A qualitative study of thesaurus integration for end-user searching

Dorothee Blocks

A submission presented in partial fulfilment of the
requirements of the
University of Glamorgan/Prifysgol Morgannwg
for the degree of Doctor of Philosophy

July 2004

Acknowledgements

I would like to thank my supervisors, Dr. Doug Tudhope and Dr. Daniel Cunliffe, for the opportunity to conduct this research, for the contacts they provided with participants and of course for their supervision and encouragement.

I am grateful to Ceri Binding, who developed the FACET system. His work enabled me to conduct the two in-depth studies and the logging he was willing to include made it so much easier for me to conduct some of the analysis.

Many thanks are also due to the participants of all studies – without their time and cooperation, this work would not have been possible.

I would also like to thank my family and friends, in particular Arnaud, Carol and Mark, for supporting me in so many ways. I am so glad to know you!

A qualitative study of thesaurus integration for end-user searching

The research conducted for this thesis investigates the impact of thesauri on users' information searching behaviour, and the integration of thesauri into information searching interfaces in order to support the searching process. The work reported here consists of two preliminary studies, which served to refine research questions and explore the methodology, followed by two in-depth studies. The two in-depth studies were conducted with FACET, an experimental system developed at the University of Glamorgan in the context of an EPSRC (Engineering and Physical Sciences Research Council)-funded project focussing on faceted information retrieval in indexed multimedia collections. These studies were conducted with the participation of museum and library professionals. The author used a combination of qualitative and quantitative methods, such as "think aloud" protocols, questionnaires, application log files and content analysis.

The in-depth studies resulted in important findings regarding the FACET interface which were considered in its further development. Findings also relate to thesaurus use in general - it was for example observed that behaviours such as browsing the thesaurus are not suitable at all search stages, so that users need to be guided in their choice of tools. One of the main findings was that conceptual problems encountered by searchers with little formal search training caused more difficulties than those related to interface design. This resulted in the conclusion that the information searching process needs to be supported by the interface and that more extensive use of thesauri in the form of support tools is possible at different stages of the search process.

Based on literature on the information searching process and data from the first FACET study, the basis for a model of information searching in controlled vocabulary enhanced systems was developed and subsequently refined with data from the second in-depth study. This model aims in particular at facilitating the design and development of such systems. It consists of a textual and graphical representation of the search stages and an account of potential problems, their causes, possible detrimental effects on the further progress of the search and suggestions on how these can be avoided, under particular consideration of tools based on controlled vocabularies.

Etudes qualitatives d'intégration des thesaurus dans la recherche d'informations

Cette thèse analyse l'impact d'un thesaurus sur le comportement en recherche d'informations, ainsi que l'intégration d'un thesaurus dans des interfaces de recherche d'informations avec le but de faciliter ce processus. Ce travail a été complété en deux études préliminaires, qui ont servi à développer les méthodes utilisées et à clarifier les questions de recherche, suivi par deux études approfondies. Ces dernières ont été exécutées avec FACET, un système expérimental conçu à l'University of Glamorgan et financé par une bourse EPSRC (Engineering and Physical Sciences Research Council) pour un projet sur la recherche d'information dans des collections multimédia utilisant une approche facettée. Des professionnels du secteur des musées et des bibliothèques ont été impliquées dans ces études. Nous avons combiné des méthodes qualitatives et quantitatives, comme le « think aloud », c'est à dire que le participant exprime ses impressions pendant la recherche d'informations à haute voix; questionnaires, un registre automatique des activités dans l'interface et l'analyse du contenu des expressions des participants.

Les études approfondies ont révélé des résultats important pour l'interface de FACET et ont influencés son développement. Certains résultats étaient liés à l'usage du thesaurus en particulier. Par exemple, nous avons trouvé que la navigation du thesaurus n'est pas convenable à toutes les étapes d'une recherche d'informations. Ainsi, les utilisateurs doivent être guidés dans leur choix d'outils. Une importante découverte a été que les problèmes conceptuels des participants, qui avaient peu de formation formelle en recherche d'informations, étaient plus sérieux que les difficultés liées au design de l'interface. Nous avons donc conclu que le procès même de la recherche d'information doit être supportée par l'interface. Par ailleurs, nous avons trouvé d'autres options potentielles afin de supporter la recherche d'informations en utilisant davantage le thesaurus, par exemple par la création des outils qui serviraient aux différentes étapes de la recherche.

D'après la littérature scientifique et les donnes recueilli pendant la première étude avec FACET, une première version d'un modèle de la recherche d'informations dans systèmes augmentés avec terminologie contrôlé a été conçu. Ce-lui a été amélioré grâce aux donnes de la deuxième étude approfondie. Ce modèle a pour but principal de faciliter la conception et le design d'un système pour la recherche d'informations. Il contient une description et une représentation graphique des étapes du modèle et une énumération des problèmes potentielles, leurs causes, leurs possibles effets négatifs pour la suite de la recherche d'informations et des suggestions pour les éviter. Les solutions proposées sont spécifiquement adaptées aux systèmes de recherche d'informations intégrant les thésaurus.

Qualitative Studien zur Einbindung von Thesauren in Benutzeroberflächen zur Informationssuche

Diese Arbeit behandelt die Auswirkung eines Thesaurus auf das Benutzerverhalten bei der Informationssuche und die Unterstützung dieses Vorgangs durch die Einbindung von Thesauren in Oberflächen zur Informationssuche. Es wurden zwei vorbereitende Studien, die zur Präzisierung der Fragestellung und zum Ausprobieren der Methodik dienten, gefolgt von zwei weiteren, ausführlicheren Studien durchgeführt. FACET, ein experimentelles System, finanziert durch das EPSRC (Engineering and Physical Sciences Research Council) und entwickelt an der University of Glamorgan, wurde für die ausführlichen Studien verwendet. Das FACET Projekt beschäftigt sich mit der Benutzung facettierter Klassifikationen zur Datenabfrage in indizierten multimedialen Datenbanken. Es beteiligten sich Museumsmitarbeiter und Bibliothekare an diesen Studien. Die Autorin kombinierte qualitative und quantitative Verfahren, zum Beispiel „lautes Denken", Fragebögen, Anwendungsprotokolle und Inhaltsanalyse, zur Erfassung und Auswertung der Daten.

Die ausführlichen Studien führten zu aussagekräftigen Ergebnissen in Bezug auf die Benutzeroberfläche von FACET, die in deren Weiterentwicklung berücksichtigt wurden. Weitere Ergebnisse hatten mit der Benutzung des Thesaurus im Allgemeinen zu tun. Es wurde zum Beispiel festgestellt, dass manches Verhalten, wie das Durchstöbern („Browsen") des Thesaurus, nicht für alle Suchphasen gleich geeignet ist und dass Benutzer daher in der Wahl der Arbeits-/Hilfsfunktionen unterstützt werden müssen. Eine der wichtigsten Erkenntnisse war, dass diejenigen Benutzer, die nicht gezielt in der Informationssuche unterwiesen worden waren, größere Schwierigkeiten wegen eigener Verständnisproblemen hatten als mit Schwächen in der Gestaltung der Benutzeroberfläche. Diese Beobachtung führte zu der Schlussfolgerung, dass der Ablauf der Informationssuche durch die Benutzeroberfläche unterstützt werden sollte. Es wurde auch festgestellt, dass die Informationssuche außerdem durch ausgiebigere Nutzung des Thesaurus, zum Beispiel durch weitere Hilfsfunktionen für verschiedene Suchphasen, gefördert werden könnte.

Unter Zuhilfenahme der Fachliteratur über den Ablauf der Informationssuche und Daten von der ersten FACET Studie wurde ein vorläufiges Modell der Informationssuche in Systemen mit Terminologiemanagement entwickelt und mit Daten von der zweiten ausführlichen Studie weiter entwickelt. Dieses Modell soll im Besonderen den Entwurf und die Entwicklung Systeme der genannten Art vereinfachen. Es besteht aus einer inhaltlichen und einer graphischen Beschreibung der Suchphasen und einer Auflistung möglicher Schwierigkeiten, deren Ursachen und deren eventuelle negative Auswirkungen auf die weitere Suche, sowie Vorschlägen dazu, wie diese vermieden werden könnten, wobei hier besonders Hilfsfunktionen, die auf Thesauren basieren, berücksichtigt wurden.

Table of content

Chapter 1 – Introduction

1.1 Introduction

The work conducted for this thesis concerns information searching by non-expert searchers in controlled vocabulary enhanced search systems. Originally, computerised information searching was a complicated activity which had to be carried out by trained professionals, using dial-up connections to access bibliographical databases which contained document surrogates. Since then, the volume of information that is published is increasing rapidly due to technological advances. More paper-based information is digitised or published in electronic format in the first place. Whole runs of journals are for example now available in full text format. However, this does not only apply to commercial databases containing scientific literature, but also to newspapers, magazines, literature and ordinary people's personal expression in the form of web pages, now possible through the advent of the Internet.

This growing volume of data makes information searching more challenging (Batty 1998), particularly for end-users who search by themselves without training or the assistance of information professionals. However, a number of approaches to providing more structured access to resources have emerged, for example in the form of educational or subject-specific gateways (see for example a listing at the University of Glamorgan web pages[1]). One well-known gateway is BUBL, which provides through "BUBL LINK" access to information resources on a wide range of academic subjects. Another gateway, ADAM (see also section 2.4.1), provides access to art and media resources. Furthermore, initiatives to use metadata on the Internet have resulted in the development of standards and technologies such as RDF[2] and XML[3], which are used to describe and exchange digital data. In addition to information on authors, date of publication, etc., records can also contain keywords or indexing terms, which is particularly common for digital libraries and museum collections. Examples of indexed databases available on the Internet are the National Museum of Science and Industry's picture collection, and the Educational Resources Information Center (ERIC) collection discussed in more detail in section 5.3.1. In order to index data collections, controlled vocabularies are required. This approach is not without problems, and Bates (2002) describes a number of the most common pitfalls of Internet Information Retrieval, many of which are related to the

[1] Urls are listed at the end of the chapter.
[2] Resource Description Framework - http://www.w3.org/RDF/
[3] Extensible Markup Language - http://www.w3.org/XML/

preparation of the data and design of the systems, leading to sub-optimal use of controlled vocabularies.

Conceived and applied appropriately though, controlled vocabularies, and more specifically, thesauri, can be used to resolve a number of difficulties in end-user searching, for example false hits, and the need to try out different synonyms as query terms (discussed in section 1.3). Thus, adopting this approach for non-professional searchers promises success in making it easier for them to find appropriate information.

This thesis investigates user-side issues related to the use of thesauri. After an investigation of methods for integrating thesauri in searching interfaces and two exploratory studies, two in-depth studies were carried out with the FACET system, developed during an EPSRC (Engineering and Physical Sciences Research Council) project at the University of Glamorgan. Difficulties during information searching were identified and possible solutions suggested. These aspects were drawn together in a model of information searching in controlled vocabulary enhanced systems.

1.2 Effects of data explosion and accessibility on information searching

Searching techniques have also evolved due to the technological advances mentioned above. Due to the popularisation of this technology through low-cost Internet access and home computers, the group of people who have access to digital information is growing. Professional searchers still perform searches for clients, but more and more end-users search themselves. This happens for example in library Online Public Access Systems (OPACs), on CD-ROMs and on the Internet in general or in specific databases, such as digital library collections, which might require a subscription. These data sources encompass a range of data types, for example full text documents, images and other media files and metadata, records from museum collections databases or indexed web resources such as the Resource Discovery Network (RDN), also referred to as subject gateways. Most of these end-users are not professional searchers, and have little, if any, search training. Some of them might even have little experience of using information technology of any kind. Search tools therefore need to be designed to be simple and easy to use without requiring extensive training, which often reduces their potential. This can for example be observed with web search engines, which in their standard interface hide any underlying functionality and thus often do not permit the searchers to manipulate parameters. A well-known problem with many search tools is the fact that they cannot distinguish between different contexts in which terms appear and also require searchers to try out all possible variations and synonyms of terms. The amount of books published on searching the Internet can be seen as an indication that current search interfaces do not provide the answer to all questions. The information explosion both on the Internet and within databases means on the other hand that simple mechanisms will indiscriminately retrieve large numbers of results. This requires either investment of much effort in order to select the appropriate results (Batty 1998), or sometimes unjustified trust that the first few items are indeed the best and that searchers can ignore the remaining ones without missing important information. In an international context, for example within the European Union, users might benefit from accessing materials published in languages other than those they can easily use for searching.

Systems providing access to information thus need to reflect the change in user groups and circumstances. They need to be easy enough to use not to require searchers to complete extensive search training, but should still provide efficient retrieval mechanisms which can locate satisfactory results. This project is particularly interested in opening up the opportunities of thesaurus-based retrieval to end-users who, in the worst case, do not even know what a controlled vocabulary is.

1.3 The role of thesauri in information searching

Controlled vocabularies are essentially collections of agreed-upon terms of which the grammatical form and spelling are controlled. Thesauri are a specific type of controlled vocabulary in which terms have been arranged hierarchically. More information on controlled vocabularies can be found in chapter 2. One of the main issues with these tools is that users often need to understand their structure before being able to employ them effectively. Sometimes, it can be difficult to locate the terms which describe required concepts and alternatives have to be selected. General accessibility and usability issues, such as having to look terms up separately from constructing the search, can also make it difficult or awkward for searchers to use controlled vocabularies.

