also enabled some further subdivisions **within** subfunctions to be specified.

The intermediary interviews reinforced the results obtained from the transcript analyses, and provided further, detailed information on the interaction of the User Model with the other functions and subfunctions. The interview with the third intermediary from a different search service demonstrated that the User Model appeared to be applicable to information interactions within **other** academic institutions, and that similar interactions took place.

The goal of the validation described in Chapter 7 was to demonstrate that the User Model and its formalism is independent of the data on which it is based. This aim was largely fulfilled, and is discussed in Section 8.1.6, this chapter.

In conclusion, the combination of methodologies used in this research, ie, discourse analysis, interviewing of experts, and a subsequent validation, has proved useful: the principal method of discourse analysis elicited the essential structure and behaviour of the User Model, and the interviews and verification confirmed its validity and provided further detailed information.

8.1.5 A Formalism for the User Model

A frame-based formalism for the User Model has been proposed, which is appropriate both for the type of knowledge that the User Model employs, and for implementation within a blackboard system. It has been possible to derive a standard set of frames (the complete hierarchy of frames is presented in Appendix F) embodying the user knowledge utilised by human intermediaries in the academic search services studied in this project, and the interactions identified between the User Model, its own subfunctions and the other system functions have been represented as procedures or demons. Although certain difficulties with this representation technique have been identified in Chapter 6, such as the possible need for specialised inheritance techniques and the question of instantiation of multiple types, it should soon be possible to attempt an implementation of a frame-based User Model.

The frame-based formalism has the advantage of adaptability: a standard set of frames can be derived for a particular type of search service, with perhaps individual modifications, such as additional slots, procedures or values, being made to the core set of frames according to the characteristics of each individual service or institution. Similarly, the present set of frames can easily be incremented or reduced, in the event of further subdivisions or additional interactions being identified as a result of further research.

8.1.6 The Validation of the User Model

The validation largely succeeded in demonstrating that the User Model which has been proposed can adequately account for a **new** situation, ie, the User Model is applicable to a dialogue which had **not** been used as data for the original analysis. At the same time, the utility of the formalism was demonstrated, in that it was able to handle the majority of the user modelling utterances within the dialogue. The validation revealed that it may be necessary to create some further slots or slot values within the UGOAL and IRS subfunctions, an issue which will be straightforward to resolve.

However, it is accepted that the general issue of validation is still a problem: at this stage in the research, it is impossible to judge the generality of the User Model as only one attempt at validation has been undertaken. It would certainly be advisable to apply the User

Model to a dialogue recorded at a **different** academic search service, ie, one which had not participated in any way in the original research project. Furthermore, the goal of the overall system is not analytic but generational,ie, it is not the intention of the system to **analyse** dialogues retrospectively, but to generate them in a similar fashion to a human intermediary. Therefore, a more appropriate means of validating the User Model may be to attempt to use it to **generate** an interaction, for example, it may be possible to find out whether a simulated interaction similar to Interview 131185pda could be generated from the first few foci of that dialogue, but this kind of investigation would require the complete functional analysis of the remaining interface functions, together with the detailed specification of their interactions. It seems likely that some kind of scheduler would also be required, in order to monitor the behaviour of the functions and to assign priorities to each at different stages of the dialogue.

Nonetheless, the validation has at least demonstrated that the User Model is independent of the data on which it is based, and that it can interpret or analyse a new interaction successfully.

8.2 Conclusions

8.2.1 Further Research on the User Model

The research described in this thesis is limited in that only **one** type of information interaction has been investigated, and it is therefore not possible to assess the generality of the User Model. It would be useful to carry out some further investigation into the interactions which take place in other settings, both in other document retrieval systems and in other types of information provision mechanisms, such as public libraries, advisory services etc. in order to find out what types of user models are constructed, whether they comprise similar subfunctions, and how they are used.

It seems likely that the User Model which has been proposed as a result of this research will need a method to resolve conflicts between competing inferences and values, and to represent how confident it is that certain values are accurate. Grundy (Rich 1979a, 1979b) uses ratings representing how confident it is that certain pieces of information are correct, and to overcome the problem encountered in the validation described in Chapter 7, ie, it was necessary to accept a default value which was almost certainly inaccurate, confidence ratings or certainty values should be attached to slot values which could be adjusted as the dialogue proceeds. If such an improvement is to be implemented, then a method for empirically determining these values or ratings will be required.

A further capability that the User Model is likely to require is that of learning. It has been emphasized that the User Model is **dynamic**, and is continually updated and adjusted throughout the interaction. It would be useful if the system could both adjust the structure of the User Model as its experience of users and their attributes increases over sessions, and also revise its models of individual users in those cases where a user interacts with the system on more than one occasion. The user model proposed for Knoesphere (Lenat et al. 1983) will be capable of learning and updating its models of both **individual** users and of **groups** of people as well, for example, if mathematicians become less averse to examples over a period of years, the mathematician model would gradually change to reflect this. It will clearly be more straightforward to implement a capability for updating individual user models: if a user is known to have interacted with the system last week, for example, it should be easy to increment the amount of IRS experience that she has.

At present, the IRS subfunction interacts heavily with the Explanation function: if a user has little, or no, IRS experience, then an appropriate explanation should be issued. However, this issue needs further investigation and in particular, it is probably necessary to attempt to discover exactly how the intermediary, whether human or automatic, is able to construct a model of the user's model of the system. This model is no doubt constructed partially from information from the IRS subfunction, but probably also requires input from elsewhere, perhaps from the UGOAL subfunction of the User Model and perhaps from the CAPAB subfunction.

The User Model is constructed not only from explicit user modelling utterances, but from other sources such as nonverbal communication, initiative-taking in the dialogue, the user's use of specialised terms and vocabulary, and extralinguistic phenomena such as silences and hesitations. The human intermediaries readily assimilate and utilize such information, but the recognition of such information by a **machine** intermediary will be immensely difficult. However, in order to construct a complex and complete User Model, it will be helpful if the interface can take account of such phenomena. The problem is that in order to recognize a wide range of user properties, goals, attitudes, etc. the system will need to deploy extensive world, domain and user knowledge.

8.2.2 Further Research on the Intelligent Interface

The work described here concentrated on just one of the functions that must be performed by an intelligent interface for document retrieval systems. The Problem Description and Retrieval Strategy functions have been investigated by Brooks (1986), but the remaining functions are largely unexplored and need to be specified to a similar level of detail.

Furthermore, the overall human-computer dialogue structure needs to be investigated, ie, a means for driving and interpreting a human-computer dialogue in the information interaction context needs to be investigated.

Finally, it has been suggested that the ideal form of interaction with the system is natural language, in which case we need to consider the problem of automating the interpretation of input utterances to determine what their topics are.

8.2.3 Implementing an Intelligent Interface

The analysis of the User Model has revealed that an additional layer of complexity amongst the distributed functions, ie, the subfunction level, will need to be taken into account for implementation, and that the system will require a wide range and large amount of knowledge. These are certainly non-trivial problems for any implementation.

The implementation of some, or all, of the prototype functions would involve the following:

- developing the software for carrying out the tasks each function must perform. Mechanisms must be developed which will simulate the problem structures observed in terms of goals and subgoals and the means for achieving these goals must be provided. As Sparck Jones (1987) has pointed out, a way of determining that the functions' goals have been satisfied is required. It is likely that some method of scheduling an agenda of tasks is needed, depending on an evaluation of the various functions' claims and on the consequences of overall constraints, such as the system should work on Problem Description before it works on Retrieval Strategies. The problem is that the system is both data **and** function driven: it needs to respond flexibly to user input, yet attempt to pursue its own goals.

- implementing the knowledge resources required by each function. This includes software for representing the knowledge in the formalism most appropriate for that function. Software for easy entry, verification and consistency checking of specific components of the expert knowledge must also be built.

- software to enable the functions to use and apply their knowledge resources must be constructed. This would, on the whole, take the form of implementing the inferencing mechanisms associated with each representation type.

- developing mechanisms for carrying out the model building activities that need to be performed by whichever functions are to be implemented.

- defining and developing the exact nature of the input/output communication (ie, function to blackboard, and blackboard to function) for each function. As it is quite likely that any expert may communicate with any other expert, a general purpose language may be required. The messages passed can be complex (eg, see Brooks 1986) and Sparck Jones (1987) suggests that they will therefore need to be encoded in a language which approaches the complexity of natural language. This issue is further complicated by the fact that the different functions will no doubt be represented in different formalisms, for example, frames for the User Model and partitioned semantic networks for Problem Description.

8.2.4 Concluding Remarks

Still much remains to be done before the long-term goal of implementing an intelligent interface for document retrieval systems is achieved. However, in addition to the primary goal of developing the user modelling function, this research has demonstrated that the intermediary's task is a complex one and that a considerable amount of time and effort is required in order to acquire this complex, detailed knowledge. Although Sparck Jones (1987) warns that

> ...it is clear that building the all-singing all-dancing expert intermediary is a major enterprise

this research has achieved some insights in other areas, notably cognitive model building in information interaction, approaches to training human online intermediaries in interview techniques, the application of expert system techniques to complex domains, and an appreciation of the skills and expertise of human intermediaries.

REFERENCES

ALTY, J.L. and COOMBS, M.J. (1980) Face-to-face guidance of university computer users - I: a study of advisory services. *International Journal of Man-machine Studies*, vol. 12, 1980, pp. 390-406.

BARR, A. and FEIGENBAUM, E.A. (eds) (1981) *The Handbook of Artificial Intelligence.* Vols. 1-3. Los Altos, C.A., William Kaufmann, 1981.

BARTLETT, F.C. (1932) *Remembering: a study in experimental and social psychology.* Cambridge, England, CUP, 1932.

BELKIN, N.J. (1977) *A concept of information for information science.* Ph.D Thesis. University of London, 1977.

BELKIN, N.J. (1980) Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, vol. 5, 1980, pp. 133-143.

BELKIN, N.J. (1984) Cognitive models and information transfer. *Social Science Information Studies*, vol.4, 1984, pp. 111-129.

BELKIN, N.J., BROOKS, H.M. and DANIELS, P.J. (1986) Knowledge Elicitation using Discourse Analysis. In: *Knowledge Acquisition for Knowledge-Based Systems Workshop.* Sponsored by the American Association for Artificial Intelligence, Banff, Canada, November 3-7th, 1986 (publication forthcoming).

BELKIN, N.J. and HAPESHI, K. *Representation and matching of anomalous states of knowledge and document texts.* Final report on BLR&D department grant no. SI/G/566. London, the City University, Department of Information Science (in preparation).

BELKIN, N.J., HENNINGS, R-D. and SEEGER, T (1984) Simulation of a distributed expert-based information provision mechanism. *Information Technology: Research Development Applications*, vol. 3, no. 3, 1984, pp. 122-141.