However, these tools can assist in resolving certain retrieval issues, for example, both searcher and indexer can select from the same controlled set of terms when several synonyms could be used. Controlled vocabularies can also assist users in term selection by providing an overview of the domain (e.g. Brajnik et al. 1996; Spink and Saracevic 1997). Thesauri are here particularly useful. Since they are based on a hierarchical structure, searchers can decide on the specificity of terms and retrieval mechanisms can also make use of it for expanding the query (Rada et al. 1989; Tudhope and Taylor 1997). These techniques can even be applied to searching in collections which have not been indexed. A number of associated problems exist however for systems designers, for example displaying excessively large numbers of thesaurus terms. The very concept of a thesaurus needs to be understood by users and this poses a certain obstacle to facilitating searching. Therefore, thesauri are not necessarily a useful tool for infrequent untrained searchers, but tend to be more suitable for user groups who search for information on a regular basis in a context where the appropriate training can be conducted.

Up to this point in time, much research on information searching, which covers various different areas such as the context of the search and personal characteristics, has been reported in the literature. Despite these efforts, our understanding of the complex processes that constitute information searching is still limited. In many studies a controlled vocabulary is considered as one tool amongst others (e.g. Spink and Saracevic 1997) and few researchers have focused specifically on interactions with thesauri, exceptions including Bates (1986), Beaulieu (1997) and Fidel (1991a-c) who looked in particular at the processes of selecting thesaurus terms and situations in which thesaurus terms were used for searching. Considering the increase in electronic data and the associated difficulties for end-user searchers to locate relevant information, the possibilities of integrating controlled vocabularies efficiently into search systems is an important issue however. Recently, some novel thesaurus-based approaches to web searching interfaces have attracted attention. For example Flamenco (Yee et al. 2003; Hearst et al. 2002, see 2.4.4) moves away from traditional ways for searching and uses browsing-type interactions with a thesaurus for query construction.

1.4 Development of research questions

The author was interested in investigating ways in which thesauri could be rendered more usable or useful to non-professional searchers who search indexed collections. Two preliminary studies using the web-based OVID Medline and ERIC wizard interfaces were conducted which focused specifically on thesaurus interaction. Following these initial studies, research questions were revised so that the two further in-depth studies with the FACET system captured the search process overall and the ways in which the thesaurus can make a contribution. This required the investigation of specific problems in interacting with the thesaurus and related functionality, the potential effects the thesaurus has on the search approach and more general information searching difficulties which could possibly be supported though thesaurus-based tools.

A qualitative, inductive approach using triangulation, i.e. the combination of several data collection methods, was taken, which permitted the collection of rich data on the participants' information searching behaviour. The methods employed were screen and audio recording, observation, application log files and questionnaires. As development of a suitable research methodology formed an important objective of the research, related issues are thus discussed in detail for each study. For the in-depth studies, an efficient methodology was developed so that a detailed data analysis was possible which enabled significant modifications of the FACET interface.

The research reported in this thesis is focused on the following research issues: It aims at exploring end-users' interactions with controlled vocabularies in information searching, at identifying the problems they encounter during this process and at proposing solutions in order to resolve or reduce these associated difficulties. Alongside the consideration of aspects of efficient training, this research required the development of techniques around the thesaurus which can be used to provide additional functionality and the investigation of the possibilities for supporting the information search process, for example though provision of learning opportunities. For some of the studies, additional research questions, for example in the FACET studies relating to specific functionality, were also relevant and are discussed in the respective chapters. Findings from studies are processed and fed into a preliminary model of information searching in thesaurus-enhanced systems. This model draws the readers' attention to problems which potentially occur at specific points in the search process and suggests solutions at these Risk Points in documentation which forms part of the model. Through the model, it is attempted to make the findings from this thesis practically applicable to support the development of thesaurus-enhanced information searching systems for end-users as well as the evaluation of systems and the set-up of user studies.

1.5 Overview of chapters

Chapters 2 and 3 review the literature on controlled vocabularies with a particular focus on thesauri and user behaviour in information searching, in order to put this research into context and to provide a background on controlled vocabularies and information searching behaviour. Chapter 4 investigates research methods with a particular focus on qualitative, inductive approaches which were found to be most appropriate considering the research aims and limitations within which the research was conducted, for example limited access and number of participants.

In chapter 5, the two preliminary studies are described which aimed at testing the methods and permitting the author to gain some experience in setting up such studies, collecting and analysing data as well as interacting with participants. These studies also served as a way to review the research questions.

Chapter 6 describes the FACET system developed at the University of Glamorgan and a study conducted with this system. The data collected in the study with version 1 of the system served to analyse user behaviour as well as evaluating the system. The significant revisions made to the system are described in chapter 7. Chapter 8 reports on set-up and data analysis of a study using the revised version of the system and chapter 9 presents the results from this study.

The findings from the two FACET studies in particular have been used to develop a model of information searching in controlled vocabulary enhanced systems, which is presented in chapters 10 and 11. Due to the limited amount of data available at the time of its development, the scope of this model is at present incomplete in terms of wider influences on the search process. However, the Risk Points documentation in chapter 11 shows problems which searchers can encounter at different points of the search process, how these can influence the further progress of the search and how detrimental effects can potentially be avoided.

Chapter 12 draws together conclusions on methods and findings from the four studies and identifies the original contribution to knowledge, together with recommendations and suggestions for further research to be conducted on information searching in controlled vocabulary based systems.

1.6 Urls for Chapter 1

ADAM: http://www.adam.ac.uk/sindex.html
BUBL LINK: http://www.bubl.ac.uk/link/
ERIC collection: http://searcheric.org/
Extensible Markup Language (XML): http://www.w3.org/XML/
FACET project: http://www.glam.ac.uk/soc/research/hypermedia/facet_proj/index.php
Gateway listings, University of Glamorgan:
http://www.glam.ac.uk/findit/elib/internet/gateways.php
National Museum of Science and Industry picture gallery:
http://www.nmsi.ac.uk/piclib/index.asp
Resource Description Framework (RDF): http://www.w3.org/RDF/
Resource discovery Network (RDN): http://www.rdn.ac.uk

Chapter 2 – Thesauri in information searching

2.1 Introduction

A controlled vocabulary is essentially a list of agreed-upon terms in which grammatical form and spelling are controlled. For a thesaurus, these terms are arranged in a hierarchical structure and they thus provide a large potential for information searching and retrieval, as will be discussed later. Thesauri can be used in combination with various retrieval algorithms such as Boolean logic and ranked searching, and they can be applied to searching database fields or full-text documents. Thesauri can also be used alongside other retrieval techniques, such as Natural Language Processing (NLP) and free text retrieval, which both work with non-controlled phrases submitted by users (see Salton 1983 for different techniques). Thesauri can be used as indexing and/or searching tools. For indexing, records are described using controlled terms. When using the thesaurus for searching, searchers manually select terms or the system automatically includes synonyms and alternate forms of query terms. All variations of a term are then matched against character strings representing the records. When the thesaurus is used for indexing and searching, searchers select terms from the thesaurus which are matched against indexing terms in the collection. The FACET system employs this last approach of using the thesaurus for searching and indexing (see chapters 6-9). Although searchers' experience (e.g. Batty 1998; Chamis 1991; Fidel 1991a) suggests that a combination of free text and thesaurus searching can aid retrieval especially in terms of recall, this is outside of the scope of this research. This thesis specifically focuses on the ways in which searchers interact with thesauri, in order to identify problems and further possibilities to support search behaviour. To provide the necessary background information, section 2.2 reports on the literature of thesaurus application. Section 2.3 goes into some detail on thesaurus characteristics. In the final section (2.4), integration of thesauri into retrieval interfaces is discussed using examples from online interfaces and the literature. Due to their relevance to the work of this thesis, the Flamenco, CHIN Artefacts Canada (Humanities) and OKAPI systems are discussed more comprehensively. However, this chapter primarily aims at discussing different options for thesaurus integration. For a survey of implementations, see Shiri and Revie (2000).

2.2 Thesaurus functionality, limitations and benefits

Certain disadvantages associated with thesaurus use have not yet been resolved. Indexing executed by humans for example is time consuming and thus costly. Inter-indexer differences arise despite training and experience and due to a number of user and retrieval-related considerations (e.g. Fidel 1994). Depending on indexing and implementation of search algorithm, especially in post-coordinated systems (section 2.3.3), thesauri cannot avoid all false hits, i.e. items which are retrieved not because they are relevant but because they contain an apparently matching term or phrase, which occurs in a different context or with a different meaning (called a homonym[4]). Thus, a record for a cast-iron chair might be represented falsely as a match on two query terms of a query on "wooden chair with cast-iron feet". In a pre-coordinate system, this would be avoided by selecting "cast-iron feet" as an indexing term.

From the searchers' perspective, thesaurus-enhanced systems can be more complex to use, especially compared to standard web search engines, which apparently offer no or few query options. Nevertheless, thesauri are a powerful tool which can certainly make searching more efficient. Batty (1998) argues that the effort invested at the data preparation stage saves searchers' time because human understanding allows indexing that makes concepts explicit which are only implied in the text, or in the case of an object, could not be identified efficiently though other techniques. A thesaurus for searching that can easily include synonyms in the query improves recall and precision. Hierarchical structures allow searchers to select terms at the appropriate level of specificity or generality. Thesauri also make it possible to retrieve items across databases, which can even be in different languages, as long as the appropriate switching languages have been established between the thesauri in question (e.g. Doerr 2001). Terminology control, i.e. the use of a controlled vocabulary, saves searchers from running multiple queries, and they potentially search faster because fewer search moves are required (e.g. Fidel 1991a, b; Manglano et al. 1998).

The semantic aspects of the thesaurus structure are also helpful in presenting a domain to unfamiliar users, and can thus help them clarify task requirements. After execution of a query, thesaurus structures can also be used to present records in a meaningful way.

2.3 What exactly is a thesaurus?

This section describes the main characteristics of a thesaurus by comparing it to other types of controlled vocabularies and illustrating examples with the Art and Architecture Thesaurus (AAT) developed and maintained by the Getty Research Institute[5]. This thesaurus of nearly 120,000 terms serves to index and describe images, arts and

[4] "Bank" is an example of a homonym. It can, amongst others, signify a financial institution, an embankment or, in a German text, a bench.
[5] www.getty.edu/research/conducting_research/vocabularies/aat

decorative objects as well as architecture and documents from antiquity to present day. It is used by a number of museums and cultural institutions for the indexing of their collections, and is employed in working retrieval interfaces such as the CHIN Artefacts Canada (Humanities), which is discussed in section 2.4.3. It was also loaded into the FACET system for the studies described in chapters 6, 8 and 9.

2.3.1 Hierarchical and relational links between terms

Types of controlled vocabularies differ in the structure imposed on the terms. Some are only alphabetical, for example glossaries and dictionaries, but many have a hierarchical structure. Subject headings such as the Library of Congress Subject Headings (LCSH) or Medical Subject Headings (MeSH) formally only contain shallow hierarchies, i.e. the distance between the terms and the root or top-level term is limited. Classification Schemes, Taxonomies and Categorization Schemes group terms under broader categories and often do not follow the ISO Standards for the construction and management of thesauri (ISO 2788, ANSI/NISO Z39.19 and ISO 5964 for multi-lingual thesauri). They do not contain non-hierarchical relationships or links between terms as are found in thesauri (discussed in the next paragraph). Apart from thesauri, semantic networks and ontologies also include these types of relationships. Semantic networks are often not hierarchical and lack rules and axioms which can be found in ontologies. This means, in ontologies more specific types of relationships are defined and at times, inference rules or definitions of roles can be found, both of which distinguishes the latter from thesauri. (Hodge 2000)

The three types of hierarchical and relational links are thus one of the main characteristics of thesauri. **Equivalence** describes the relationship between synonyms; **Hierarchy**, those links which establish the hierarchical structure and **Association** the connections between more loosely-related concepts in the thesaurus. Equivalence of terms is described using for example "Use For" (UF), which denotes Alternate terms to the Preferred Terms used in indexing. Non-preferred terms can help searchers in accessing the thesaurus, and some researchers suggest that end-user searching will be greatly facilitated by a large number of alternate terms which function as entry points to the thesaurus (Bates 1986).

Narrower Terms (NT) for the subclass and Broader Terms (BT) for the superclass establish the hierarchical structure of the thesaurus (Aitchison et al. 2000). Hierarchical relationships between Broader and Narrower Terms are mostly generic, i.e. of the "is-a" type, for example "great chairs" is-a "armchairs". In specialised thesauri, "whole-part" relationships also exist. In terms of the AAT, this could for example mean arranging "arms" under "chairs" instead of under "<furniture components>". A third type of hierarchical relationship describes instances, for example specific lakes would be considered as an instance of "lakes", rather than being related via a generic, "is-a" relationship.

An association connects terms between which a semantic relationship exists, but that are not hierarchically related. The term "chairs" for example has several Related Terms (RT), "daybeds", "stools" and since recent updates, "chair-makers", i.e. the agent who produces the object "chairs".

<single seating furniture>
  Chairs
    <chairs by form>
      armchairs
      ax chairs
      backstools
      Barcelona chairs
      barrel chairs
      ...

Fig. 2.1 Excerpt from the AAT: the term "chairs" in its hierarchy and the guide term "<chairs by form>"

The Broader Term of the AAT term "chairs" is "<single seating furniture>", and its Narrower Terms are "<chairs by design>", "<chairs by form>" (see Fig. 2.1 for an excerpt), "<chairs by function>" and "<chairs by location>". The angular brackets indicate that these terms are guide terms. Guide terms help in breaking down long lists of terms into categories, based on different characteristics of division or perspectives, in this example the 95 Narrower Terms of "chairs". Guide terms are thus not used in indexing.

Term relationships can be used for the searchers' benefits. Equivalence is particularly useful when it comes to locating thesaurus terms, especially when searchers do not know the Preferred Term for a concept. Associative and hierarchical relationships allow searchers to explore the subject area and because they reflect the domain's perspective on relationships of concepts, they can help searchers gain an understanding of the topic and thus select more appropriate query terms and refine their search task. Hierarchical and associative relationships can also be used to include additional terms in queries. When "expanding" a term along the hierarchical relationships, a certain number of NTs and BTs is included by the system in the query. One of the main advantages here is that searchers do not need to include terms individually and, as it is the case in the FACET system, are less obliged to select a precise term in order to retrieve the desired results.

2.3.2 *Facets*

Facets are mutually-exclusive fundamental categories, i.e. terms are arranged according to their most prominent feature or characteristic and appear only once in the thesaurus. The AAT consists of 7 facets - Associated concepts, Physical attributes, Styles and periods, Agents, Activities, Materials and Objects. "Chairs" is for example in the Objects facet, and "chair-makers" in Agents. Each facet is broken down further into hierarchies, and some facets such as "Activities" contain several hierarchies (Disciplines, Functions, Events, Physical and mental activities, Processes and techniques), where as others, such as "Materials" only contain one (Materials). Facets can themselves be arranged hierarchically. One of the advantages of using facets is that branches can be rearranged easily and terms can be added without influencing the rest of the thesaurus.

2.3.3 *Pre- versus post-coordination*

Thesauri are more or less detailed depending on the requirements of their application area. In broader thesauri, terms normally focus on a specific field but also cover concepts linking the domain to adjacent ones. In the example of the AAT, many terms related to the creation of works of art and architecture can be found, but it also contains a relatively small number of terms related to watercraft. The Waterways thesaurus[6] on the other hand is an example of a specialised thesaurus which contains only terms related to inland waterways. Specialised thesauri sometimes contain more compound terms, i.e. thesaurus terms constructed from several concepts, so that one term is used in indexing where otherwise, several would be necessary. Although an exception, a few of these terms such as "upholstered seat" can be found in the AAT. This phenomenon is also referred to as "pre-coordination". Use of compound terms can reduce the amount of false hits further (section 2.2), but makes thesauri more costly to compile and maintain. Post-coordination on the other hand requires searchers to construct concepts from several thesaurus terms, in this example "upholstering" and "seats" or "chairs". The advantage of this approach is that fewer thesaurus terms are needed overall, and the scope to express concepts is larger as the searcher does not rely on the thesaurus developer to include all possible options. The necessary breaking down of concepts into facets in order to represent these terms can sometimes increase the cognitive load of identifying the most appropriate thesaurus terms (see section 6.7.5 for a discussion of observed incidents).

2.4 Thesauri integrated into information retrieval interfaces

Benefits of thesauri in information retrieval are best explored by seeing them in use. As part of the preparation for the subsequent studies (chapters 5, 6, 8 and 9), the author investigated two thesaurus browsers (AAT and Iconclass) and 13 retrieval interfaces integrating thesauri on the Internet, which were mainly available freely, although some no longer exist, and others require a subscription. Interface descriptions found in the

[6] Urls are listed at the end of the chapter.

literature cited later were also considered, but the investigation was limited to the information provided in the respective papers. In some cases, interfaces are adapted for several data collections, but they were not analysed separately. This is for example the case with Ovid for Inspec and Medline, where however differences in browsing arise from the use of a thesaurus with the former and subject headings with the latter due to the depth of the hierarchies.

The main aim of this investigation was to gain an overview of existing thesaurus interfaces with particular focus on thesaurus integration and potential user interactions. The author used a template which included fields for the system or interface and the thesaurus used, target user groups, other tools or assistance and their relationship to the thesaurus (some tools, such as permutated indices, are based on controlled vocabularies), further potentially useful features, presentation of the thesaurus and documents in the interface, information on query formulation and options to compile and compare data. This scheme was in parts based on the Networked Knowledge Organisation Systems (NKOS) Registry help document[7] and was extended during the analysis of the interfaces according to other categories which the author deemed important. The data collected can be used to establish an overview of the options, for example "Potentially useful functionality" or different ways of presenting the controlled vocabulary. Appendix 2.1 shows the scheme in which the sections considered relevant to this investigation contain information on one interface. In addition to the scheme, the author also took notes in an informal manner of benefits and shortcomings of the interfaces. These are not reported here, as the main objective was to identify possibilities for thesaurus integration and to see the implication of their application. Different options are illustrated with examples from the interfaces investigated. Three interfaces, CHIN, Flamenco and OKAPI, make specific use of the thesaurus in order to assist searchers in constructing their queries. This makes them particularly relevant to this research and they are thus discussed more extensively. The ERIC and OVID interfaces are described in more detail in chapter 5 in connection with the preliminary studies. These interfaces, apart from OKAPI, were also used to test the model of information searching in controlled vocabulary enhanced systems (chapter 10).

Important points to consider in integrating a thesaurus in an information searching interface are for example accessibility or the effort involved in using the thesaurus. This applies to viewing it as well as including terms from the thesaurus and from retrieved records in the query. Some systems present the thesaurus as an "advanced" tool, although

[7] http://nkos.slis.kent.edu/Thesaurus_Registry.html

less experienced searchers would particularly benefit from terminology control and gaining access to a representation of the subject domain. When thesaurus browsing is possible, rules from hypertext environments apply regarding orientation and navigation. It appears also that all approaches to integrating a thesaurus have drawbacks linked to the specific implementation, but to address these here would mean broadening the discussion beyond the scope of this research.

2.4.1 Thesaurus displays and integration in the interface

The author identified four distinct approaches to integrating thesauri into search systems. For the first one, the thesaurus is accessed by explicit request, for example clicking a button, as in the Netherlands Institute for Scientific Information Services (NIWI) interface or in OVID through checking a box (fig. 2.2, on colour plates, and section 5.4.1), which results in the controlled vocabulary being displayed. For the second form of integration, a view of the thesaurus is shown together with the query form on the same screen as can be seen in the ERIC (fig. 5.2) or BHInet interfaces (fig. 2.3). In implementations of these two approaches, the thesaurus can be used at different points in the search, but it is mainly intended for consultation when constructing the first query or after retrieving the results as part of query reformulation. "Revealing things" (fig. 2.4) at the Smithsonian website, or the multibrowser on the ADAM website (fig. 2.5) are examples of the third form of thesaurus integration where the thesaurus is the main retrieval tool. Pollitt et al. developed HIBROWSE for EMBASE and OPACs, which specifically makes use of the faceted structure of the controlled vocabulary and thus falls into this category. Mutually constraining views of each concept, which only show existing combinations of terms (using Boolean AND), allow searchers to modify their query by selecting terms from the thesaurus display or by adding further concepts (Pollitt et al. 1996; Pollitt 1997). Figure 2.6 shows the interface with two open views.

The fourth approach hides the thesaurus from the searchers. The system selects thesaurus terms based on the submitted free text query either with or without consulting the searchers. Different levels of system control which apply in this context are discussed in chapter 3.5 with regards to user support.

The actual presentation of the thesaurus can also vary both in interactivity and the amount of information shown. Aitchison et al. (2000) distinguish between alphabetical, systematic and graphical displays, though graphical displays will often also be systematic. Alphabetical displays are useful for merging thesauri, but systematic displays make discovery of new terms easier due to the presentation of relationships between thesaurus terms. The main type of systematic displays are hierarchical displays, often in the form of trees which demonstrate semantic closeness through spatial proximity. These range from traditional family trees to computationally-expensive three dimensional displays as seen online in "Revealing things" (fig. 2.4) or Plumbdesign's WordNet interface (fig. 2.7) and in the literature in Card, Robertson and Mackinlay's and Hearst's work on cone trees (Robertson et al. 1993; Hearst 1999). The ITIS fisheye (fig. 2.8) demonstrates some of the advantages of these more flexible displays – it is for example possible to show large amounts of terms simultaneously. At the same time, disadvantages such as the difficulty in returning to previously viewed terms after jumping to different sections of the thesaurus become obvious. Expandable hierarchies similar to those of Windows Explorer are more familiar and thus simpler to oversee and navigate.

Johnson and Cochrane (1995) developed one of the earliest interfaces that attempts to resolve the issue of showing Related Terms, which cannot logically be integrated into a tree structure, by displaying them separately as a cloud (fig. 2.9). Information visualisation is in fact a large field of research in itself and the reader should refer to Storey et al. (1999) for layouts, Huang et al. (1998) for types of fisheye views, and Sebrechts et al. (1999) or Herman et al. (2000) for a discussion on two and three-dimensional displays. Although well conceived displays can help users in viewing information, query formulation remains the main objective in information retrieval.

In terms of interactivity, some interfaces simply show the thesaurus terms (CHIN, fig. 2.11) while others allow searchers to browse terms, hierarchies are for example expanded by double-clicking or term information is shown when following a hyperlink (e.g. OVID interfaces, fig. 5.8). Other differences in the display lay mainly in the amount of information available on the individual thesaurus terms.

From the perspective of usability, interfaces also employ different approaches to how thesaurus terms can be included in the query. In the first, searchers select a term, for example by checking a box and pressing a button to include the term in the query (ERIC, OVID). In the second approach, queries are executed dynamically when searchers select a category/term. The "Revealing things" interface simply retrieves associated records, while interfaces designed for more purposeful searching (e.g. Flamenco), allow searchers to make further modifications, for example combination with other terms, after the query has been executed. The third approach uses the thesaurus as a filter, where the selection of further facets or narrower terms reduces the number of results until the searchers are satisfied with the number and choose to view the records (HIBROWSE, ADAM).

2.4.2 OKAPI

The OKAPI system is based on an OPAC retrieval system which was initially conceived at the Polytechnic of Central London (now Westminster University) and then developed further at City University. Key features, which are maintained in all versions of OKAPI, are:

- Probabilistic term weighting
- Relevance feedback
- Query expansion using terms extracted from relevant documents

Over time, the interface moved from a character-based display to a graphical user interface. The initial version (VT100) of OKAPI used Automatic Query Expansion (AOE). The next version (Xokapi) used Interactive Query Expansion (IQE) as it had been found that users had problems identifying additional relevant documents using the expansion facility as they had no control over it. The Enquire project version finally employed incremental query expansion, a combination of the two, which required fewer documents to be selected. Terms are automatically extracted from relevant documents but are only included in the query when they reach a threshold. Searchers can still remove these terms from the query if they wish.

A version of OKAPI is available online, and the reader is refer to the literature on these interfaces, especially Beaulieu et al. (1997) on the Enquire OKAPI project which includes screen dumps of the interface, and Beaulieu (1997) for a more in-depth explanation of the above summary on the series of studies of thesaurus-based retrieval systems, which have consistently investigated the integration of thesauri into information searching interfaces.

2.4.3 CHIN Artefacts Canada (Humanities)

The Canadian Heritage Network (CHIN)'s Artefacts Canada was one of the first web-based interfaces to integrate the AAT. It gives access to a collection of records from museum and galleries related to archaeology, ethnography, history, decorative and fine arts, indexed with the AAT. The interface is available to the general public and allows users to find pointers to more information, as well as discovering information for personal interests, study or research.

When selecting the option "Advanced search – by browsing the AAT" (fig. 2.10), the facets of the thesaurus are shown on the left hand side. When expanding the top-level categories, all terms are shown (fig. 2.11). By clicking a thesaurus term, a preliminary query on the collection is triggered. The number of records which will be retrieved are then displayed broken down using "Who, what, where, when, how" in order to aid users with searching record fields, on the right hand side (fig. 2.12, top right).

The searchers then select the appropriate category or field to search for the term by clicking the number of results associated and thus executing the query. When the results are displayed (fig. 2.13), the user has the option to reformulate the query, but does not have any further access to the thesaurus. Indexing terms are shown in the records, but cannot be navigated. Records can be viewed or printed, but not saved or processed in any other form.

Another search option, "By browsing indices", lets users search with an index in the form of an assisted search. The index functions effectively as a controlled vocabulary for searching. Figure 2.14 shows on the left hand side of the screen a list of words extracted from the selected field, marked on the right hand side by a red dot. Checking a term moves it automatically into the respective field. The searchers can jump to specific terms by entering them in the text box above the list.

2.4.4 Flamenco

The Flamenco system was developed at the University of Berkeley specifically to investigate searching in large collections, with a particular focus on browsing and the use of facets (see Yee et al. 2003; Hearst et al. 2002 for more information). The database accessed in this discussion is SPIRO, which contains 32,000 records with images from the UC Berkeley Architectural Visual Resources Library Image Database. These images had been indexed with terms from the AAT, and only thesaurus terms used in indexing are shown in the interface. The interface allows users to search using specific terms or browse AAT categories to find records. When searching, the users enter a free text phrase into the search box (fig. 2.15, top left) and are presented with potentially suitable categories, which contain the free text phrase. All results are displayed below (fig. 2.16).

Results are displayed alongside 10 main facets, which were already available for browsing on the Main screen (fig. 2.15) and those subcategories which apply to the retrieved records. Searchers can thus refine their query by selecting additional categories, broaden the search by clicking on Breadcrumbs displayed above the results (in the example "1720-1729", fig. 2.17) or by deleting previously included query terms by clicking the "x" next to them (figures 2.17 and 2.18, top centre). The records contain descriptors (fig. 2.18) and clicking on these will execute a new query.

2.5 Summary

In this chapter, the most prominent thesaurus features, namely the hierarchical, associative and equivalence relationships between terms, have been explained. We have seen that thesauri can help improving recall and precision, for example by giving searchers ideas for refining their queries, taking care of alternative phrases or grammatical forms and preventing false hits. The hierarchical structure can also be used by the system in supporting users in their searching, for example by expanding a query. However, using a controlled vocabulary means an increased cognitive load, which designers try to counteract in their implementations. Different approaches to integrating thesauri into information searching system have thus been developed, and these have been presented here using the example of existing interfaces. Three interfaces which focus in particular on using thesauri were presented in order to give readers an idea of how these tools can be applied in practice. The next chapter now investigates how searchers approach information searching.