BELKIN, N.J. and KWASNIK, B.H. (1986) Using structural representations of anomalous states of knowledge for choosing document retrieval strategies. In: *Proceedings of the joint ACM/BCS International Conference on R & D in IR.* Pisa, 1986 (in press).

BELKIN, N.J., ODDY R.N. and BROOKS, H.M. (1982a) ASK for information retrieval, part I. *Journal of Documentation*, vol. 38, 1982, pp. 61-71.

BELKIN, N.J., ODDY, R.N. and BROOKS, H.M (1982b) ASK for information retrieval, part II. *Journal of Documentation*, vol. 38, 1982, pp. 145-164.

BELKIN, N.J., SEEGER, T. and WERSIG, G. (1983) Distributed expert problem treatment as a model for information system analysis and design. *Journal of Information Science*, vol.5, 1983, pp. 153-167.

BELKIN, N.J. and VICKERY, A. (1985) *Interaction in information systems: a review of research from document retrieval to knowledge-based systems.* LIR report no. 35. London, The British Library, 1986.

BELKIN, N.J. and WINDEL, G. (1984) Using MONSTRAT for the analysis of information interaction. In: *Representation and Exchange of Knowledge as a Basis of Information Processes.* Ed. by H.J. Dietshmann. Amsterdam, Elsevier, 1984, pp. 359-382.

BERRY, D.C. and BROADBENT, D.E. (1984) On the relationship between task performance and associated verbal knowledge. *Quarterly Journal of Experimental Psychology*, vol. 36A, no. 2, 1984, pp. 209-291.

KAPLAN, R.M., KAY, M, NORMAN, D.A., THOMPSON, H, and WINOGRAD, T

BOBROW, D.G. et al (1977) GUS, a frame-based dialog system. *Artificial Intelligence*, vol.8, 1977, pp. 155-173.

BOGURAEV, B.K. (1985) User modelling in cooperative natural language front ends. In: *Social action and artificial intelligence.* Ed. by G. Gilbert and C. Heath. London, Gower Press, 1985, pp. 124-143.

BORGMAN, C.L. (1984) Psychological research in human-computer interaction. *Annual Review of Information Science and Technology*, vol. 19, 1984, pp. 33-64.

BROOKS, H.M. (1985) Developing and using problem descriptions. In: *IRFIS 6: Intelligent information systems for the information society.* Frascati, Italy, Sept., 1985. Amsterdam, North Holland, 1986, pp. 141-161.

BROOKS, H.M. (1986) *An intelligent interface for document retrieval systems: developing the problem description and retrieval strategy components.* Ph.D Thesis, the City University, London, 1986.

BROOKS, H.M. and BELKIN, N.J. (1983) Using discourse analysis for the design of information retrieval mechanisms. In: *Research and Development in Information Retrieval.* Proceedings of the Sixth Annual International ACM SIGIR conference, Washington D.C., 1983, New York, ACM, 1983, pp. 31-47.

BROOKS, H.M., DANIELS, P.J. and BELKIN, N.J. (1985) Problem Descriptions and User Models: developing an intelligent interface for document retrieval systems. In: *Informatics 8: Advances in Intelligent Retrieval.* London, ASLIB, 1986, pp. 191-214.

BUCHANAN, B.G. (1982) New research on expert systems. In: *Machine Intelligence 10.* Ed. by J.E. Hayes, D. Michie and Y.H. Pao. Edinburgh, Edinburgh University Press, 1982, pp. 269-299.

BARSTOW, D, BECHTEL, R, BENNETT, J., CLANCEY, W, KULIKOWSKY, C, MITCHELL, T and WATERMAN, D.A.

BUCHANAN, B.G. et al (1983) Constructing an expert system. In: *Building Expert Systems.* Ed. by F. Hayes-Roth, D.A. Waterman and D.B. Lenat. Reading, Mass. Addison-Wesley, 1983, pp. 127-168.

BURTON, R.R. (1982) Diagnosing bugs in a simple procedural skill. In: *Intelligent Tutoring Systems.* Ed. by D. Sleeman and J.S. Brown. London, Academic Press, 1982, pp. 157-182.

BURTON, R.R. and BROWN, J.S. (1982) An investigation of computer coaching for informal learning activities. In: *Intelligent Tutoring Systems.* Ed. by D. Sleeman and J.S. Brown. London, Academic Press, 1982, pp. 79-98.

CAMPBELL, J.A. (1986) On artificial intelligence. *AI Review*, vol. 1, 1986, pp. 3-9.

CARBERRY, S. (1983) Tracking user goals in an information-seeking environment. *Proceedings of the National Conference of the American Association for Artificial Intelligence*, 1983, pp. 59-63.

CLANCEY, W.J. (1982) Tutoring rules for guiding a case method dialogue. In: *Intelligent Tutoring Systems.* Ed. by D. Sleeman and J.S. Brown. London, Academic Press, 1982, 201-225.

HOPKINS, C., COLE, I., HOCKLEY, A., and ARSHAD, F

CLOWES, I et al. (1985) User modelling techniques for interactive systems. In: *People and computers: designing the interface.* Proceedings of the conference of the BCS Human Computer Interaction Specialist Group, University of East Anglia, 17-20 September, 1985, pp. 35-45.

COCHRANE, P. (1981) *Presearch interview project: executive summary, project bibiography and sample data.* Report no. ED 205 190, Syracuse, N.Y., School of Information Studies, Syracuse University, 1981.

COOMBS, M.J. and ALTY, J.L. (1980) Face-to-face guidance of university computer users - II: characterizing advisory interactions. *International Journal of Man-machine studies*, vol. 12, 1980, pp. 407-429.

COOMBS, M.J. and ALTY, J.L. (1984) Expert systems: an alternative paradigm. *International Journal of Man-machine studies*, vol. 20, 1984, pp. 21-43.

CROFT, W.B. (1985) An expert assistant for a document retrieval system. In: *RIAO-85. Actes* of the conference: recherche d'informations assistee par ordinateur, Grenoble, March 1985. Grenoble, I.M.A.G., 1985, pp. 131-149.

CROFT, W.B. and THOMPSON, R.H. (1985) *An expert system assistant for document retrieval systems.* COINS Technical Report 85-05. Computer and Information Science Department, University of Massachusetts, 1985.

DANIELS, P.J. (1985) The user modelling function of an intelligent interface for document retrieval systems. In: *IRFIS 6: intelligent information systems for the information society.* Frascati, Italy, Sept., 1985. Amsterdam, North Holland, 1986, pp. 162-176.

DANIELS, P.J. (1986) Cognitive models in information retrieval - an evaluative review. *Journal of Documentation*, vol. 42, no. 4, December 1986, pp. 272-304.

DANIELS, P.J., BROOKS, H.M. and BELKIN, N.J. (1985) Using problem structures for driving human-computer dialogues. In: *RIAO-85.* Actes of the conference: recherche d'informations assistee par ordinateur, Grenoble, March 1985. Grenoble, I.M.A.G., 1985, pp. 645-660.

DE MEY, M. (1977) The cognitive viewpoint: its development and its scope. *Communication and Cognition*, vol. 10, no. 2, 1977, pp. 7-23.

ERMAN, L.D. and LESSER, V.R. (1979) The Hearsay-II system: a tutorial. In: *Trends in Speech Recognition*. Ed. by W.A. Lea. U.S.A., Prentice-Hall, 1979, pp. 361-381.

ERICSSON, K.A. and SIMON, H.A. (1980) Verbal reports as data. *Psychological Review*, vol. 87, no. 3, 1980, pp. 215-251.

FIDEL, R. (1986) Towards expert systems for the selection of search keys. *Journal of the American Society for Information Science*, vol. 37, no. 1, 1986, pp. 37-44.

FIKES, R. and KEHLER, T. (1985) The role of frame-based representation in reasoning. *Communications of the ACM*, vol. 28, no. 9, 1985, pp. 904-920.

FOX, E.A. and FRANCE, R.K. (1986a) *A knowledge-based system for Composite Document Analysis and Retrieval: Design issues in the CODER project*, TR-86-6, Virginia Tech Dept. of Computer Science, March 1986.

FOX, E.A. and FRANCE, R.K. (1986b) *Architecture of an Object-oriented Expert System for Composite Document Analysis, Representation and Retrieval*, TR-86-10, Virginia Tech Dept. of Computer Science, April 1986.

GAINES, B.R. (1986) Foundations of knowledge engineering. In: *Research and Development in Expert Systems III: Proceedings* of Expert Systems 86, the Sixth Annual Technical Conference of the BCS Specialist Group on Expert Systems, Brighton, England, 15-18th December, 1986. Ed. by M.A. Bramer, pp. 13-24.

GAMMACK, J.G. and YOUNG, R.M. (1985) Psychological techniques for eliciting expert system knowledge. In: *Research and Development in Expert Systems.* Proceedings of the 4th Technical Conference of the BCS Specialist Group on Expert Systems, University of Warwick, 18-20th December 1984. Ed. by M.A. Bramer. Cambridge, Cambridge University University Press, 1985, pp. 105-112.

GENESERETH, M.R. (1982) The role of plans in intelligent teaching systems. In: *Intelligent tutoring systems.* Ed. by D. Sleeman and J.S. Brown. London, Academic Press, 1982, pp. 137-155.

GERSHMAN, D. (1981) Figuring out what the user wants: steps towards an automatic yellow pages assistant. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, Vancouver, 1981, pp. 423-425.

GOLDSTEIN, I.P. (1982) The genetic graph: a representation for the evaluation of procedural knowledge. In: *Intelligent tutoring systems.* Ed. by D. Sleeman and J.S. Brown. London, Academic Press, 1982, pp. 51-77.

GOLDSTEIN, I.P and ROBERTS, R.B. (1979) Using frames in scheduling. In: *Artificial Intelligence: an MIT Perspective*, vol. 1. Ed. by P.H. Winston and R.H. Brown. The MIT Press, 1979, pp. 257-284.

GOLDSTEIN, I.P and ROBERTS, R.B. (1980) NUDGE, a knowledge-based scheduling program. In: *Frame conceptions and text understanding.* Ed. by D. Metzing. New York, Walter de Gruyter, 1980.

GROSZ, B.J. (1977) *The representation and use of focus in dialogue understanding.* Menlo Park CA, Stanford Research Institute, July 1977. Technical Note 151.

GROSZ, B.J. (1978) Discourse knowledge. In: *Understanding spoken language.* Ed. by D.E. Walker. New York, Elsevier-North Holland, 1978, pp. 229-346.

GUIDA, G. and TASSO, C. (1983) IR-NLI: an expert natural language interface to online databases. In: *Proceedings of the Conference on Applied Natural Language Processing*, Santa Monica, California, February 1-3rd, 1983, USA, ACL, 1983, pp. 31-38.