2.6 URLs for chapter 2

AAT: http://www.getty.edu/research/conducting_research/vocabularies/aat
ADAM: http://www.adam.ac.uk/sindex.html - select "multi option browser" from the list
BHInet (obsolete): http://www.bhinet.co.uk/content/free/welcome.asp
CHIN: http://www.chin.gc.ca/Artefacts/e_artefacts_canada.html
ERIC: http://searcheric.org/
Flamenco: http://bailando.sims.berkeley.edu/flamenco.html
Iconclass: http://www.iconclass.nl/
ITIS (obsolete): http://habanero.nhm.ukans.edu/aves/default.htm
Netherlands Institute for Scientific Information Services (NIWI) (registered users only): http://www.niwi.knaw.nl/nl/SWL/intro.htm
OKAPI Pack: http://www.soi.city.ac.uk/~andym/OKAPI-PACK/index.html
OVID for Inspec (registered users only): http://inspec.edina.ac.uk/
OVID for Medline (registered users only): http://gateway2.uk.ovid.com/
Plumbdesign's visual thesaurus: http://www.visualthesaurus.com/online/index.html
Revealing things: http://www.si.edu/revealingthings
Waterways thesaurus: http://www.mda.org.uk/waterw/index.htm

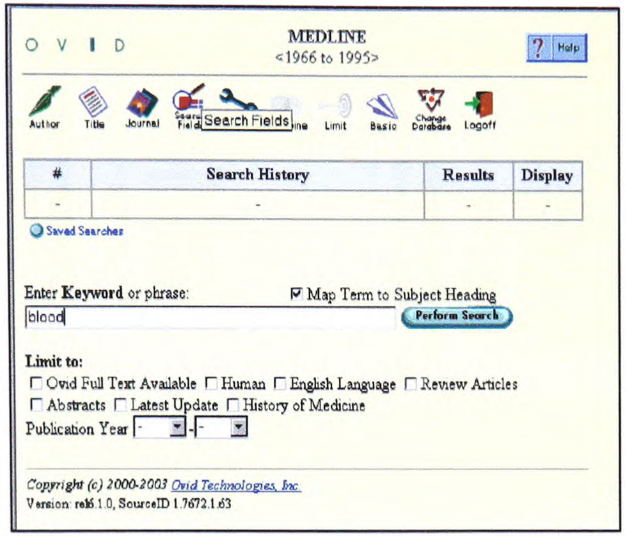


Figure 2.2 – Ovid for Medline

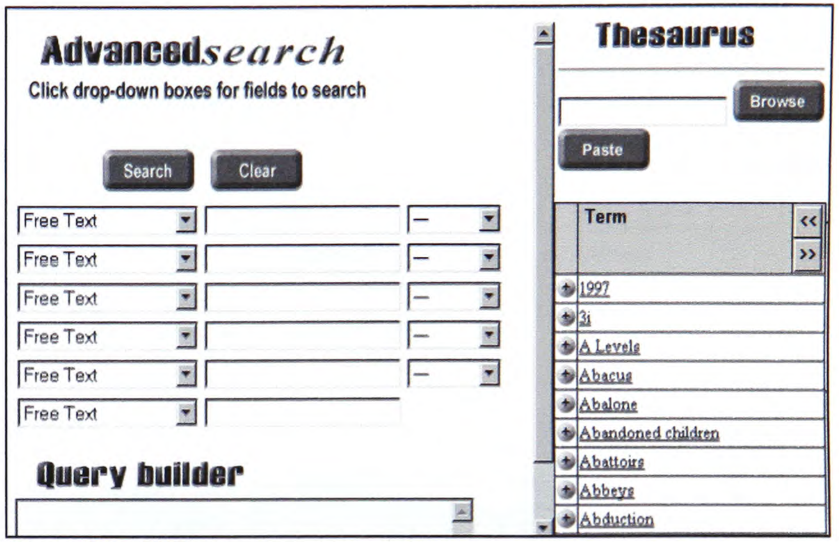


Figure 2.3 – BHInet

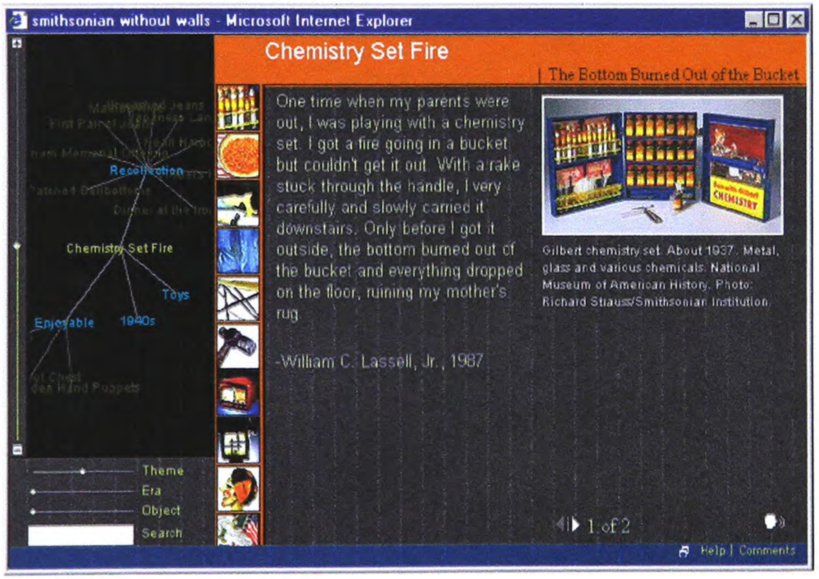


Figure 2.4 – Revealing things

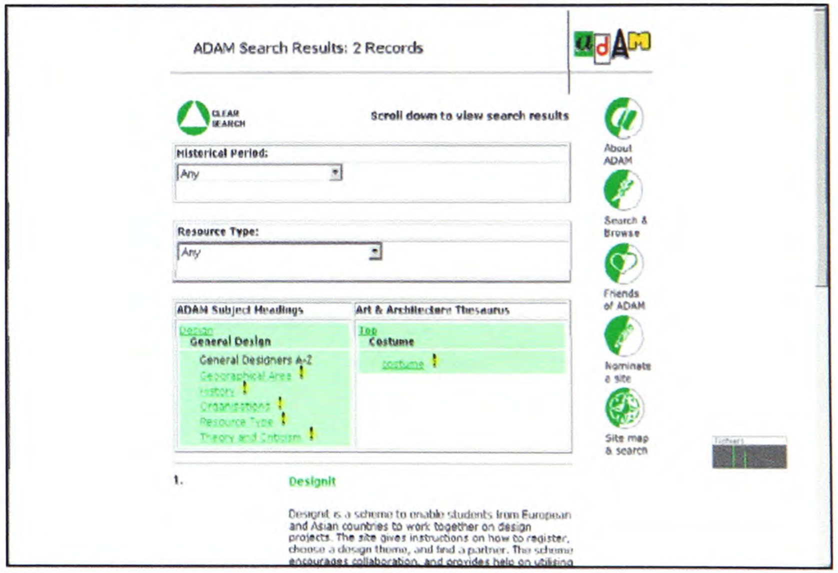


Figure 2.5 – ADAM multi browser

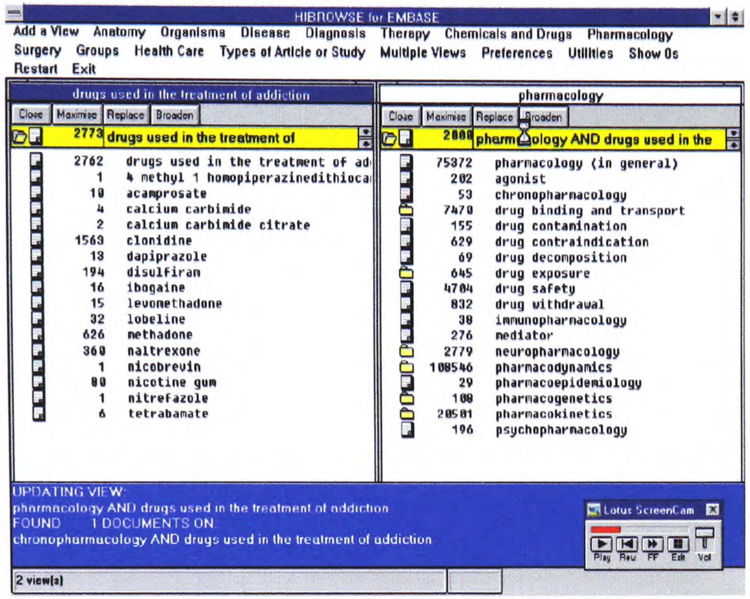


Figure 2.6 – HIBROWSE for EMBASE

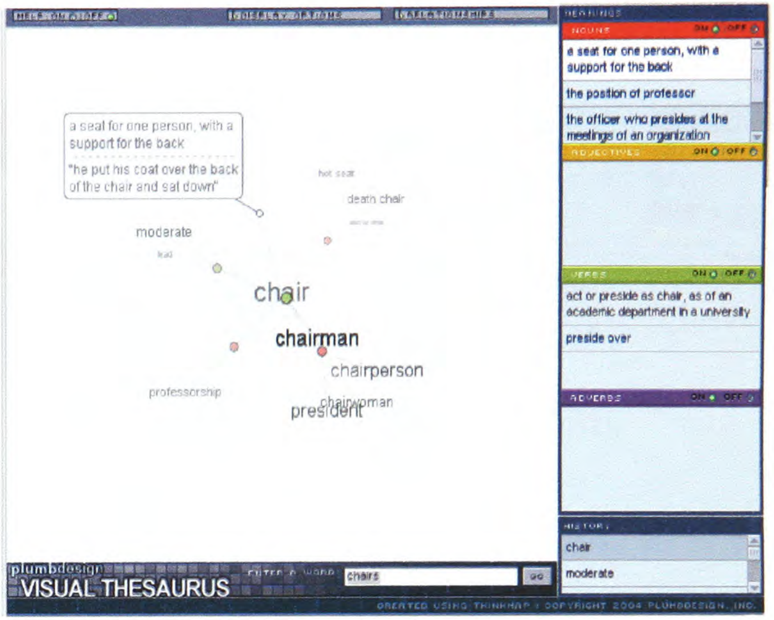


Figure 2.7 – Plumbdesign Visual Thesaurus

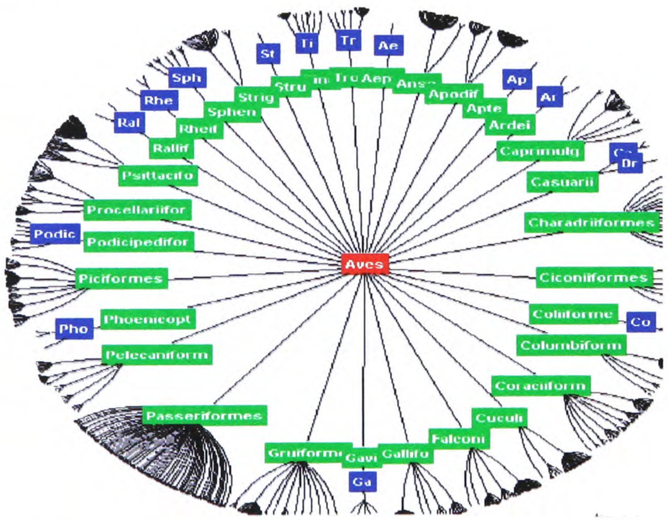


Figure 2.8 – ITIS fisheye

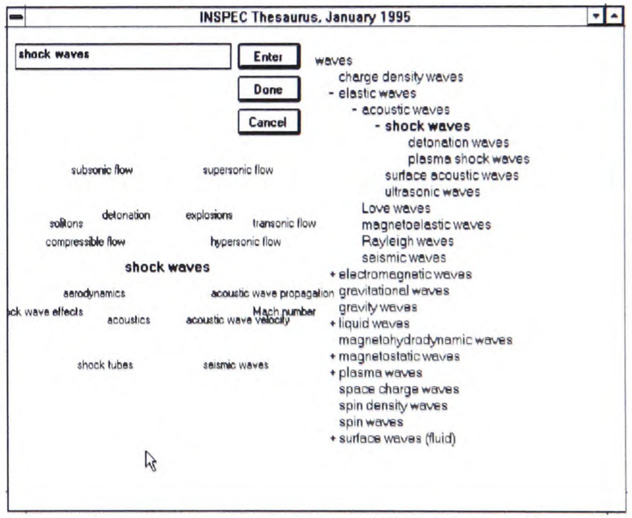


Figure 2.9 – Johnson and Cochrane's interface

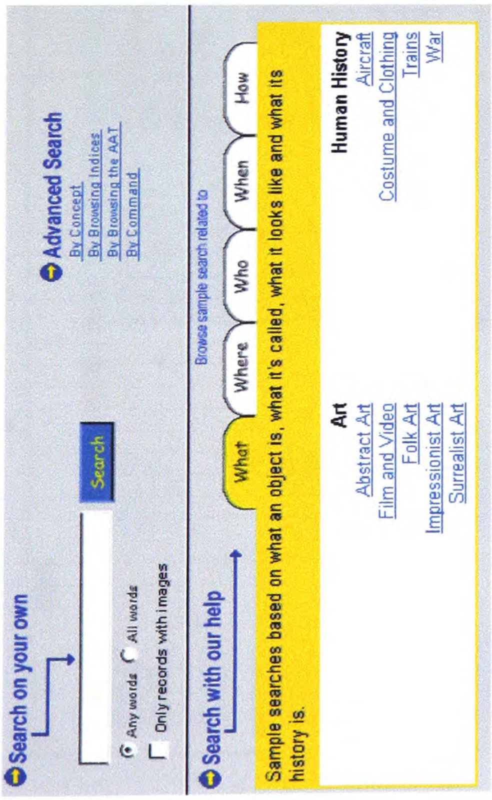


Figure 2.10 – CHIN – Selecting type of search

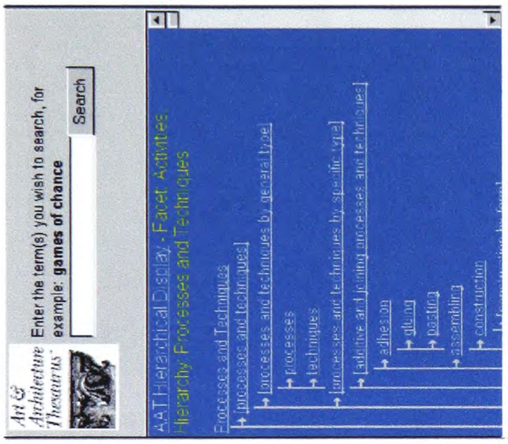


Figure 2.11 – CHIN – Display of the AAT

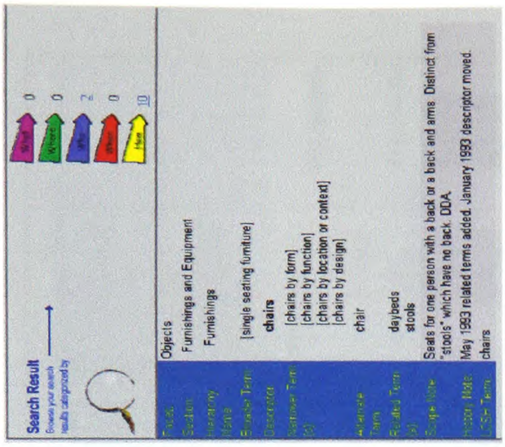


Figure 2.12 – CHIN – records per category

Figure 2.13 – CHIN – results list

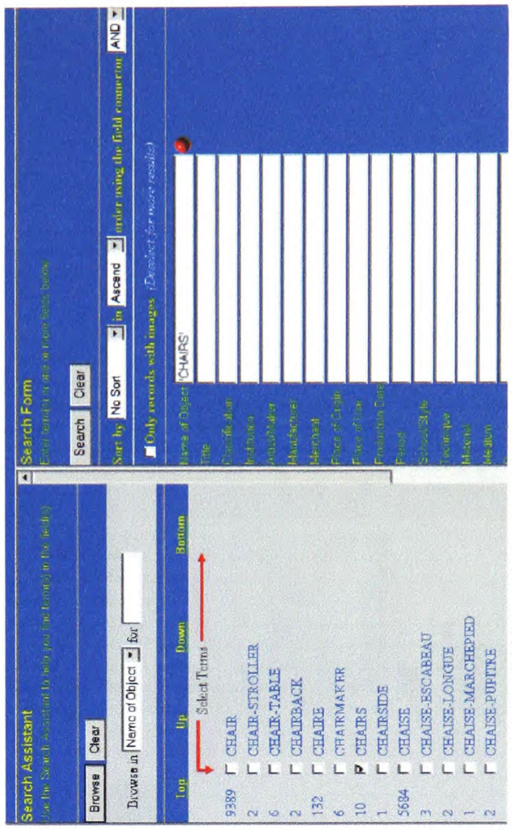


Figure 2.14 – CHIN – Using the indices

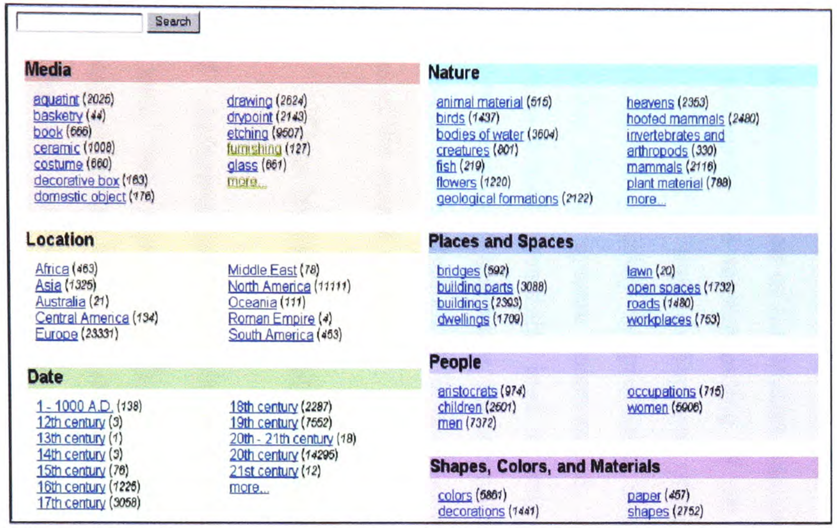


Figure 2.15 – Flamenco – Start page

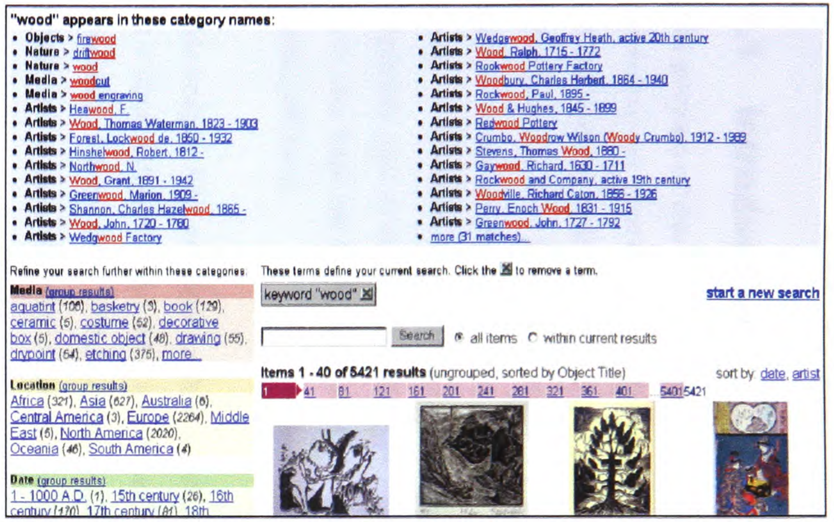


Figure 2.16 – Flamenco – Candidate categories

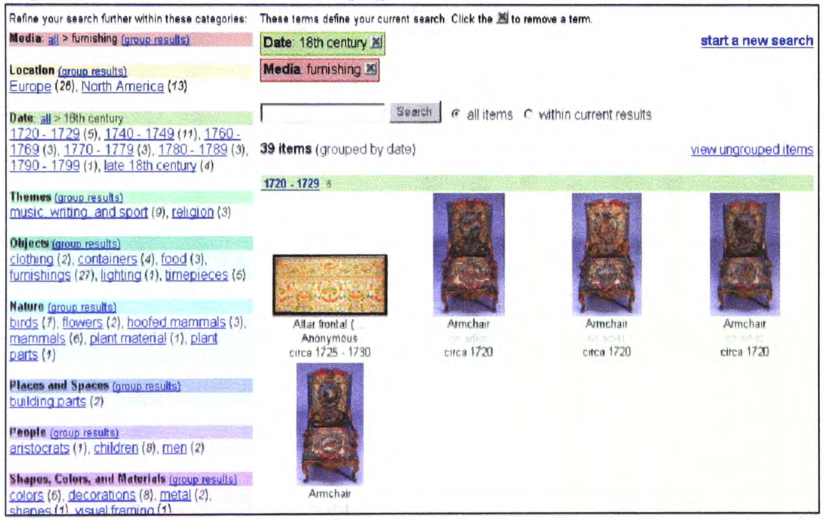


Figure 2.17 – Flamenco– Results

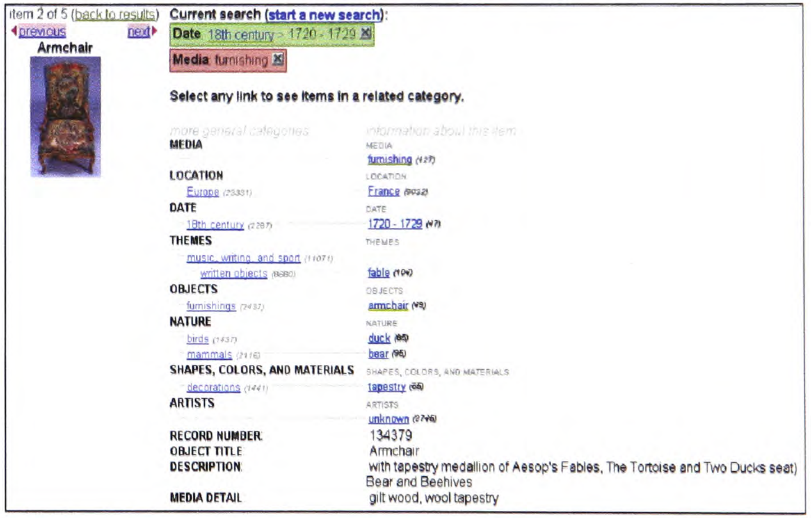


Figure 2.18 – Flamenco– Record

Chapter 3 - User behaviour in information seeking

3.1 Introduction

The purpose of this chapter is to put the research conducted for this thesis into context by providing an overview of information searching behaviour, with particular focus on online systems and end-users. Although the research is focused on the use of controlled vocabularies, this literature review also includes aspects of free text searching from studies where researchers did not necessarily make this distinction and which potentially apply to controlled vocabulary searching. Controlled vocabularies and thesauri form implicitly part of most information seeking models (for example referred to as a tool) but have also been the main topic of some research studies, for example those conducted by Fidel (1991a-c).

This chapter first presents models of information searching and seeking (3.2.1), followed by a description of types of interactions which provide searchers with different ways of accessing information (3.2.2). Term selection and thesaurus use are then discussed as important aspects of query construction (3.2.3). Section 3.3 is dedicated to influences on the information searching process, such as the user types, information need which drives the search and the search outcome. Error and support are discussed in section 3.4 because during the execution of the studies reported in chapters 5, 6, 8 and 9, it became apparent that, alongside the understanding of the search processes, these are important aspects when it comes to the development of systems for information searching. Problematic situations can possibly be prevented either through design which channels the users' behaviour or by system intervention. The system can monitor behaviour, identify situations which potentially lead to error or frustration, for example a too large or too small result set and intervene, for example by suggesting strategies. In order to be able to implement this efficiently, it is necessary to be aware of the influences on the various stages of the search process.