HAPESHI, K. and BELKIN, N.J (1985) Developing and evaluating an online information retrieval system based on user problem statements. In: *RIAO-85.* Actes of the conference: recherche d'informations assistee par ordinateur, Grenoble, 1985. Grenoble, I.M.A.G., 1985, pp. 681-698.

HENDRIX, G.G. (1979) Encoding knowledge in partitioned networks. In: *Associative Networks.* Ed. by N.V. Findler. New York, Academic Press, 1979, pp. 51-92.

HITCHINGHAM, E.E. (1979) Online interviews: charting user and searcher interaction patterns. In: *Proceedings of the 42nd Annual ASIS Meeting*, vol. 16. White Plains, N.Y., Knowledge Industry Publications, 1979, pp. 66-74.

CHRISTALER, T.H, MARBURGER .H., MURIK, K., NEBEL, B., O'LEARY, M., and WAHLSTER, W.
HOEPPNER, W. et al (1983) Beyond domain-independance: experience with the development of a German language access system to highly diverse background systems. In: *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Kahlsruhe,18-21st Aug., 1983, pp. 588-594.

HOLLNAGEL, E. and WOODS, D.D. (1983) Cognitive systems engineering: new wine in new bottles. *International Journal of Man-machine Studies*, vol. 18, 1983, pp. 583-600.

INGWERSEN, P. (1982) Search procedures in the library - analysed from the cognitive point of view. *Journal of Documentation*, vol. 38, no.3, Sept. 1982, pp. 165-191.

INGWERSEN, P. (1986) Cognitive analysis and the role of the intermediary in information retrieval. In: *Intelligent Information Systems: progress and prospects.* Ed. by R. Davies. Ellis Horwood, 1986, pp. 206-237.

INGWERSEN, P. and KAAE, S. (1980) *User-librarian negotiations and information search procedures in public libraries: analysis of verbal protocols.* Final Research Report. DB-TEK-50, 1980. Copenhagen, Royal School of Librarianship.

JERRAMS SMITH, J. (1985) SUSI: a smart user-system interface. In: *People and computers: designing the interface. Proceedings* of the Conference of the BCS Human Computer Interaction Specialist Group, University of East Anglia, 17-20th Sept., 1985. Ed. by P. Johnson and S. Cook. Cambridge, Cambridge University Press, 1985, pp. 211-220.

KELLY, G.A. (1955) *The psychology of personal constructs.* New York, Norton, 1955.

KIDD, A.L. (1985) The consultative role of an expert system. In: *People and Computers: Designing the interface. Proceedings* of the BCS HCI Specialist Group Conference, University of East Anglia, 17-20th Sept., 1985. Ed. by P. Johnson and S. Cook. Cambridge, Cambridge University Press, pp. 248-258.

KONOLIGE, K. (1983) A deductive model of belief. In: *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, 18-21st Aug, 1983, pp. 377-381.

KUIPERS, B. and KASSIRER, J.P. (1983) How to discover a knowledge representation for causal reasoning by studying an expert physician. In: *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, 18-21st Aug, 1983, pp. 49-56.

LABOV, W. and FANSHEL, D. (1977) *Therapeutic Discourse.* New York, Academic Press, 1977.

BORNING, A, McDONALD, D., TAYLOR, C., and WEYER, S.
LENAT, D.B. et al (1983) KNOESPHERE: building expert systems with encyclopedic knowledge. In: *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, 18-21st Aug, 1983, pp. 229-235.

MAIDA, A.S. (1983) Knowing intensional individuals and reasoning about knowing intensional individuals. In: *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, 18-21st Aug, 1983, pp. 382-384.

MARCUS, R.S. (1981) An automated expert system for information retrieval. In: *Proceedings of the 44th Annual ASIS meeting*, vol. 18. White Plains, N.Y., Knowledge Industry Publications, 1981, pp. 270-273.

MARCUS, R.S. (1985) Development and testing of expert systems for retrieval assistance. In: *Proceedings of the 48th Annual ASIS meeting*, vol. 22. White Plains, N.Y., Knowledge Industry Publications, 1985, pp. 289-292.

MICHALSKI, R.S., CARBONELL, J. and MITCHELL, T. (eds) (1983) *Machine learning: an artificial intelligence approach.* Palo Alto, Tioga, 1983.

MINSKY, M. (1975) A framework for representing knowledge. In: *A psychology of computer vision.* Ed. by P. Winston. New York, McGraw-Hill, 1975, pp. 211-277.

MITTMAN, B. and BORMAN, L. (1975) *Personalized data base systems.* Los Angeles, Melville Publishing Company, 1975.

NEWELL, A. and SIMON, H.A. (1972) *Human Problem Solving.* Englewood Cliffs, N.J., Prentice-Hall, 1972.

ODDY, R.N. (1977) Information retrieval through man-machine dialogue. *Journal of Documentation*, vol. 33, 1977, pp. 1-14.

HIRSCHBERG, J and WEBBER, B
POLLACK, M.E. et al (1982) User participation in the reasoning processes of expert systems. In: *Proceedings of the national conference of the American Association for Artificial Intelligence*, 1982, pp. 358-361.

POLLITT, A.S. (1981) An expert system as an online search intermediary. In: *Proceedings of the 5th International Online Information Meeting.* London, 1981, Learned Information Inc., pp. 25-32.

POLLITT, A.S. (1984) A "front-end" system: an expert system as an online search intermediary. *ASLIB Proceedings*, vol. 36, no. 5., 1984, pp. 229-234.

POLLITT, A.S. (1985) The CANSEARCH approach to end-user information retrieval. In: *Proceedings of IRFIS 6: Intelligent Information Systems for the Information Society*, Frascati, Italy, Sept., 1985. Amsterdam, North-Holland, 1986, pp. 100-108.

PRICE, L.E.T. (1983) *Functional and satisfaction analyses of information interaction dialogues.* M.Sc Thesis. Dept. of Information Science, the City University, London, 1983.

QUILLIAN, M.R. (1968) Semantic memory. In: *Semantic Information Processing.* Ed. by M. Minsky. Cambridge, Mass., MIT Press, 1968, pp. 227-270.

QUINLAN, J.R. (1979) Discovering rules by induction from collections of examples. In: *Expert systems in the microelectronic age.* Ed. by D. Michie. Edinburgh, Edinburgh University Press, 1979, pp. 33-46.

RICH, E. (1979a) User modelling via stereotypes. *Cognitive Science*, vol. 3, 1979, pp. 329-354.

RICH, E. (1979b) *Building and exploiting user models.* Ph.D Thesis, Technical Report CMU-CS-79-119, Carnegie-Mellon University, 1979.

RICH, E. (1983) Users are individuals. *International Journal of Man-machine studies*, vol. 18, no. 3, 1983, pp. 199-214.

ROBERTS, R.B. and GOLDSTEIN, I.P. (1977) *The FRL Reference Manual.* MIT AI Laboratory Memo 409, 1977.

ROSENBERG, S.T. (1980) Frame-based text processing. In: *Frame Conceptions and Text Understanding.* Ed. by D. Metzing. New York, Walter de Gruyter, 1980.

SCHANK, R. and ABELSON, R. (1977) *Scripts, plans, goals and understanding.* Hillsdale, NJ, Laurence Erlbaum Ass., 1977.

SHAW, M.L.G. (1984) Knowledge engineering for expert systems. In: *Interact '84. Proceedings* of the IFIP Task Group on Human-computer interaction. Imperial College, London, 4-7th Sept. 1984, pp. 328-332.

SHOVAL, P. (1985) Principles, procedures and rules in an expert system for information retrieval. *Information Processing and Management*, vol. 21, no. 6, 1985, pp. 475-487.

SIDNER, C. (1979) Developing support systems for information analysis. In: *Artificial Intelligent: an MIT perspective*, vol. 1. Ed. by P.H. Winston and R.H. Brown. The MIT Press, 1979, pp. 231-252.

SIDNER, C.L. and ISRAEL, D.J. (1981) Recognizing intended meanings and speakers' plans. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, 1981, pp. 203-208.

SINCLAIR, J. McH. and COULTHARD, R.M. (1975) *Towards an analysis of discourse. The English used by teachers and pupils.* Oxford, OUP, 1975.

SLEEMAN, D. (1985) UMFE: a user modelling front end subsystem. *International Journal of Man-machine studies*, vol. 23, 1985, pp. 71-88.

SMITH, L.C. (1979) *Selected artificial intelligence techniques in information retrieval systems research.* Ph.D Thesis. School of Information Studies, Syracuse University, N.Y., 1979.

SPARCK JONES, K. (1984) *User models and expert systems.* University of Cambridge Computer Laboratory, Technical Report No. 61, 1984.

SPARCK JONES, K. (1985) Natural language interfaces for expert systems: an introductory note. In: *R & D in expert systems. Proceedings* of the 4th Technical Conference of the BCS Specialist Group on Expert Systems, University of Warwick, 18-20th December, 1984. Ed. by M.A. Bramer. Cambridge University Press, 1985, pp. 85-94.

SPARCK JONES, K. (1987) *Architecture problems in the construction of expert systems for document retrieval.* Paper presented at Distributed Expert-Based Information Systems Workshop, Rutgers University, N.J., March 1987.

STANSFIELD, J.L. (1979) Developing support systems for information analysis. In: *Artificial Intelligence: an MIT Perspective*, vol. 1. Ed. by P.H. Winston and R.H. Brown. The MIT Press, 1979, pp. 289-314.

STEELS, L. and CAMPBELL, J.A. (eds) (1985) *Progress in Artificial Intelligence.* Ellis Horwood, 1985.

STINTON, C. (1984) *Intermediary knowledge resources in information interaction.* M.Sc Thesis. Dept. of Information Science, the City University, London, 1984.

BROOKS, H.M., ROBINSON, B.A., and VICKERY, B.C.
VICKERY, A. et al. (1986) *An expert system for referral.* Final report to BLR&D on project SI/G/625. February, 1986. London, Central Information Service, University of London, 1986.

WELBANK, M. (1983) A review of knowledge acquisition techniques for expert systems. *British Telecom Research Laboratories Report.* Martlesham Heath, Martlesham Consultancy Services, December, 1983.

WILENSKY, R., ARENS, Y. and CHIN, D. (1984) Talking to UNIX in English: an overview of UC. *The AI Magazine*, vol. 5, 1984, pp. 29-39.

WOOLF, B. and McDONALD, D.D. (1983) *Human-computer discourse in the design of a PASCAL tutor.* COINS Technical Report, University of Massachusetts, 27th November, 1983.

YIP, M.K. (1981) *An expert system for document retrieval.* M.S. Thesis. MIT, Cambridge, Mass. February, 1981.

APPENDICES

Appendix A

Transcript of Interview 120684hba

I we should be ok now (......) emm (.) what's the (.) problem? /(1)
U

I   mm hm /(3)
U ok I'm just beginning (.) a research project /(2) err (.) I'm a
----

I   yeah /(5)
U research student at LSE /(4a) (...) and um (...) working in the
----

I
U Geography department and I'm err (.) /(4b) doing a (.) thesis (.)

I
U beginning a thesis on err (.) forestry (.) and err the impact of
----

I   interesting /(6)   oh that's
U recreation (.) of conflicts of recreation and forestry /(4c)

I interesting (.) mm (5 secs) (inaud........) /(7)
U   and err (...) one of the

I
U things I'm aiming to do (.) eventually is to look at the cost
----

I
U benefits of different err (.) management schemes (.) for recreation
----

I   mm (8 secs) have you found very much (.) so far? /(9)
U (...) /(8)   I've um

I
U (.) really not looked in the area so far /(10a) cos I've- I've been
----

I
U looking more in err (.) just the err (.) theoretical aspects of cost
----

I   right (.) mm hm /(11)
U benefit (.) /(10b)   and conflict (.) those sorts of things /(12)

I so th- the angle that you're going to take on it is it is it more
U

I the (.) say the cost benefit economics (.) /(13)   or (.) is it

```
U                                                mm hm /(14)
----------------------------------------------------------------------
----------------------------------------------------------------------
I looking (.) looking at the area as a whole what sort of recreation
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I goes on (.) err (.) /(15) how much land is given to forestry (.) and
                 err no I think /(16)
----------------------------------------------------------------------
----------------------------------------------------------------------
I so on /(17)
U       no (.) what I'm interested in- /(18a) err I'm doing this (,)
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                    ah (.) yeah (.) mm /(19)
U in Nova Scotia (.) when I finish (.) /(18b) um (.) and err (.) I
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U spent four years with the Government as the co-ordinator of outdoor
----------------------------------------------------------------------
----------------------------------------------------------------------
I                    mm /(21)
U recreation so I (.) /(20) have a pretty good handle on what kinds of
----------------------------------------------------------------------
----------------------------------------------------------------------
I             yes (.) /(23) I see (.) yes /(25)
U recreation (.) /(22) at this stage (.) /(24a) but what I'm looking
----------------------------------------------------------------------
----------------------------------------------------------------------
I             mm /(26)
U at now is (.) /(24b) err (.) the conflict and the cost benefit (.)
----------------------------------------------------------------------
----------------------------------------------------------------------
I                             fine (.) mm /(28)
U aspects of it (.) although they I s- (.) /(27) I will have to look
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                                yes /(30)
U at um (...) some aspects of um (.) recreation (.) /(29) but I think
----------------------------------------------------------------------
----------------------------------------------------------------------
I          but you- you- you've got an idea of what sort of things go
U I can /(31)
----------------------------------------------------------------------
----------------------------------------------------------------------
I on (laughter) in the forest as it were (.)  /(32) yeah (.) um (.)
U                               yeah (.) /(33) yeah /(34)
----------------------------------------------------------------------
----------------------------------------------------------------------
I now (.) are you going to (.) look at this on a sort of world- (.)
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I world wide level I mean obviously it would only go on in certain
U
----------------------------------------------------------------------
```

```
----------------------------------------------------------------------
I countries but (.) /(35) are you concentrating on (.) just Canadian
U               um /(36)
----------------------------------------------------------------------
----------------------------------------------------------------------
I or- or are you looking at the whole area (.) wherever it (..) /(37)
U                                                    well (...) /(38)
----------------------------------------------------------------------
----------------------------------------------------------------------
I recreation is going on? /(39)                    mm /(41)
U                       in terms of methodology (.) /(40) I- I have no
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                       mm (.) yeah /(43)   yeah /(45)
U particular restrictions geographically /(42)     ok? /(44) but
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U obviously I'm (.) I want to look at um (...) err Nova Scotia in
----------------------------------------------------------------------
----------------------------------------------------------------------
I                              mm /(47)
U particular in fact err /(46) (.) the more North American (..) err
----------------------------------------------------------------------
----------------------------------------------------------------------
I                          yes (.) is that because it- it err it sort
U literature I can look at (..) /(48) (inaud..) /(50)
----------------------------------------------------------------------
----------------------------------------------------------------------
I of began there? (.) or (.) /(49) is it just that- that you're
U                     umm /(51)                   I th- /(53)
----------------------------------------------------------------------
----------------------------------------------------------------------
I concentrating your (.) research /(52)           mm /(56)
U                I th- /(54)       I think the um (5 secs) wh- (.) I'm
----------------------------------------------------------------------
----------------------------------------------------------------------
I
U not sure I understand the question in terms of what began where/(55)
----------------------------------------------------------------------
----------------------------------------------------------------------
I no sorry i- is it because um: say- say Nova Scotia's had more e- a
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I longer experience of (.) providing recreational facilities? /(57)
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I              (inaud.) /(59)              yeah (laughter) you've
U no /(58a) its just because I live there /(58b)
----------------------------------------------------------------------
----------------------------------------------------------------------
I done y- you've done the err um the practical experience (.) /(60)
U                                                         yes (.)
----------------------------------------------------------------------
----------------------------------------------------------------------
I right (inaud..) um:: so (.) really we- (.) we've (.) got
```

```
U yeah /(61)
----------------------------------------------------------------------
----------------------------------------------------------------------
I possibilities of looking at this from the economics side (..) /(62)
U                                                         mm hm /(63)
----------------------------------------------------------------------
----------------------------------------------------------------------
I but at the same time (.) um: if we can get some information on- on:
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I (.) America /(64a) I would imagine that ther wouldn't be (.) a
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I fantastic amount of information on (.) um (.) recreation in- and
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I forestry linked anyway (,)/(64b) and if there were (.) hopefully they
U                                                   mm /(65)
----------------------------------------------------------------------
----------------------------------------------------------------------
I would mention (.) something on the economics side as well /(64c)
U                                                           mm (.)
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                                        mm /(67)
U well th- there's a growing literature the last ten years I would
----------------------------------------------------------------------
----------------------------------------------------------------------
I                                                yes /(68)       yeah
(.) U think in terms of (.) recreation and forestry (.) /(66) and err
----------------------------------------------------------------------
----------------------------------------------------------------------
I cos i- there's one (.) um: (.) know if you start to think of- (.)
U (...) /(69) (inaud.) /(71)
----------------------------------------------------------------------
----------------------------------------------------------------------
I sources really of information there's the (.) Commonwealth Agricult-
U
----------------------------------------------------------------------
----------------------------------------------------------------------
I ural Bureaux (.) /(70) and- are you familar with their (.) public-
U                           mm hm /(72)
----------------------------------------------------------------------
----------------------------------------------------------------------
I ations? (.) /(73) they've got an abstracting journal that looks at
U          no (.) /(74)  no /(76)
----------------------------------------------------------------------
----------------------------------------------------------------------
I leisure and recreation (.) /(75) and (.) um: I can remember from (.)
U                         mm hm /(77)
----------------------------------------------------------------------
----------------------------------------------------------------------
I previous searches we have had things like for example (.) footpaths
U
----------------------------------------------------------------------
```

```
------------------------------------------------------------------------
I err sort of wear and tear (laugh) (.) /(78) that having loads of
U                                 mm hm /(79)
------------------------------------------------------------------------
------------------------------------------------------------------------
I people trot- trotting round on these (.) /(80) public available
U                                        yes /(81)
------------------------------------------------------------------------
------------------------------------------------------------------------
I footpaths on (.) um: special sites (.) th- that sort of thing (,)
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I /(82) so in other words I think you might get a bit on- on the
U mm hm /(83)
------------------------------------------------------------------------
------------------------------------------------------------------------
I conflict there (..) /(84) err that's one source /(86)
U                right (.) /(85) my interest is not so much (.) on
------------------------------------------------------------------------
------------------------------------------------------------------------
I                 no (.) /(88) its the overall thing (.) /(90)
U specific sites (.) /(87) its um on the /(89) overall management
------------------------------------------------------------------------
------------------------------------------------------------------------
I yeah (.) /(92) therefore we ought to look at (.) um: (.) economic
U of forests /(91)
------------------------------------------------------------------------
------------------------------------------------------------------------
I sides as well /(93)          um (.) there's- there's Social Sciences
U               mm hm (.) yeah /(94)
------------------------------------------------------------------------
------------------------------------------------------------------------
I Citation Index (........) /(95a) it soun- sounds as though its- err
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I its (,) very much Social Sciences but its- its a very broad based
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I um: (.) indexing source (.) /(95b) and I think that would be worth
U                          mm hm /(96)
------------------------------------------------------------------------
------------------------------------------------------------------------
I trying (.) um: (.) /(97) what we've got to do there is concentrate
U                  ok /(98)
------------------------------------------------------------------------
------------------------------------------------------------------------
I on just on title words (.) so its not so easy (,) you know so if
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I you say forestry and they talk about (.) a larch plantation or
U
------------------------------------------------------------------------
------------------------------------------------------------------------
I something you wouldn't actually get it (.) /(99) but (.) um I think
```

U mm /(100)

I if we look at CAB (.) and we look at the Social Sciences (,) and I
U

I think its worth looking at some of the economics ones as well /(101)
U

I um: (...) let's just have a look through (..) the Dialog list
U right /(102)

I /(103) (44 secs) (inaud..) /(104)
U

Appendix B

Subfunction Analysis of Interview 120684hba

Focus	Utterance	Speaker	Subgoal
1 [USER]	1	I	RES
	2	U	PDIM
	3	I	ph
	4a	U	USER
	4b	U	USER
	4c	U	UGOAL..PDIM..RES
	5	I	ph
	6	I	ph
	7	I	ph
	8	U	RES
2 [PREV]	9	I	PREV
	10a	U	PREV
	10b	U	PDIM..RES
	11	I	ph
	12	U	RES
3 [RES]	13	I	RES
	14	U	ph
	15	I	RES
	16	U	RES
	17	I	RES
	18a	U	RES
	18b	U	UGOAL
	19	I	ph
	20	U	BACK
	21	I	ph
	22	U	KNOW
	23	I	ph
	24a	U	KNOW
	24b	U	RES
	25	I	ph
	26	I	ph
	27	U	RES
	28	I	ph
	29	U	RES
	30	I	ph
	31	U	failed
	32	I	MATCH:KNOW
	33	U	MATCH:KNOW:CONFIRM
	34	U	MATCH:KNOW:CONFIRM
4 [RES]	35	I	RES
	36	U	ph
	37	I	RES
	38	U	ph
	39	I	RES
	40	U	RES
	41	I	ph
	42	U	RES