3.2 The information seeking process

3.2.1 Information behaviour and seeking processes models

The term "information seeking" is used here to refer to the pursuit of information in general, while "information searching" refers to searching online, either as an end-user or working with an intermediary. Although this chapter focuses on information searching, some information seeking models are also presented in order to put this behaviour into a wider context. Information seeking takes into account environmental variables, for example the users' profession or organisational structure within which the information seeking takes place. It also refers to the acquisition of information from non-electronic sources such as colleagues and records created within the organisation. Examples of models of information seeking are described in Ingwersen (1996) and Wilson (1999). In this context, researchers have also examined stress/coping and risk/reward theory (Wilson's model in Wilson 1999) as well as underlying reasons for information seeking. Dervin developed the Sense-Making approach which explains how users can understand change or overcome gaps in their knowledge, for example through information searching. The approach also proposes structures for reference interviews (used by librarians when helping clients find information) and other contexts where users encounter information, which will maximise the impact and usefulness of the information (refer to Dervin 1999 for an overview).

Ellis and Kuhlthau describe the basic elements of information searching. Ellis (1989) describes the following types of behaviour which can occur in any order when somebody tries to resolve an information problem:

- **Starting** behaviours include activities for initial searching, for example attempts to get an overview of an area of research.
- **Chaining** refers to behaviour where users follow up references in order to discover new material.
- **Browsing** describes semi-directed search behaviour, for example going through lists of authors or journals.
- **Differentiating** refers to filtering of information by criteria such as the main topic of the document.
- **Monitoring** activities are those which concentrate on viewing new material on a previously researched topic.
- **Extracting** means working through materials in a systematic way.

Kuhlthau (1991) looked at information searching as a linear process, and identified the following stages of the Information Seeking Process (ISP):

- **Initiating** refers to the very first stage of information searching where the person becomes aware of their information need.
- Whilst **selecting**, searchers identify the area on which they require information.
- Selecting is followed by **exploring**, where users increase their knowledge on the general area. However, they are still unsure exactly which information they required.
- At the **formulating** stage, users express their information need more and more precisely.

- **Collecting** describes the stage where interaction with an information system is most beneficial to fulfilling the Information Need (IN), this behaviour thus refers to the actual searching and identifying of useful information.
- **Presenting**, the final stage, refers to organising the information found and resolving the problem which started the information searching process.

Marchionini (1995, p. 50; section 6.6.5) formulates similar stages to those identified by Kuhlthau as an iterative process. Some of the iterations here include returning to formulating the query after examining the results. His model refers more explicitly to interactions with information systems, for example "execute query". This model can therefore be considered to be on a more specific level than those of Kuhlthau and Ellis.

Other research has focused on the details of the interactions that take place while users search (Fidel 1991a-c). Fidel (1991a) developed a hierarchical routine based on professional searchers' controlled vocabulary and free text term selection behaviour. Bates (1979a, b) and Fidel (1985) identified a number of tactics or moves respectively which are employed in order to modify queries with a specific purpose, for example to enlarge or reduce the result set or to modify the search topic.