	43	I	ph
	44	U	MATCH:RES:CONFIRM
	45	I	MATCH:RES:CONFIRM
	46	U	RES
	47	I	ph
	48	U	UGOAL/DOCS
	49	I	SUBJ
	50	U	ph
	51	U	ph
	52	I	RES
	53	U	failed
	54	U	failed
	55	U	MATCH:RES:QUERY
	56	I	ph
	57	I	SUBJ..RES
	58a	U	MATCH:RES:DISCONFIRM
	58b	U	BACK
	59	I	ph
	60	I	MATCH:PREVNON:QUERY
	61	U	MATCH:PREVNON:CONFIRM
5 [TOPIC]	62	I	TOPIC
	63	U	TOPIC
	64a	I	TOPIC
	64b	I	SLIT
	64c	I	DOCS
	65	U	ph
	66	U	SLIT
	67	I	ph
	68	I	ph
	69	U	failed
6 [DB]	70	I	DB
	71	U	failed
	72	U	ph
	73	I	IRS
	74	U	IRS
	75	I	EXPLAIN:DB
	76	U	IRS
	77	U	ph
	78	I	EXPLAIN:DB
	79	U	ph
	80	I	EXPLAIN:DB
	81	U	ph
	82	I	EXPLAIN:DB
	83	U	ph
	84	I	EXPLAIN:DB
	85	U	MATCH:DB:CONFIRM
	86	I	DB
7 [RES]	87	U	RES
	88	I	RES
	89	U	RES

	90	I	RES
	91	U	RES
	92	I	ph
	93	I	TOPIC
	94	U	MATCH:TOPIC:CONFIRM
8	95a	I	DB
[DB]	95b	I	EXPLAIN:DB
	96	U	ph
	97	I	DB
	98	U	MATCH:DB:CONFIRM
9	99	I	EXPLAIN:STRAT
[STRAT]	100	U	ph
	101	I	STRAT:DB
	102	U	MATCH:STRAT
	103	I	DISPLAY

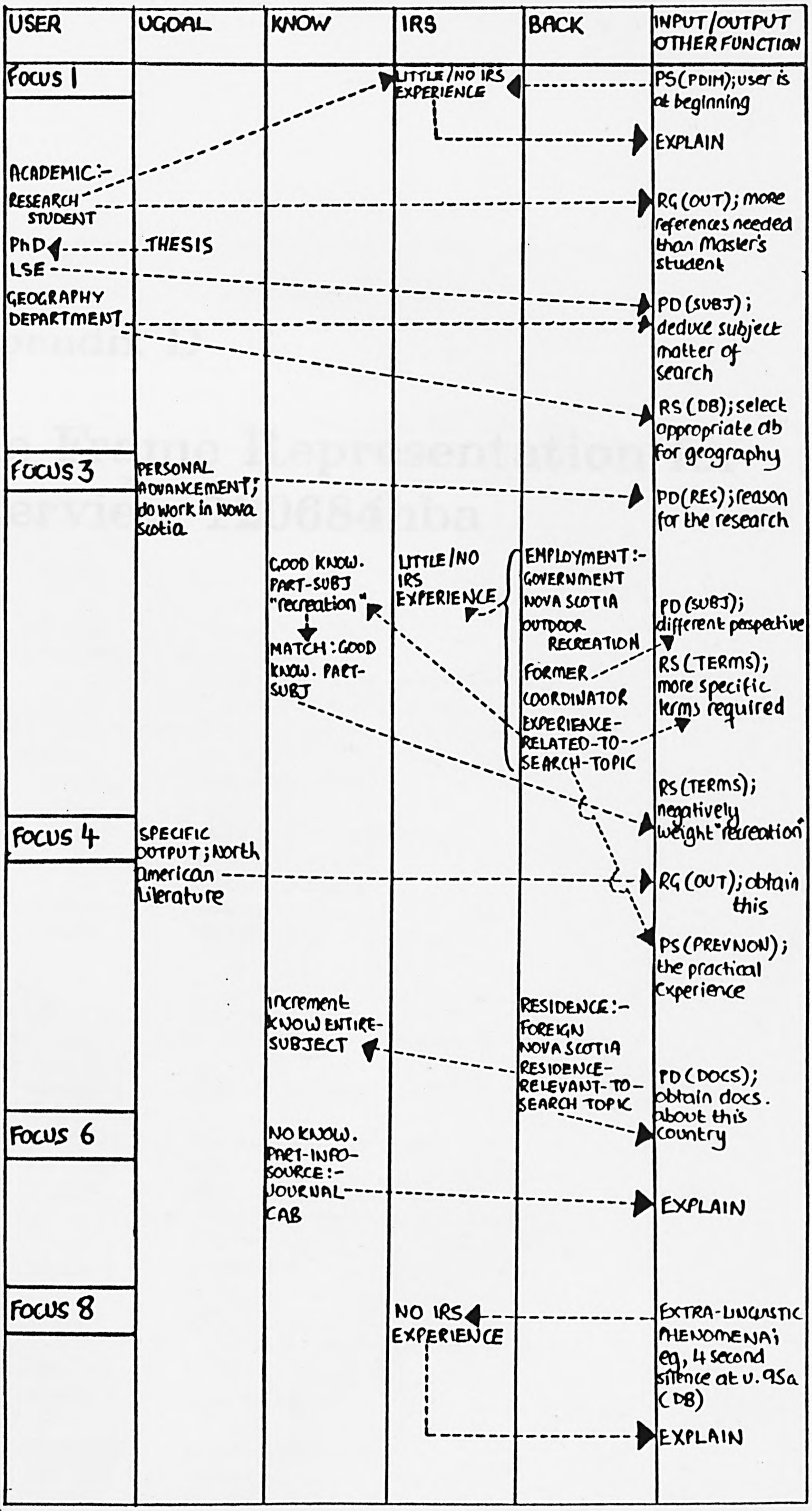
The following instances of disagreement were found in the independent analyses by the two team members:

FOCUS	UTTERANCE	SPEAKER	SUBFUNCTION
4	58b	U	BACK(PJD) USER(HMB)
6	73	I	KNOW(PJD) IRS(HMB)
6	74	U	KNOW(PJD IRS(HMB)
9	101	I	PLAN(PJD) STRAT(HMB)

The agreed final categorisation of these utterances can be seen above.

Appendix C

Interaction of the User Model with the other Functions for Interview 120684hba


USER
UGOAL
KNOW
IRS
BACK
INPUT/OUTPUT OTHER FUNCTION
FOCUS 1
LITTLE/NO IRS EXPERIENCE
PS (PDIM); user is at beginning
EXPLAIN
ACADEMIC:-
RESEARCH STUDENT
RG (OUT); more references needed than Master's student
PhD
THESIS
LSE
GEOGRAPHY DEPARTMENT
PD (SUBJ); deduce subject matter of search
RS (DB); select appropriate db for geography
FOCUS 3
PERSONAL ADVANCEMENT; do work in Nova Scotia
PD (RES); reason for the research
GOOD KNOW. PART-SUBJ "recreation"
LITTLE/NO IRS EXPERIENCE
EMPLOYMENT:- GOVERNMENT NOVA SCOTIA OUTDOOR RECREATION
PD (SUBJ); different perspective
MATCH: GOOD KNOW. PART-SUBJ
FORMER COORDINATOR
RS (TERMS); more specific terms required
EXPERIENCE-RELATED-TO-SEARCH-TOPIC
RS (TERMS); negatively weight "recreation"
FOCUS 4
SPECIFIC OUTPUT; North American Literature
RG (OUT); obtain this
PS (PREVNON); the practical experience
increment KNOW ENTIRE-SUBJECT
RESIDENCE:- FOREIGN NOVA SCOTIA RESIDENCE-RELEVANT-TO-SEARCH TOPIC
PD (DOCS); obtain docs. about this country
FOCUS 6
NO KNOW. PART-INFO-SOURCE:- JOURNAL CAB
EXPLAIN
FOCUS 8
NO IRS EXPERIENCE
EXTRA-LINGUISTIC PHENOMENA; eg, 4 second silence at u. 95a (DB)
EXPLAIN