Some researchers developed models specific to information seeking behaviour on the Internet. Choo et al. (2000) for example combined Ellis' model (1989) with different types of interaction as identified by Aguilar (1967) and Weick and Daft (1983, 1984)[1]. The research conducted in this thesis concentrates however on searching in a closed system, such as a database, where specific tools like controlled vocabularies are used. While the models mentioned so far effectively concentrate on the user side of the information searching process, other models represent the users' relationship with the system, which was traditionally modelled as consisting of two prongs. The user side is represented by one, the computer side by the other prong. They converge only in the comparison of the user's query formulation and the system's object representations. More recent models also allow for interaction between the user and the system (Saracevic 1996a, 1997). Saravecic's stratified model (Saracevic 1996a, 1997) for example provides an understanding on how hidden changes in cognition can affect observable changes on the surface, for example in the shape of query reformulation. The model consists of three layers. Changes on one of these will result in changes on other levels. The Surface level equates effectively to the interface. The Cognitive level encompasses users' interactions

---

[1] Cited in Choo et al. 2000

with query results and the clarification of the Information Need (IN). The third level is the Situational level. This includes the assessment of information in relation to the contextual situation from which the IN arose. Other researchers view the interaction between users and system as a dialog where the user and the system take it in turns to communicate (Belkin et al. 1995).

3.2.2 Types of information searching interaction

While searching might have been initially understood as entering a query and taking away a results set, a number of different interactions employed in seeking information have been identified over the years. Hawkins (1996) groups them into three categories from goal-oriented to potentially non-purposive:

- Hunting, i.e. searching for information
- Grazing, i.e. "information on demand" and
- Browsing, i.e. serendipitous encountering of information.

Wilson (1997) identified levels of activity ranging from passive to pro-active. Passive attention takes place outside of actually interacting with an information system, for example when watching TV or listening to the radio. In the context of mass media and museums, Treinen (1993) even identified "active dozing" – "A purposeless, planless activity that looks for gaining and maintaining permanent stimulation/diversion". A passive search is one where relevant information is found while looking for something else. Then there are active searches and ongoing searches which consist of re-running a search and updating an existing framework of knowledge of the literature. Some researchers refer to these latter activities as "monitoring" (Choo et al. 2000; Ellis 1989).

While users are required to enter their queries into the system to run a standard search, systems can be designed so that executing a query becomes close to browsing. This implementation is called "Query by navigation" (Cunliffe et al. 1997; Ter Hofstede et al. 1996). When the users follow links, the system dynamically compiles a query and retrieves appropriate results. From the users' perspective, this type of query formulation resembles browsing and like browsing, it can be easier than searching as it makes use of recognition rather than a priori definitions of the requirements. As opposed to searching, browsing has a number of functions, for example orientation or identification.

At the design stage, the available types of interactions are defined for the individual system. The choice should depend on logical approaches to the identification of the required information. Ellis (1989) for example makes suggestions of implementations that will support chaining (a browsing behaviour), which is particularly useful when dealing with articles as they contain references along which users can browse.

3.2.3 Term selection and controlled vocabulary use

Term selection can be studied from the algorithmic or the human perspective. The algorithmic approach includes techniques such as generating additional query terms, weighting terms, and creating artificial intermediaries (e.g. Fidel and Efthimiadis 1995). The human approach, which is summarised here, focuses more on the cognitive side of the search process. These studies revolve mainly around the interaction between user and search intermediary (called dyadic interaction) but can also involve the computer (triadic interaction) (Spink and Saracevic 1997). It should be noted that the selection of search terms is not a simple translation of a concept into a usable phrase, but means that the context of the search, the discourse of intermediary searcher and end-user are explored in order to identify suitable terms (Iivonen and Sonnenwald 1998).

Term selection is a concept particularly discussed in connection with systems that incorporate controlled vocabularies because of the potential obstacle these represent due to the additional cognitive load. However, for example Iivonen and Sonnenwald (1998) identified six sources for query terms in their study of the pre-online discussion between an intermediary and a searcher. These sources are the subject domain, the indexing practise, the search request, databases, controlled vocabularies and the searcher's experience. Spink and Saracevic (1997) conducted a real-life study in DIALOG on the contribution of sources to query terms in mediated (i.e. with the help of an intermediary) searching. It appears that in this study, where free text phrases from various sources as well as thesaurus terms could be used in the query, the controlled vocabulary made a relatively small contribution to the query terms selected (19% compared to 61% for the user's request forms and interaction with the intermediaries). However, the researchers' source sequence analysis indicates that once the thesaurus has been consulted, it is unlikely that the query statement or interaction between user and intermediary will yield the next query term. This thus allows us to conclude that concepts which the clients suggest are exploited first in order to identify terms. Once these have been exhausted, other sources are used to provide additional terms, but at this point, most of the suitable terms have probably already been identified. Fidel's (1991b) research seems to support this hypothesis as she found that professionals preferred to use text words under certain conditions, for example if they aimed for increased recall and the term had been provided by the end-user.

Fidel (1991b) also identified that database characteristics have a major influence (48% as opposed to 32% influence of the query) on the terms professionals used for their queries. If they for example do not trust the indexing or the quality of the thesaurus, they are less likely to use a thesaurus. On the other hand, searchers might use thesaurus terms when they assume the term has been included into the controlled vocabulary or if the term is a thesaurus term in a different database. However, Fidel also found that the thesaurus does indeed play an important role in term selection as in this study of 47 professional searchers, 75% of terms were looked up in the controlled vocabulary when one was available, despite the reservations mentioned above.

The thesaurus has not only been used as a passive source for query term selection. The OKAPI project encompasses a number of studies in which the thesaurus as part of a feedback mechanism assists users with their searches. After an automatic (Hancock-Beaulieu and Walker 1992) and an interactive version (Hancock-Beaulieu et al. 1995), the researchers established that a combination of the two, using incremental expansion, was most effective in assisting users in their search without being too prescriptive (Beaulieu 1997; Beaulieu et al. 1997 and section 2.4.2.).

3.3 Influences on the information searching process

This section deals with influences that form the direct context of the search session. These include the information sources available, the purpose of the information, the (personal) background and motivation of the users, the setting, for example the users' occupational role and status, the domain and the effect information use will have for the company (Marchionini 1995; Siatri 1999). For example Leckie et al. (1996)'s review of the literature clearly demonstrates the different requirements linked to the domains of medicine, law and engineering. Especially reports 1, 3 and 4 of the studies conducted by Bates, Siegfried and Wilde in connection with the Getty Art History Information Program (Bates 1994; Bates et al. 1993; Bates et al. 1995) are of interest here as they show that humanities scholars conducted their research differently from scientists, which influences their interaction with information search systems, for example in the form of queries submitted. Behaviour is also influenced by the fact that often, the search only forms one part of an ongoing process of information seeking (see e.g. Spink et al. 2002 for reasons behind and characteristics of successive search sessions).

The searchers themselves logically have an influence on the search process, and certain traits and typical behaviours have been identified, which are discussed in section 3.3.1.

Fidel (1988)[2] however points out that at least for professional searchers, the influence of the search in its context is greater than that of personal characteristics. As far as the task's influence is concerned, the higher the level to which the task is defined (i.e. task complexity), the easier it will be to complete it. Knowledge of more variables and their relationships (which form the problem structure) will also make it easier to resolve (Vakkari 1999). Information Need is presented here in connection with Kuhlthau's model of the Information Searching Process because it describes how various aspects of information searching are related (section 3.3.2). Lastly, the search outcome and its relevance are discussed (section 3.3.3) as these are also reflected in the model developed in the course of this thesis.

3.3.1 Types of users

People who search for information can be divided according to their search experience into three basic groups of users. End-users are those with no or little formal search training to speak of; novice searchers are training to become professionals, but lack experience and professional searchers themselves. Information searching as a profession has a vested interest in research into the information searching behaviour of novices and professionals as it serves as feedback on the training of searchers. Other researchers are interested in it for the development of support mechanisms. Spink et al. (1995) investigated how intermediaries interact with users in order to extract information on aspects end-users should consider when searching on their own. Fidel (1991a-c) also investigated professionals' term selection for the design of support for end-users.

End-user searchers are of particular relevance in the context of this research. The frequency with which this user group search can vary greatly. They might search regularly or intermittently, i.e. regularly, but with long gaps between sessions, or they might encounter a system for the first time. End-users are able to construct simple queries after little training (Yuan 1997) and their behaviour is potentially influenced by their training (e.g. Siegfried et al. 1993; Pask and Scott 1972 and Witkin et al. 1977 in Ford et al. 1994). As their experience grows, the end-users' searching behaviour changes in terms of the types of moves, for example use of a greater range of commands and more advanced features (Siegfried et al. 1993; Yuan 1997). However, they are likely to miss opportunities, especially in neglecting controlled vocabularies (Wildemuth and Moore 1995).

[2] Cited in Fidel (1990)

Nicholas (1996) investigated stereotypes of end-users and found that it is the lack of training and the "professionalism" that distinguishes them from professional searchers. Similar findings are reported in Bates (1977) in respect to indexing-specific knowledge in a study with three groups of end-users and novice searchers.

Identified types of users

Although several studies have examined user characteristics, only one set of significant domain-independent searcher profiles exists so far. Fidel (1990) identified "operationalist" and "conceptualist" searchers by examining the types of moves professional searchers made and categorised them according to whether these represented changes to the topic (conceptual) or changes which did not modify the meaning of the request (operational). Operationalist searchers have the tendency to use text words and place less emphasis on recall and this type of behaviour, although not directly influenced by the work environment, is more likely to be displayed by science searchers and searchers without a specific domain. The relevance of the different searching styles to this research is that they provide detailed information on approaches to reformulating queries which was used for data analysis, described in chapter 8 (FACET 2 study).

3.3.2 Information Need

Information seeking is generally considered to be triggered by the searchers' realisation that they require data or information. This requirement is thus a benchmark against which results of queries and the search outcome overall are judged. Smithson (1994), Kennedy et al. (1999) and Kuhlthau (1988) all investigated changes in Information Need over the duration of school projects. It is not quite clear whether the fact that these projects were imposed on the participants restricts the generalisability of the findings. Most people probably have to resolve questions without choosing to do so, thus the situations differ essentially in the implications of not finding the required information.

While Smithson does not specifically address this aspect, Kuhlthau and Kennedy agree that an early focus on an assignment topic can be problematic. Kennedy identified that the pre-focus stage thus requires high recall to allow searchers to get an overview; the focused stage, high recall and precision for them to refine their topic; and the post-focus stage particularly high precision in order to retrieve the immediately-relevant documents. Information systems need to be developed in a way that they at least make searchers aware of the stage, the associated problems and risks and potential strategies of dealing with these. In user-intermediary interaction, the intermediary is probably capable of identifying the users' search stage, and conducting the interview and search accordingly, although Kuhlthau feels that even intermediaries can contribute more during the later stages of the process. Kuhlthau's model draws attention to the fact that first of all, information systems design also has to consider less tangible influences on the affective levels. Furthermore, different types of Information Need and the associated searching behaviours need to be supported, rather than implying that users always need to "gather and complete regardless of the state of the problem" (Kuhlthau 1991 p. 370). Her model established the correspondence between different levels of Information Need (Taylor 1968), Mood and Phases of construction (Kelly's Construct Theory in Kuhlthau 1991), Levels of Specificity of tasks (Belkin) and expression (Belkin's ASK; Taylor 1968). She then established which types of feelings, thoughts and actions users go through at the different stages of the Information Search Process. (Kuhlthau 1991 p. 367, figure 3.1)

Stages in ISP	Feelings common to each stage	Thoughts common to each stage	Actions common to each stage	Appropriate task according to Kuhlthau Model
1. Initiation	Uncertainty	General/Vague	Seeking background information	Recognise
2. Selection	Optimism			Identify
3. Exploration	Confusion/Frustration /Doubt		Seeking relevant information	Investigate
4. Formulation	Clarity	Narrower/Clearer		Formulate
5. Collection	Sense of direction/Confidence	Increased interest	Seeking relevant or focused information	Gather
6. Presentation	Relief/Satisfaction or Disappointment	Clearer or focused		Complete

Figure 3.1 Feelings, thoughts, actions at different stages of the ISP, from Kuhlthau 1991

The author observed the influence of affective factors such as frustration and confusion during the studies conducted for this thesis. This model is thus of particular relevance to future research and further analysis of the data collected in order to develop the model described in chapter 10 further, for example in order to identify risks more clearly and to develop more advanced suggestions for support at the different search stages.

Related research explains how levels of Information Need and Specificity can evolve. A hierarchical structure of Goals allows shifts in the users' intentions to cause changes which cascade from the least to the most persistent goal, and which also require changes in search strategies (e.g. Xie 2002). Robins (2000) shows through the analysis of end-user-intermediary discourse how for example the end-users' uncertainty results in more shifts in focus (i.e. topic) as they explore more options.

3.3.3 *Search outcome and relevance of results*

Relevance of results is highly context-dependant. It is not the only factor which influences selection of results though. Apart from the topic or subject, data type or format, and the authority of the producer of the item, presentational issues, such as its position in the results list are also important (for example Schamber 2000; Barry 1994;

Cuadra 1967 cited in Ellis 1996). There is also a cut-off point beyond which searchers do not even view documents from a list. Fowkes and Beaulieu (2000) found that 85% of documents viewed were from the first 25, and 28% of users did not actually look at any document beyond these. This does of course not mean that no other documents were actually relevant. Other issues can also be important, for example the cost to obtain or use the item and its availability. In the library context, users often require only an approximate match which will then allow them to select the best item by physically inspecting books with the same shelf reference (Hancock-Beaulieu 1990). Relevance changes with the Information Need, and while users learn more about the topic during the search, the Information Need will change. Relevance assessments form an important part of the search process because only through them can users gauge whether their Information Need has been fulfilled, whether it can be fulfilled, or whether they need to continue searching elsewhere.