Appendix D

The Frame Representation for Interview 120684hba

```
FOCUS 1
------------------------------------------------------------
[IRS
AKO      $value       user-model-slot

type     $require     none some
         $value       NONE
         $if-needed   IRS-demon
                      (IF PDim = Beginning THEN None AND
                        IF USER-STATUS Student THEN None
                        AND IF BACK Acad Former THEN
                        None)]
------------------------------------------------------------
[NONE
AKO      $value       IRS

type     $value       terminal
         $if-added    None-demon
                      (TRIGGER EXPLAIN; comprehensive
                         explanation required)]
------------------------------------------------------------
[USER-STATUS
AKO      $value        user-model-slot

type     $require      current-acad former-acad
                       prospective-acad
                       current-non-acad
         $default      current-acad
         $value        CURRENT-ACAD

position
 -slot   $require      position]
------------------------------------------------------------
[POSITION
AKO      $value        user-status-slot

type     $require      end intermediary team
         $default      end
         $value        END]
------------------------------------------------------------
[END
AKO      $value        position

type     $value        terminal]
------------------------------------------------------------
[CURRENT-ACAD
AKO      $value        user-status

type     $require      staff student
         $default      student
         $value        STUDENT
```

```
institution-
   slot   $require        institution

funding-source-
  slot    $require        funding-source

time-slot $require        time]
--------------------------------------------------------------
[STUDENT
AKO       $require        current-acad

type      $require        undergrad postgrad
          $default        postgrad
          $value          POSTGRAD
          $if-added       student-demon
                          (TRIGGER PS(Pdim);find out what
                           stage user is at AND
                           TRIGGER KNOW Subj Evalue; less
                           than Staff AND
                           TRIGGER IRS Experience = None)]
--------------------------------------------------------------
[POSTGRAD
AKO       $value           student

type      $require         taught research
          $default         research
          $value           RESEARCH}
          $if-needed       postgrad-demon
                           (IF UGOAL Specific-Intentions =
                             Thesis THEN Research)]
--------------------------------------------------------------
[RESEARCH
AKO       $value          postgrad

type      $require        MPhil PhD
          $default        PhD
          $value          PhD
          $if-added       research-demon
                          (TRIGGER RG(Out); more required
                           than Taught AND
                           TRIGGER UGOAL Specific
                           Intentions Thesis)]
--------------------------------------------------------------
[PhD
AKO       $value           research

type      $value           terminal)]
--------------------------------------------------------------
[INSTITUTION
AKO       $value           current-academic-slot

type      $require         London-uni UK-uni foreign-uni
          $default         London-uni
```

```
          $if-needed     instit-demon
                          (IF BACK Residence THEN
                            foreign-uni)
          $value          LONDON-UNI

department-
   slot   $require       department]
------------------------------------------------------------
[LONDON-UNI
AKO       $value         institution

type      $require       a-London-college
          $value         LSE]
------------------------------------------------------------
[LSE
AKO       $value         London-uni

type      $value         terminal
          $if-added      a-London-college-demon
                          (TRIGGER PD(Subj); deduce)]
------------------------------------------------------------
[DEPARTMENT
AKO       $value         institution-slot

type      $require       any-valid-subject
          $value         GEOGRAPHY
          $if-needed     dept-demon
                          (TRIGGER PD(Subj); deduce
          $if-added      dept-add-demon
                          (TRIGGER PD(Subj); deduce
                           subject matter of search
                          AND TRIGGER RS(db); select
                           appropriate db)]
------------------------------------------------------------
[GEOGRAPHY
AKO       $value         department

type      $value         terminal]
------------------------------------------------------------
[UGOAL
AKO       $value         user-model-slot

type      $require       current-search-goal
                         goals-leading-to-search
                         specific-intentions
                         general-goals
          $value         SPECIFIC-INTENTIONS]
------------------------------------------------------------
[SPECIFIC-INTENTIONS
AKO       $value         ugoal

type      $prefer        review-article thesis
                         dissertation report paper
```

```
                      teaching-materials
                      advise-others
                      fulfill-task-for-other
          $value      THESIS
          $if-needed  specific-intentions-demon
                       (IF USER-STATUS Postgrad
                        THEN Thesis OR Dissertation)]
---------------------------------------------------------------
[THESIS
AKO       $value      specific-intentions

type      $value      terminal
          $if-added   thesis-demon
                       (TRIGGER USER-STATUS
                         Postgrad = Research)]
---------------------------------------------------------------
FOCUS 3
---------------------------------------------------------------
[UGOAL
AKO       $value      user-model-slot

type      $require    current-search-goal
                      goals-leading-to-search
                      specific-intentions
                      general-goals
          $value      GENERAL-GOALS]
---------------------------------------------------------------
[GENERAL-GOALS
AKO       $value      ugoal

type      $prefer     personal-advancement
                      build-up-knowledge-in-own
                                          -country
                      idealistic-goals
          $value      PERSONAL-ADVANCEMENT
          $if-needed  general-goal-demon
                       (IF BACK Employment =
                                     Prospective
                        THEN Personal-advancement AND
                        IF BACK Residence
                        THEN build-up-knowledge-in
                                     -own-country)]
---------------------------------------------------------------
[PERSONAL-ADVANCEMENT
AKO       $value      general-goals

type      $value      terminal)]
---------------------------------------------------------------
[BACK
AKO       $value      user-model-slot

type      $require    employer residence academic-back
                      temporary-features
```

```
        $value          EMPLOYER]
-------------------------------------------------------------
[EMPLOYER
AKO     $value          back

type    $require        any-employer
        $value          GOVERNMENT

time-dimension-slot
        $require        time-dimension

location-slot
        $require        any-location
        $value          NOVA-SCOTIA

level-of-responsibility-slot
        $require        level-of-responsibility

employ-nature-slot
        $require        any-employ-nature
        $value          OUTDOOR-RECREATION

related-to-search-topic-slot
        $require        related-to-search-topic]
-------------------------------------------------------------
[GOVERNMENT
AKO     $value          employer

type    $value          terminal]
-------------------------------------------------------------
[TIME-DIMENSION
AKO     $value          employer-slot

type    $require        current former prospective
        $value          FORMER
        $if-needed      time-dimension-demon
                         (IF USER-STATUS Part
                           THEN Current)]
-------------------------------------------------------------
[FORMER
AKO     $value          time-dimension

type    $value          terminal
        $if-added       former-demon
                         (TRIGGER PD(Subj); may need
                           different perspective on
                           subject matter)]
-------------------------------------------------------------
[NOVA-SCOTIA
AKO     $value          employer-slot

type    $value          terminal]
-------------------------------------------------------------
```

```
[LEVEL-OF-RESPONSIBILITY
AKO        $value          employer-slot

type       $require        any-level
           $value          CO-ORDINATOR]
--------------------------------------------------------------
[CO-ORDINATOR
AKO        $value          level-of-responsibility

type       $value          terminal
           $default        consultant
           $if-added       any-level-demon
                            (IF Senior AND
                             Related-to-search-topic = Yes
                             THEN increment KNOW
                                       Entire-subj)]
--------------------------------------------------------------
[OUTDOOR-RECREATION
AKO        $value          employer-slot

type       $value          terminal]
--------------------------------------------------------------
RELATED-TO-SEARCH-TOPIC
AKO        $value          employer-slot
                           academic-back-slot
                           residence-slot

type       $require        yes no
           $value          YES
           $default        yes
           $if-needed      experience-related-demon
                            (IF USER-STATUS Part
                             THEN yes AND
                             IF Time-dimension = Current
                             THEN yes)]
--------------------------------------------------------------
[YES
AKO        $value          Related-to-search-topic

type       $value          terminal
           $if-added       yes-demon
                            (TRIGGER KNOW Entire-subj
                             Evaluate; Increment AND
                             TRIGGER RS (Terms); use more
                             specific terms)]
--------------------------------------------------------------
[KNOW
AKO        $value          user-model-slot

type       $require        part-subj entire-subj
                           all-info-sources
                           part-info-source
           $value          PART-SUBJ
```

```
evaluate-slot
          $require        evaluate)]
------------------------------------------------------------
[EVALUATE
AKO       $value          know-slot

type      $require        none poor average good complete
                          sufficient
          $value          GOOD]
------------------------------------------------------------
[GOOD
AKO       $value          evaluate

type      $value          terminal)]
------------------------------------------------------------
[PART-SUBJ
AKO       $value          know

type      $value          terminal
          $if-added       part-subj-demon
                           (IF Evaluate = Good OR Complete
                            THEN TRIGGER RS(Terms);
                            do not search on those terms)]
------------------------------------------------------------
FOCUS 4
------------------------------------------------------------
[UGOAL
AKO       $value          user-model-slot

type      $require        current-search-goal
                          goals-leading-to-search
                          specific-intentions
                          general-goals
          $value          CURRENT-SEARCH-GOAL]
------------------------------------------------------------
[CURRENT-SEARCH-GOAL
AKO       $value          ugoal

type      $prefer         specific-output
                          preliminary-search
                          exhaustivity
          $value          SPECIFIC-OUTPUT]
------------------------------------------------------------
[SPECIFIC-OUTPUT
AKO       $value          current-search-goal

type      $value          terminal
          $if-added       specific-output-demon
                           (TRIGGER RG(Out)]
------------------------------------------------------------
[BACK
AKO       $value          user-model-slot
```

```
type       $require        employer residence academic-back
                           temporary-features
           $value          RESIDENCE]
------------------------------------------------------------
[RESIDENCE
AKO        $value          back

type       $require        foreign-country
                           geographical-region town
           $value          FOREIGN-COUNTRY
           $if-added       residence-demon
                            (TRIGGER USER-STATUS
                             Institution = Foreign-uni)
           $if-needed      residence-need-demon
                            (TRIGGER USER-STATUS
                             Institution; Access and
                                               deduce)
related-to-search-topic-slot
           $require        related-to-search-topic]
------------------------------------------------------------
[FOREIGN-COUNTRY
AKO        $value          residence

type       $require        any-foreign-country
           $value          NOVA-SCOTIA]
------------------------------------------------------------
[NOVA-SCOTIA
AKO        $value          foreign-country

type       $value          terminal]
------------------------------------------------------------
[RELATED-TO-SEARCH-TOPIC
AKO        $value          residence-slot employer-slot
                           academic-back-slot

type       $require        yes no
           $default        yes
           $value          YES]
------------------------------------------------------------
[YES
AKO        $value          related-to-search-topic

type       $value          terminal
           $if-added       yes-demon
                            (TRIGGER KNOW Entire-subj
                             Evaluate; Increment AND
                             TRIGGER PD(Docs); retrieve
                             docs concerning this country)]
------------------------------------------------------------
FOCUS 6
------------------------------------------------------------
[KNOW
```

```
AKO       $value        user-model-slot

type      $require      part-subj entire-subj
                        all-info-sources
                        part-info-source
          $value        PART-INFO-SOURCE

evaluate-slot
          $require      evaluate)]
------------------------------------------------------------
[PART-INFO-SOURCE
AKO       $value        know

type      $require      books journals newspapers
                        magazines
          $value        JOURNALS]
------------------------------------------------------------
[JOURNALS
AKO $value part-info-source

type      $require      all-journals a-named-journal
          $value        A-NAMED-JOURNAL]
------------------------------------------------------------
[A-NAMED-JOURNAL
AKO       $value        journals

type      $require      any-journal
          $value        CAB]
------------------------------------------------------------
[CAB
AKO       $value        a-named-journal

type      $value        terminal
          $if-added     journal-demon
                         (IF Evaluate = Good OR Complete
                          THENTRIGGER RG(Out); do not
                          retrieve refs from this
                          journal)

evaluate  $value        NONE]
------------------------------------------------------------
------------------------------------------------------------
```

Appendix E

The Questionnaire for the Interviews with the Intermediaries

Section 1 : Questions relating to the USER subfunction

(a) Why is it useful to know which **type** of degree the user is doing?
(b) If you know the type of degree, does this enable you to form hypotheses about, for example, the OUTPUT required, or the user's STATE OF KNOWLEDGE?
(c) Is it useful to know whether the degree is **full** or **part** time?
(d) Does it help to know which **department** the user is in?
(e) Is it necessary to know that the user is an **intermediary** rather than an **end** user?
(f) Is it useful to know whether a user is a **Fellow** or holds a **Research Fellowship**?
(g) Why do you ask about the user's **source of funding**?
(h) How do **non-academic** research users differ from **academic** research users?
(i) Is it helpful to know from which **institution** the user comes?

Section 2 : Questions relating to the UGOAL subfunction

(a) In general terms, how do you tend to use information about the user's **goals** and **objectives**?
(b) Why is it important to know whether someone is **starting a project** on something, as opposed to **"just literature collection"**?
(c) Why is it useful to know that someone is doing a **literature review**?
(d) Similarly, is it important to know whether someone is doing a **case study** or a **state of the art** paper or a **longitudinal study**?
(e) How useful is it to know about a user's **"idealistic"** goals?
(f) Is it useful to know if someone is trying to fulfill **someone else's goal**, ie, the user is carrying out something on behalf of someone else?
(g) How do you use the information that someone is just wanting to do a **"preliminary search"**, "to see what is around"?
(h) A foreign student is having a search carried out on "education administration" and wishes to be equipped to "advise policy makers", and to "prepare lectures for students" at the college of higher education that he runs in his own country. Do you find it useful to know about these types of goals?

Section 3 : Questions relating to the KNOW subfunction

(a) When a user apparently has **very little knowledge** about her search topic and/or her subject area, what does this imply to you? Do you have to alter your tactics?
(b) Do you distinguish between knowledge on **topic** and knowledge on **subject**?
(c) Do you form hypotheses about a user's knowledge from information that you have received about her **background**?
(d) If the user seems to have little/no idea in which **information source** the relevant documents may be found, what do you assume about the user from this information?
(e) A general question - how do you tend to **evaluate users' knowledge**, and how do you use this "knowledge of knowledge"?

Section 4 : Questions relating to the IRS subfunction

(a) In the absence of any explicit information, how do you form hypotheses about a user's experience of information retrieval systems?
(b) In one case, once the intermediary discovered that a user had done a **similar search** back at his own university in Canada, she tried to find out which systems and which files were searched. Is this usually done so that they can be avoided or updated? Are there any other reasons for finding out about a user's previous online searches?
(c) In another case, the intermediary asked whether the user was familiar with CAB publications. If someone is not familiar with the **relevant abstracting journals** do you

tend to assume that therefore they have no IRS experience? Do you make any other assumptions?
(d) In general, what are the differences between someone who has had **some** IRS experience, and someone who has had a **lot** of IRS experience?

Section 5 : Questions relating to the BACK subfunction

(a) A general question - **which features** of a user's background (eg, academic, employment, residence, temporary features) is it useful to find out about?
(b) Is it useful to know about someone's **previous employment**?
(c) Is it useful to know about someone's **current employment**?
(d) What do you assume when someone's employment background seems to be **different/unrelated** etc. to the **search** topic? Do you form a hypothesis about the user's knowledge of the topic? Or his/her goals?
(e) Do you ever try to form hypotheses about a user's **goals** using information about her/his **employment** background?
(f) What do you make of information about a user's **place/country of residence**?
(g) Do you find it useful to know about a user's **academic background**?
(h) If a user's academic background is **unrelated/outside the subject area of** their present search topic, do you form any particular hypotheses regarding their states of knowledge, their output requirements, and so on?
(i) If a user's academic training took place a **long time ago**, do you form any hypotheses about their states of knowledge, their IRS experience, etc.?
(j) Is it important to find out about what I have termed **"relevant temporary features"** of a user's background, eg, budget available, when they must depart from the UK, access to libraries, translation services, photocopying facilities, etc. ?

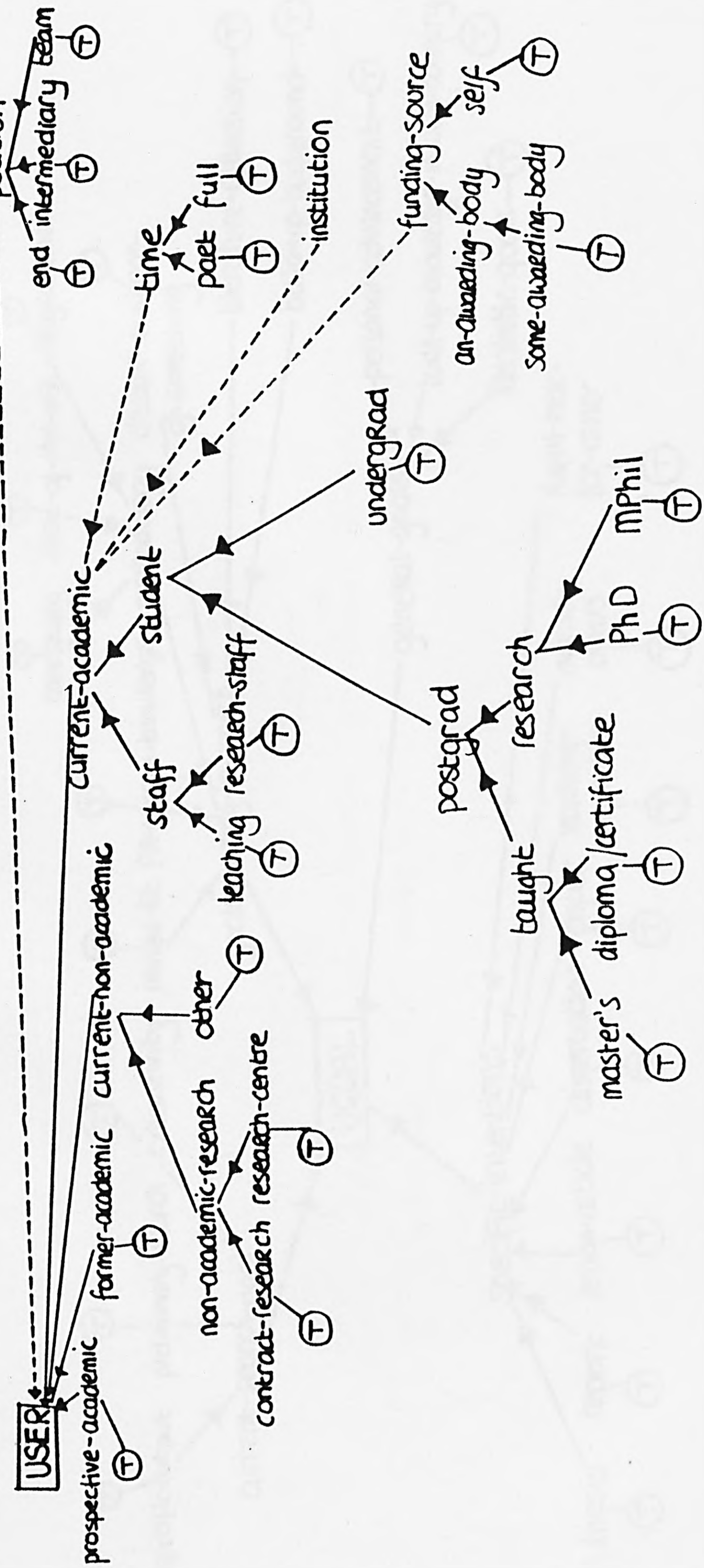
FINALLY, do you have any additional comments on users? Do these five subcategories seem exhaustive and adequate?