3.4 Errors and problematic situations

3.4.1 *Defining errors*

Generally speaking, an error can be considered something that does not conform to a standard. Recall and precision are traditional "standards" for measuring the quality of a search in information retrieval experiments. User satisfaction or relevance of the retrieved items are other measures. Due to the number of ways to express quality, search failure can also be interpreted in different ways (Tonta 1992). Certain types of error are easy to detect, for example no result (also referred to as "zero hits") strongly indicates a problem with the query. These errors can easily be recorded in log files for analysis of error types and causes. Other situations are not as clear-cut. If no bench mark or sample approach is available, problems cannot be detected by comparison.

Making a mistake is not an issue in itself as long as users can handle the situation sufficiently well to continue their search, which is not always the case (Yuan 1997). The interest in identifying problems lies thus in the fact that with an awareness of potential difficulties, systems can be designed to provide the necessary support which allows users to recover from the errors. Kontogiannis (1999) presents of a framework of error recovery in machine operation which could be applied to information searching.

3.4.2 *Common mistakes*

Numerous studies have identified different types of errors in both novice (Shaw 1986) and end-user searching (Barrett and Maticka 1989; Nahl and Tenopir 1996; Wildemuth and Moore 1995; Yuan 1997). Seemingly small mistakes such as spelling mistakes and incorrect use of Boolean Operators can lead to very problematic situations. Zero hit queries are especially complicated for users to deal with as they have no results to work with, which could otherwise provide enough information for them to adjust their query (Nahl and Tenopir 1996). Zero hits are also problematic because sometimes, users falsely believe that no appropriate records are in the data collection.

However, it is not always the users' behaviour which leads to problems. In some studies, researchers found that difficulties were related to the data itself, for example inconsistencies in author names and abbreviations (Dickson 1984). Some of these issues could be resolved through the use of controlled vocabularies, but researchers also had to realise that these can cause confusion even for people who have cataloguing knowledge (Bates 1977). Neglect of the controlled vocabulary is a common complaint in studies of interfaces which integrate such a tool (e.g. Wildemuth and Moore 1995), and even when it is used, it is not always employed to the maximum of its potential due to limitations in the implementation or the users' ignorance.

3.5 User support

Users support is essential for error recovery and avoidance. It is a wide-ranging field, so this section focuses on implementation issues and relevance feedback, which is a functionality that was first developed in connection with other IR techniques, but works well with controlled vocabularies.

User support can first of all be provided for technical or conceptual problems (Brajnik et al. 1996). Different user groups require different types of support. For example somebody who is unfamiliar with the interface is likely to require more and/or different information from an intermittent or regular user (Shneiderman 1998, p. 67ff). Further distinctions can be made between subject specialists and non-subject specialists, as well as end-users and professional searchers (Meadow et al. 1994) as discussed in section 3.3.1. Paterno and Mancini (1999) distinguish between adaptable systems which personalise by setting parameters and adaptive systems which adjust through observation of the user's interactions. They propose a system based on three fixed user models which will take into account three different levels of expertise and also includes an adaptive component.

Different levels of modality apply to any area of support. The assistance can be generic or context-dependent and can be controlled by the user or the system. System controlled help would for example ask the user to make a selection from a list of terms, as opposed to providing the opportunity to browse terms. The levels of control over the support provided can also vary: the users can prompt the system to provide assistance, the system can make suggestions to let the user choose, the system takes actions but lets the users intervene (penetrable), the users are aware of the system's actions but cannot influence them (transparent) or the system executes a strategy without the users' awareness (opaque)(Brajnik et al. 1996; Fidel and Efthimiadis 1995; Koenemann and Belkin 1996).

On the level of conceptual support, decision making, search strategies, the evaluation of information and the use of controlled vocabularies are processes which benefit from support (Meadow et al. 1994). Many studies of information searching actually aim at finding data that will allow the construction of efficient intermediary systems which provide support for the above-mentioned areas (Fidel 1991a; Fidel and Efthimiadis 1995; Harkness Connell 1995). Belkin et al. (1995) investigated common sequences of user interaction which are transformed into "scripts" and used in their MERIT system in order to support different information seeking strategies. Fidel (1991a) takes a similar approach by identifying the professional searchers' decisions when selecting search terms which can be represented as a hierarchical tree structure that can then be used in assisting end-users with this process. However, Fidel and Efthimiadis (1995) note that support of conceptualising the search question is particularly difficult to implement.

In thesaurus-based systems, the thesaurus is often used as a tool for term selection (Fidel and Efthimiadis 1995; Koenemann and Belkin 1996; OKAPI studies in section 3.2.3 - Term selection). One of the uses is relevance feedback, where the system for example reassigns weights to query terms or modifies terms included in the query based on records marked as (not) relevant. Ruthven (2003a) proposes advanced approaches to modification which take into consideration changes in relevance judgments and user behaviour, for example to improve recall or precision, which could potentially be applied in the context of thesaurus-searching. With regard to the different levels of control exercised by the system, Koenemann and Belkin (1996) found that it was important for users to be able to modify selected terms, even though this meant fewer terms were used in the queries than with versions where the system was solely responsible for term selection. Ruthven (2003b) discusses an experimental study of interactive (IQE) and automatic query expansion (AQE) based on the top 15 terms extracted from relevant documents amongst the 25 highest ranked items. He concludes that although in theory, IQE is more likely to give better results, it is difficult for human users to select the most efficient terms or to eliminate the least useful ones, and the paper makes suggestions regarding the support of the selection. Ruthven (2003a) found that information on query modifications made users search more interactively.

Fowkes and Beaulieu (2000) identified three levels of complexity of topics during an interactive TREC track. Easy-to-understand topics benefit more from automatic feedback. Those topic which require a higher level of interpretation and thus need to be defined better, or those where an understanding of the actual documents is necessary, benefit from the users' contribution to the reformulation. Brajnik et al. (1996) also found that automatic reformulation was only helpful when users had conceptualised the problem correctly. Situations where too few records are retrieved cannot easily be resolved using this method.

Training is another important aspect of user support, especially in more complex systems, where even end-users receive some search training, for example in the form of a course or seminar. Chapman (1981) compared training through humans and an intermediary system and found that searchers followed the example of their trainers in both cases, although the groups trained by humans showed more diversity in their behaviour. Various difficulties can be associated with training itself, for example the availability of courses and the volume of other materials available, for example help files. In order to counteract these, a group of researchers promote minimalist learning materials which allow users to access those functionalities they require (e.g. Lazonder and van der Meij 1993). Their research shows the benefits of more practically-oriented materials. Wiedenbeck et al. (1995) categorise these hands-on methods into two groups, "exercise" and "exploration". The former helps to learn basic functionality of the system while the latter is more open-ended and learner-driven in order to maintain motivation. They found that especially well-designed exercises can be a very valuable tool to teach even inexperienced end-users to utilize a new system.

One major problem with providing user support is that first of all, information searching has to be understood well and potential problems need to be identified. Only when this is understood can support really be helpful to searchers.

3.6 Summary

This chapter provided an introduction to information searching behaviour. To present the fundamental processes, several models of information searching behaviour were discussed. Readers were given an overview of different ways of accessing information and on the process of term selection in query formulation. Although in mediated searching where free text terms can be included in the query, controlled vocabularies might statistically-speaking not play an important role in term selection (Spink and

Saracevic 1997), they may be more significant than these numbers suggest, as they seem to be primarily consulted when end-users and intermediaries run out of ideas and thus function as a support option, even though they are not specifically integrated into the system to maximise their potential in this respect.

Influences on search behaviour discussed here are user types, Information Need and relevance. Even after little training, end-users can construct simple queries. While searching, their behaviour evolves, but the error rate tends to remain stable. The major differences between end-users' and professional searchers' behaviour seems to be connected to training. End-user behaviour is clearly influenced by the domain they are working in, whereas for professionals, the influences are more diverse. The complexity of relevance and the many ways in which it can influence searching behaviour have also been indicated. Information Need has been shown to differ depending on the search stage, and Kuhlthau's model provides an insight into how changes on the conceptual affective level are reflected in search behaviour.

In the final part of the chapter, error and user support have been addressed. During the studies discussed later in this thesis, it became clear that designers have to be aware not only of the fundamentals of information searching, but also of the searchers' possible misconceptions and mistakes. User support, even in its most passive form, should thus play a significant role in the conception of search systems. In this chapter, some general issues regarding the design of support have been addressed and potential tools mentioned briefly with the example of Relevance Feedback. The Risk Points of the model in chapter 11 describe problematic situations and potential ways to support the corresponding behaviour in more detail.

Chapter 4 – Qualitative research methods for user studies

4.1 Introduction

This chapter considers the application of user-centred research to information searching, with a particular focus on qualitative methods. Since user interfaces first became a focus of research in the mid-20th century, the way studies are conducted has evolved greatly, as researchers endeavoured to overcome methodological limitations. Information searching is an object of investigation in two main streams of research, Library Information Science (LIS) and Information Retrieval (IR). LIS as a discipline concentrated from its beginnings on helping users find information, while IR's initial aim was to improve algorithms, thus focusing on the information retrieval system. Although this later stream of research still exists in its own right, the searchers' perspective now plays an integral role in many studies concerned with information systems development.

After a brief description of the development of studies in IR and LIS from the mid-20th century to the present day, the main characteristics of qualitative research approaches and their applicability are outlined in contrast to those of quantitative approaches. Appropriate methods to conduct qualitative studies are presented in the later part of this chapter with a particular focus on those used in the research presented in this thesis.

4.2 History of evaluation in information searching

As mentioned above, two different disciplines, Information Retrieval (IR) and Library Information Science (LIS) are concerned with Information Seeking (IS). LIS was from the beginning of 1940's interested in information use and acquisition and was thus oriented towards user needs at an early stage. As a result, the initial focus of this research was very much on tools for the library context, i.e. catalogues and indices, but was then broadened to include information access of other population groups as well (Siatri 1999). IR on the other hand concentrated initially on algorithms and mathematical issues and later on linguistics before it was considered necessary to bear users in mind during evaluations. According to Robertson and Hancock-Beaulieu (1992), this process manifested itself in "revolutions" in the areas of relevance, cognition and interaction.

Early laboratory-based tests allowed in IR research generalisations on the performance of languages and indexing devices. Relations between entities were considered to be quantitative and relevance was the basis for all derived measures. Robertson and Hancock-Beaulieu (1992) report that Taylor (1968) raised an awareness that relevance should be considered in relation to Information Need, not query statements. Similarly, Belkin's perspective (1980, cited in Robertson and Hancock-Beaulieu 1992) that users search for information in an attempt to rebalance their "anormal state of knowledge" required the consideration of the user's actions and their cognitive processes in systems evaluation.

The advent of faster equipment meant that in the 80's, users were able to execute queries repeatedly and modify them online. This is when the processes between the users and the computer seem to become truly interactive. However, through the dialog between intermediary and end-user, information searching inherently had interactive elements. While the implementation of new algorithms, such as clustering, move control back from the user to the system, research into information searching examined individual differences and learning or searching styles. Hypertext and Graphical User Interfaces (GUIs) allowed more dynamic interaction and the new concept of exploration of information spaces appears (Savage-Knepshield and Belkin 1999). Since their third conference in 1994, TREC (Text REtrieval Conference)[3], which has become a major venue for large-scale IR evaluations, included an interactive track.

This new interactive, problem-solving focus required more interdisciplinary cognitive research approaches, and even in IR, users were now considered as part of the system. The interest in users' needs and requirements during the search was reinforced by the increasing number of end-users now executing their searches themselves rather than consulting an intermediary. Researchers looked for example at aspects such as instruction manuals and other supports. Brajnik et al. (1996) for example studied the different types of support users required during their search, and found that terminological support was urgently required. At this point, the definition of useful measures also had become more and more prominent in research and methodologies were questioned both in IR and LIS. Studies set in real-life situations were emphasised and the discussion around the concept of relevance (see section 4.3.3) attains major importance.

Research in both IR and LIS has thus changed significantly over the years. Although technically-oriented streams of research still exist, there are several areas of interest which focus on the users, their needs and requirements. The following sections will discuss practical issues related to these types of studies.

[3] http://trec.nist.gov

4.3 Qualitative approaches to data collection and analysis

Qualitative approaches to conducting a study can most easily be explained by contrasting r approaches, which is done throughout this section. Studies, quasi experiments can be distinguished depending on the level of control the ise (Haas and Kraft 1984). In studies, researchers have little or no e events, and observe rather than intervene. In quasi-experiments, a limited amount of control, while they exercise at times rigid control in he following, the terms have been used according to this definition, in s are not specific to a type of research the term "study" was used as an

eriment consists of the elements of comparison, manipulation, control . In the first step, two or more groups are compared by conducting a aspects such as errors or search speed. In the next step, the groups are gh for example training or modifications of the systems by including components or changes in the retrieval mechanisms. At least one group does not receive this training or continues to use the basic search system, and serves as a control group. Another test is then conducted in order to assess the effects of the manipulation. The comparison with the results of the pre-test allows generalisation on the effects of the manipulation. Several tests can be repeated over time before and after the manipulation - this is referred to as a time series design.

4.3.1 Purposes and settings

Both studies and experiments can have different purposes. Formative evaluations of prototypes feed back into the development of the system, and summative evaluations assist in the revision of finished products or evaluation of mechanisms. The latter, for example to compare two systems, can more easily be achieved using experiments. User-centred, qualitative approaches are more suitable for the identification of user requirements. Data on system use can be analysed to assess the system, i.e. to conduct an evaluation, or to investigate user behaviour. In these cases, a combination of qualitative and quantitative data should be considered.