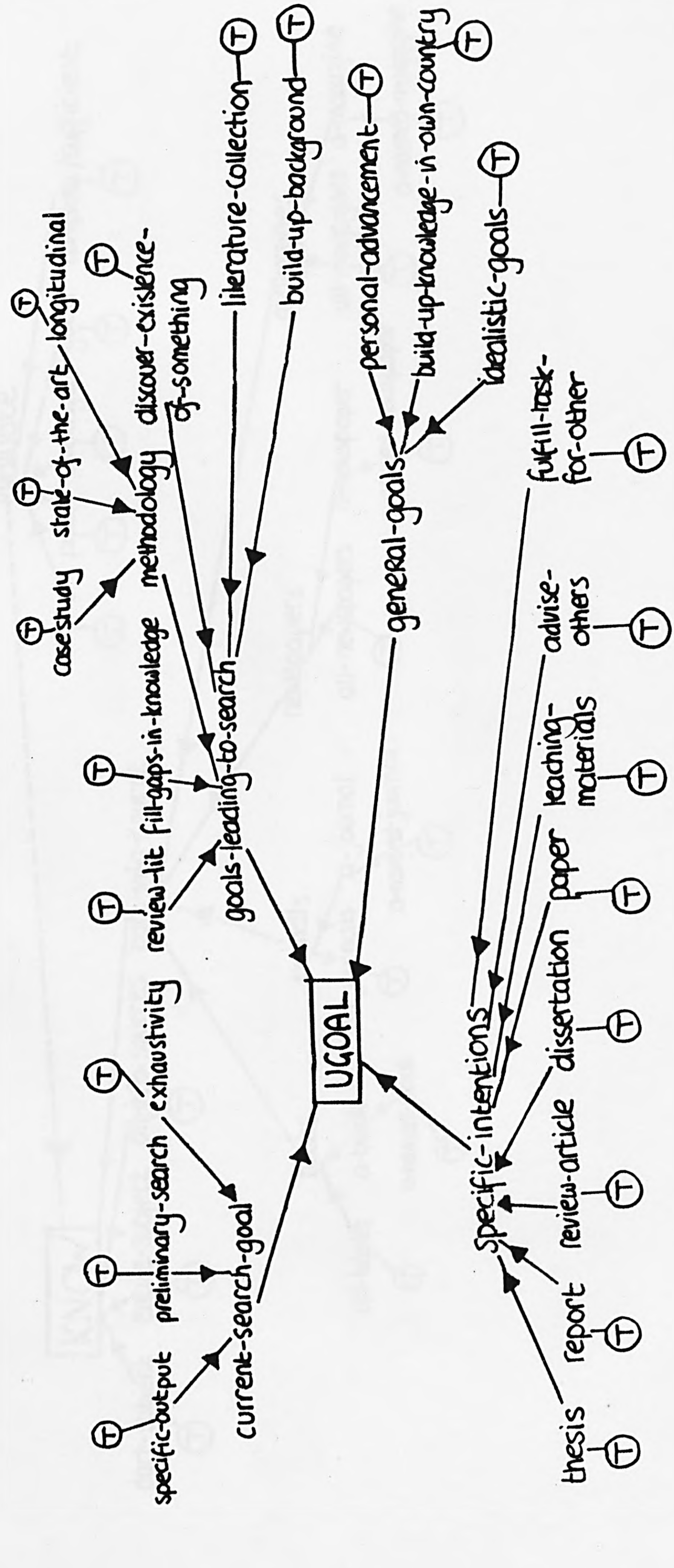
Appendix F

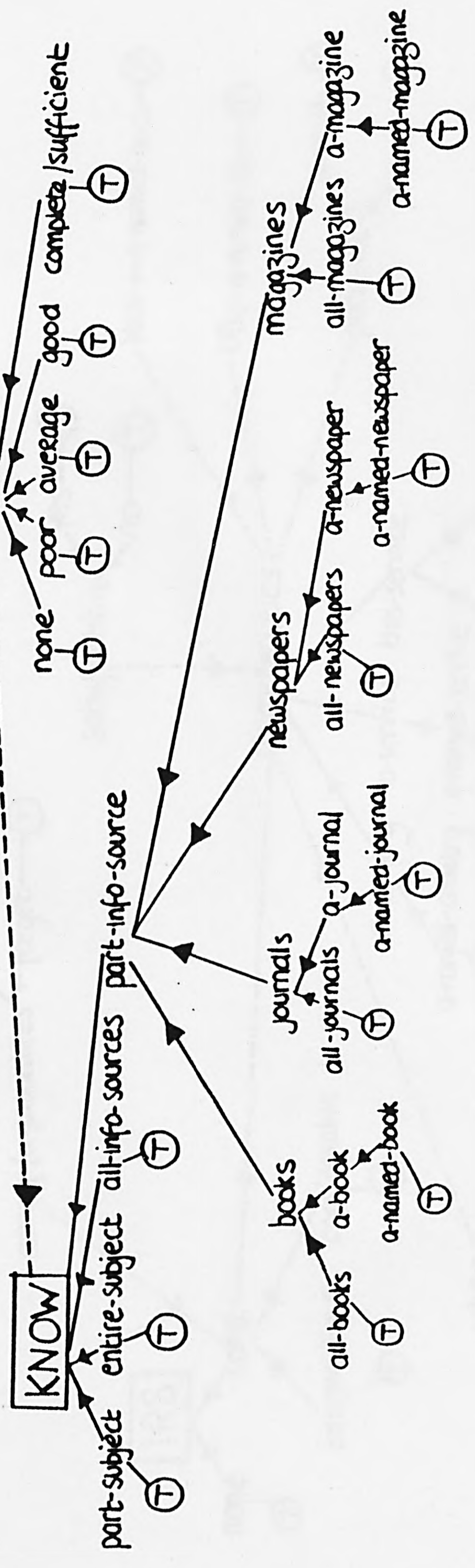
Frame Hierarchy for the User Model

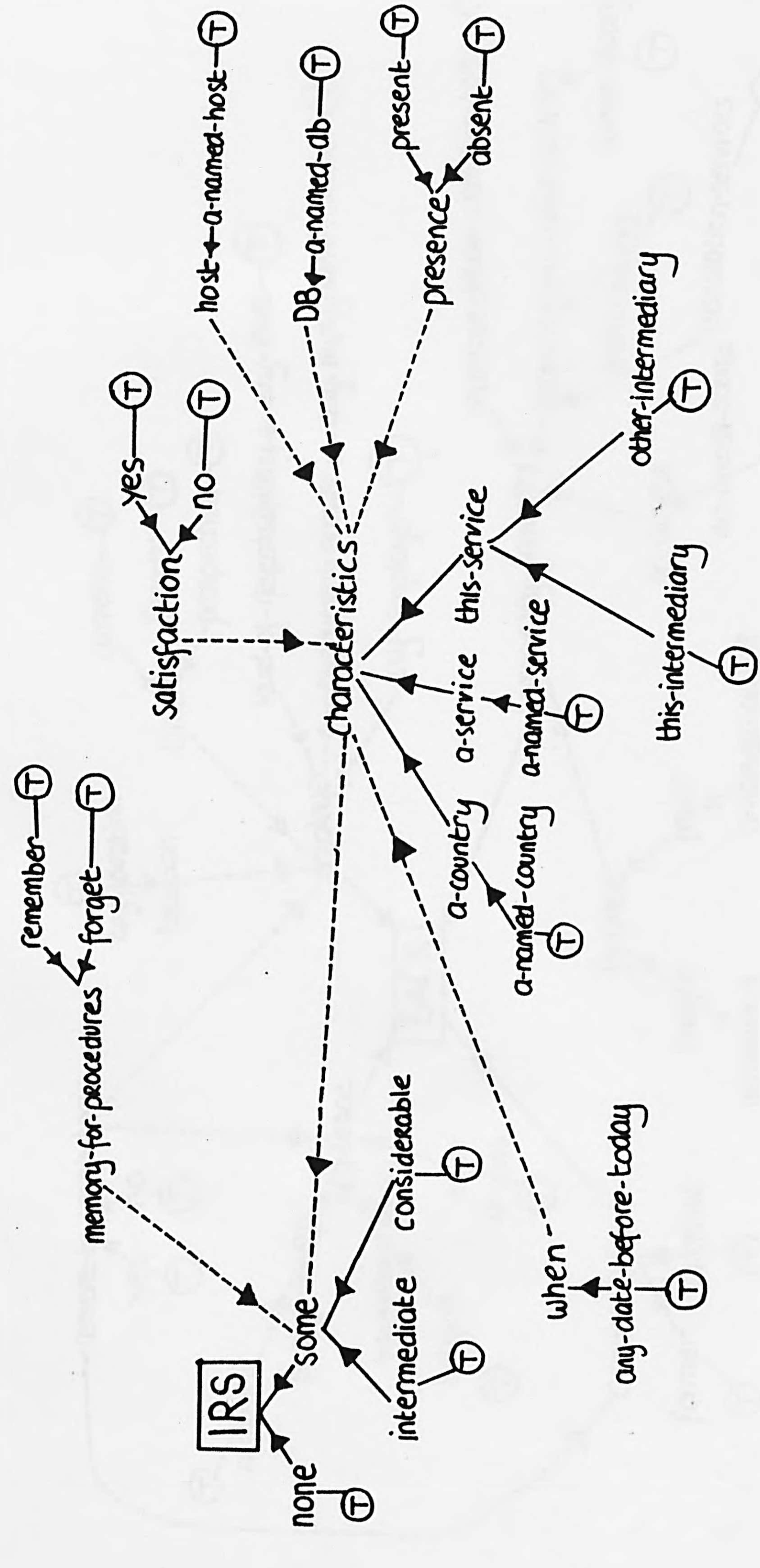
KEY:

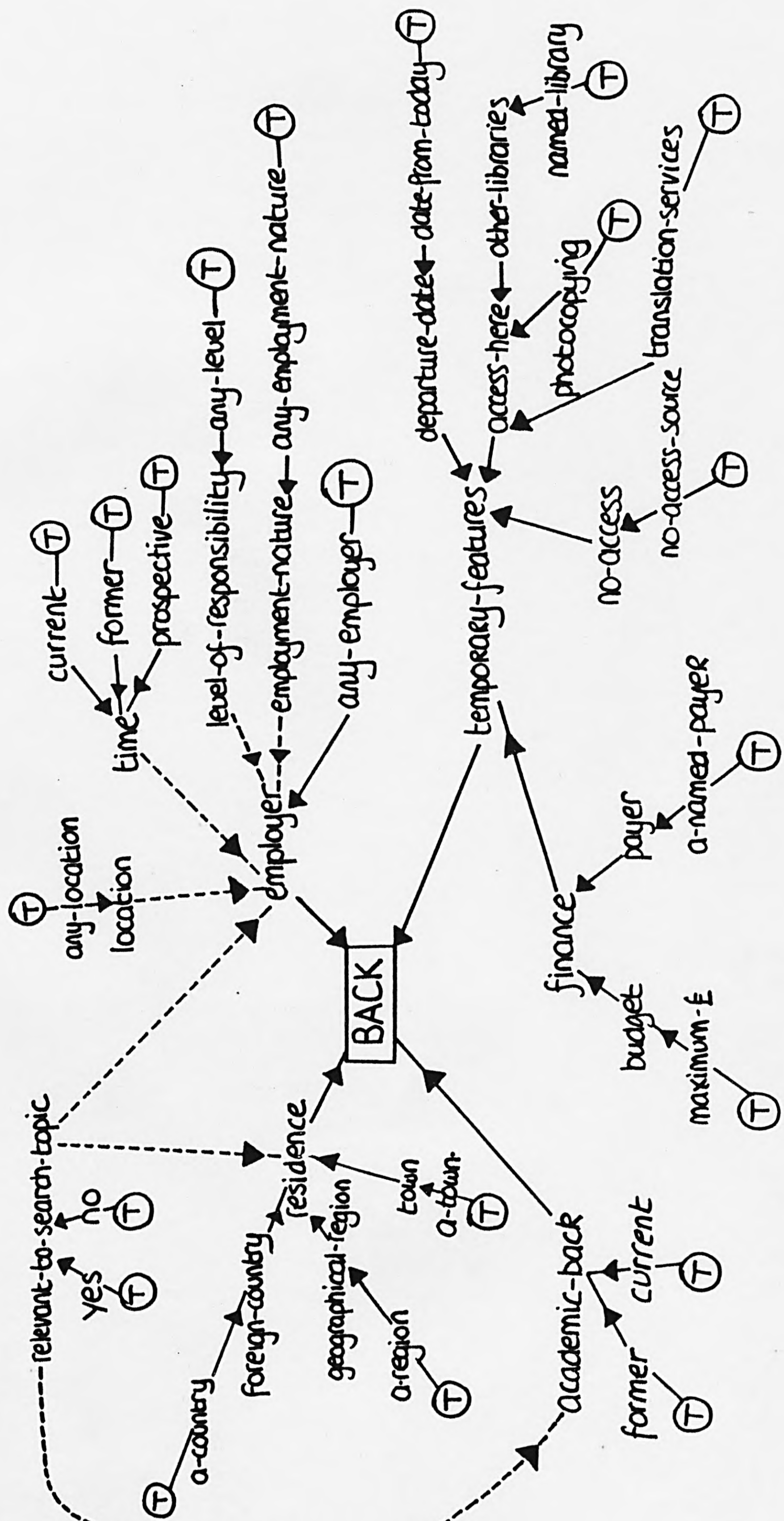
- x → y : frame x is a daughter of frame y
- x ⇢ y : frame x occurs as a slot in frame y
- x —(T) : frame x is a terminal value
- USER : determine the status of the user
- UGOAL : determine the user's goals
- KNOW : determine the user's state of knowledge in the field
- IRS : determine the user's IRS experience
- BACK : describe relevant features of the user's background


USER
prospective-academic
former-academic
current-non-academic
current-academic
position
end
intermediary
team
time
part
full
institution
funding-source
an-awarding-body
some-awarding-body
self
staff
teaching
research-staff
student
undergrad
postgrad
taught
research
master's
diploma/certificate
PhD
MPhil
non-academic-research
contract-research
research-centre
other
T


UGOAL
current-search-goal
specific-output
preliminary-search
exhaustivity
goals-leading-to-search
review-lit
fill-gaps-in-knowledge
methodology
casestudy
state-of-the-art
longitudinal
discover-existence-of-something
literature-collection
build-up-background
general-goals
personal-advancement
build-up-knowledge-in-own-country
idealistic-goals
specific-intentions
thesis
report
review-article
dissertation
paper
teaching-materials
advise-others
fulfill-task-for-other
T


KNOW
evaluate
none
poor
average
good
complete/sufficient
part-subject
entire-subject
all-info-sources
part-info-source
books
all-books
a-book
a-named-book
journals
all-journals
a-journal
a-named-journal
newspapers
all-newspapers
a-newspaper
a-named-newspaper
magazines
all-magazines
a-magazine
a-named-magazine
T


IRS
none
some
intermediate
considerable
memory-for-procedures
remember
forget
Characteristics
Satisfaction
yes
no
host
a-named-host
DB
a-named-db
presence
present
absent
this-service
other-intermediary
this-intermediary
a-service
a-named-service
a-country
a-named-country
when
any-date-before-today


BACK
employer
location
any-location
time
current
former
prospective
level-of-responsibility
any-level
employment-nature
any-employment-nature
any-employer
temporary-features
departure-date
date-from-today
access-here
other-libraries
named-library
photocopying
translation-services
no-access
no-access-source
finance
payer
a-named-payer
budget
maximum-£
residence
relevant-to-search-topic
yes
no
foreign-country
a-country
geographical-region
a-region
town
a-town
academic-back
current
former
T