Depending on the aims, studies and experiments can be conducted in different settings. For heuristic or inspectional evaluations, evaluators compare an application with a set of rules, such as interface guidelines, which have normally been applied at the design stage. These guidelines can be more or less abstract, but the former requires higher trained or experienced evaluators. To assess retrieval algorithms, test collections exist, which consist of documents, queries and relevance judgements. The documents are "loaded" into the system one wants to assess and test queries are executed. The system's performance can then be assessed by looking at the retrieved documents and their relevance. Certain types of studies can be conducted with "test subjects", who ideally correspond closely to the target user group, in usability laboratories. These allow a maximum control on environmental and system variables and the installation of all kinds of data collection equipment, for example non-obtrusive cameras and application logging (see 4.4.3). These potentially threatening environments can influence participants' behaviour, thus researchers often choose more realistic set-ups in operational environments. In order to explore some characteristics which need to be known for operational experiments, laboratory tests can be conducted first (Robertson and Hancock-Beaulieu 1992).

4.3.2 Qualitative and quantitative approaches

Information searching is a domain where researchers are still trying to understand the fundamental processes, thus many unknown factors remain and the known variables such as mood, experience and commitment to the search are difficult to control or measure. In the context of the resulting complex analysis, it is often fruitful to use holistic, inductive approaches before conducting specific experiments (Fidel 1987). These qualitative approaches investigate individual situations at great depth and account for their complexity, by analysing for example a behaviour in a specific situation. Quantitative measures are often used to inform this analysis, for example the sequence of moves or number of search cycles. Hypotheses are formed during the data analysis, not before the study. Although it can be more difficult to identify significant findings, the combination of several data collection methods, called 'triangulation', reduces the risk of missing potentially important data. Flexibility is necessary as the methods depend on the specific situation and adaptation during the data collection process can ensure better coverage and hence more fruitful analysis (Fidel 1993).

For experiments, validity and reliability are essential. Validity refers to the generalisability of findings (external validity), the question of whether causes and effects have been established correctly (internal validity) and also whether the variables measured what they were designed to measure (construct validity). Validity can easily be compromised through the design of the study if it does not prevent or measure effects such as learning, tiredness and hunger (called "maturation"). Even previous participation in studies can have an influence on behaviour as well due to familiarity with the general procedures. The changes due to the simple fact that a study takes place is also called "Hawthorne effect" after a study conducted by L. Urwick and E.F.L. Brech in 1947 where it was first noticed. Unobtrusive data collection methods exist in some settings so that this effect can be minimised. If dealing for example with real-life Information Need and intermediaries, there are often standard request forms on which users explain their information requirements (e.g. Spink and Saracevic 1997).

Reliability describes the extent to which an experiment, if repeated under the same conditions, will produce the same results. It is for example influenced by too small or inadequate samples, non-random samples or improper methods of statistical analysis (Tague-Sutcliffe 1992). It has now been generally accepted that qualitative research approaches cannot guarantee reliability.

Thus, even though more control and less uncertainty about effects possibly allows more straightforward conclusions on relationships between factors, studies should not impose restrictions solely for reasons of reliability and validity. These could prevent some searchers a priori from conducting a good search, by limiting the search strategies available (Fidel 1987). Researchers should however make every effort to be consistent and to measure for example contextual factors, which are not controlled, in order to assess their influence during data analysis. Moreover, variables should be selected and defined carefully, for example to avoid measuring effort instead of searcher performance (e.g. by counting the terms used rather than the concepts, Fidel 1987; Wildemuth et al. 1993).

4.3.3 Measures for data collection and analysis

In information searching research, traditional variables are precision, i.e. the ratio of relevant items retrieved to total items retrieved and recall, i.e. the ratio of retrieved relevant items to relevant items in the database, although these are less important for qualitative approaches. As mentioned above, it is important to capture information on the context, which can include system characteristics. Familiarity with the tools, search experience and other user characteristics should also not be neglected. Other variables commonly used relate to the search progress, for example the number of search cycles and the sequence of commands. On a more qualitative level are user judgements, characteristics of the search task and reasons for terminating a search. See for example Borgman (1996); Boyce et al. (1994); Penniman (1982) and Tague and Schultz (1989) for discussions of variables.

The particular case of relevance

The concept of relevance has a particular status. This variable has been widely used, especially in IR research because relevance judgments are necessary in order to assess precision and recall. Relevance is also important in the investigation of information searching behaviour and it is in this context that the complexity of this variable became apparent. Relevance can be considered from different perspectives (Borlund and Ingwersen 1997), for example strictly related to the information searching task (topical relevance), as it is done in simulated searches, or in the context of the searchers' real Information Need (situational relevance).

Relevance has a number of different attributes (Boyce et al. 1994; Saracevic 1996b), which can be defined for a particular context by asking the questions regarding the goal, the way relevance is to be measured, whether is it is considered to be static etc. (Boyce et al. 1994, p. 179). The searchers' assessment of relevance depends on several factors, such as the document type, the topic, the time available for the search, the order in which they are presented etc. (for example Barry 1994; Cuadra et al. 1967 in Ellis 1996; Schamber 2000; Smithson 1994 and see also section 3.3.3 – Search outcome and relevance of results). Relevance is thus highly context-dependant and difficult to define and measure. Framing relevance as "usefulness" is one option for reformulating the concept in a less ambiguous way.

4.4 Methods for data collection and analysis

The methods presented here have been selected from a practical perspective, i.e. they are appropriate and feasible for the context of the doctoral thesis within which the following research was conducted. Analytical methods tend to be used in the earlier stage of systems development and are used to establish the design and functionality. Inspection methods, such as walk throughs where an expert moves step by step through the interface identifying potential problems and statistical methods were found inappropriate for the intended investigation of user behaviour. For further information on statistics related to information retrieval research, see for example Boyce et al. (1994) and Pors (2000).

The data collection methods are thus primarily observational or aim at capturing the participants' opinions. Some of these are qualitative, for example "think aloud" protocols, which are complemented by methods which are considered to be quantitative, such as application logging. Most methods have been adapted according to the requirements of the individual studies described in the following chapters.

Some of the methods, their advantages and drawbacks are well known and are not discussed in detail, such as questionnaires. More detail is provided on those which are more specific to computer-based systems and information searching, like application logs. It is in the nature of these data that they have to be processed before they can be analysed. Data presentation is a particular issue in studies where several data collection methods are used, as large amounts of data need to be analysed in parallel. Transforming the data into more readable formats is the first step to identifying patterns. Heinecke (1995) used a chart, which represents the levels of the hierarchy through which users have navigated, and also includes time measures. Thus the data is presented in a visual way without making it too complicated, but maintaining all available information on navigation. Visual presentation is not appropriate for all data, but tables are another option. A common way of presenting data is through the use of descriptive statistics, for example graphs and values such as the median and the average (Tague-Sutcliffe 1992). The author employed a number of visual and other presentations including graphs, diagrams and tables to represent observational data, for example application log files. The choice of presentation and analysis depended due to the inductive nature of the research very much on the issues identified during initial investigations. Hence, data analysis techniques are discussed within their specific context in the study chapters (5, 6 and 8). The only analysis technique discussed here is Content or Discourse Analysis, which served in three out of four studies as a means of treating verbal data.

4.4.1 Verbal or textual data

Verbal data in different formats such as audio recordings of interviews and questionnaire responses submitted by the participants on paper plays an important part in qualitative research approaches as they are considered as a close (if not totally reliable) expression of the users' perceptions of the system and their interactions with it.

Questionnaires

Questionnaires are a common method for collecting information from users. They need to be formulated carefully, but in order to collect certain information, standard questionnaires are available commercially, for example "Questionnaire for User Interface Satisfaction" (QUIS); some of its development is described in Chin et al. (1988). The data can be both qualitative, for example answers to open-ended questions and quantitative, for example counting responses to closed questions and submitting the values to statistical analysis. Results can be compared from one user to the next, or pre- and post-test questionnaires can be administered. They can for example establish the users' goals before the task, and their satisfaction afterwards (e.g. Hansen 1998). Using a second method to verify the claims users make can counteract some of the potential bias and can result in interesting data. For example if users claim they had no trouble using the interface, an application log can reveal how many and which types of mistakes they made.

Likert scales

These scales normally consist of an uneven number of elements, normally five or seven. These range for example from "I agree completely" and "I disagree completely" and contain one neutral element. Participants are asked to rate attitudes, opinions, relevance, ease of use etc. on these scales. Some researchers have found that these gave more positive results than open-ended questions in questionnaires because in addition to the common reporting of faults and nuisances, they received some positive feedback on specific features (Hill et al. 2000).

Interviews

Interviews are another common method for collecting data (e.g. used in Fidel 1991a, Nicholas 1996). Detailed instructions on how to formulate questions and how to structure the interview overall have been developed in the Social Sciences and are applicable here. Due to the familiarity and complexity of this technique, it will not be discussed in further detail. However, time-line interviews are particularly useful in the context of information searching on a computer-based system.

Time line interviews

Time-line interviews are a way of interviewing people about past, but recent searches. First, the sequence of the search is established and written down on index cards or similar pieces of paper. Going through the process helps the searchers remember details, and it has been found that this is actually a rich data collection method (Schamber 2000). Critical incident analysis is a related method, for which participants are asked to remember a situation where they experienced problems (e.g. Choo et al. 2000). The interviewer can administer open-ended questionnaires to ensure that all important issues are covered in the interview. Schamber specifically addressed the users' relevance judgement of information and explanations of the way in which sources of information were relevant to the searches. Idealising, in this context asking participants to describe an ideal search outcome, was also part of the questionnaire-based interview. Schamber found that this was a very successful way of getting people to summarise their notions of relevance and in the context of systems development, this can be a way to clarify the users' expectations.

A disadvantage is that the questionnaires are complex to develop and that they were not easy to administer either (Schamber 2000). The researcher depends to a large extent on the users' recollection of the situation. Participants sometimes try to make themselves appear in a good light, but these cases can be identified by using a combination of different data collection methods.

Think aloud

Think aloud is a method which requires the participants to discuss their impressions while using a system. This technique is probably the most important primary source of qualitative data and requires basically no equipment, but audio taping for future reference is useful. This can also serve as a means to synchronise data collection techniques, and if the audio is recorded with a video that captures the screen at the same time, synchronisation is even more accurate. However, environmental aspects such as noise can also render some of the recording indistinguishable. Thus, information on interesting interactions can sometimes be missing, but overall, participants' comments can be very useful as they put interactions into context, explain the motivation behind them and reveal misunderstandings which would otherwise not become apparent.

Although thinking aloud creates an informal atmosphere where it can be easy to obtain an insight into what the participants really think (Shneiderman 1998 p 130; Dix et al. 1993), they can feel inhibited and may not justify their actions or not mention alternatives they might see or consider. Others might not actually voice their thoughts but their conclusions or describe their actions (Branch 2000). Here, the observer could prompt the user for comments. There are actually two ways of using this method: The observer can be a passive on-looker or interact with the searcher during the session. The latter is also called "co-operative evaluation" (Dix et al. 1993). Using the high-interaction technique, the participants are not left to struggle with the interface, which can speed up the sessions (see e.g. Tamler 1998). However, it is important that observers know to which extent they are allowed to help. A higher level of interaction can help the users criticise and make suggestions. Using low levels of interaction, the observers need to be very aware that they have to refrain from intervening, especially when the participant asks for help.

4.4.2 Content or Discourse analysis

Some of the verbal data can be used as it is, for example taking on a suggestion for the modification of an interface or the inclusion of a feature. However, no valid conclusions can be based on individual instances like this. The data need to be processed and analysed further, for example using content or discourse analysis, which is a work-intensive process borrowed from observation-focused disciplines. In oral history, complex analysis of for example use of pronouns can reveal inter-personal relationships and attitudes, which normally goes beyond the needs of information searching research. In the first step, verbal and sometimes also observational data are transcribed. By looking at transcripts of user statements, classification or coding schemes are developed. This allows work with few a priori assumptions. This process can also be used in order to fine-tune questionnaires for data collection (e.g. Schamber 2000). Content analysis normally takes place in the stages of identifying logical sections, analysing the content of the section and creating categories based on the contents by grouping logical entities (e.g. Shute and Smith 1993; Weber 1985 cited in Schamber 2000). Once these categories have been established, the grouping of the data is revised according to the final categories.

Content analysis is not intrusive, as it takes place after the data collection. It is useful for exploring issues which might be raised during the data analysis, and it provides a method for analysing verbal data to an extent that findings can potentially be generalised. On the other hand, it is complicated to develop the coding scheme and to code the data. Inter-coder inconsistencies are a common problem in classification in general and thus likely to occur even if coders have been trained and work according to rules. Working with more than one coder or re-classifying the data after a certain period of time has passed means it is at least possible to calculate the agreement and to have an indication of the reliability of the coding.

4.4.3 Observation or monitoring

Observational techniques require participants in the study to conduct their normal activities with as little restriction as possible. Video recording, "think-aloud" audio recordings and observer notes are methods closely associated with observation. However, the methods mentioned in this section refer strictly to the monitoring of activity. Mostly, these methods generate large amounts of data and they are work-intensive in that they require content analysis.

Application Log Files

A log file is a recording of what is happening on the hardware and/or software level while the subject of a study uses a computer. Logging can be done by independent software that registers hardware interrupts or by integrated features, which often include additional information into the log file, for example query terms submitted, which facilitates interpretation. The software either registers changes or states in the system (doing both would lead to redundancy). The OKAPI project (2.4.1) has used different logging tools and the related literature thus provides a good overview of different possibilities. There are certain problems such as detecting where in a GUI text was entered, and often, more data than can conveniently be analysed is captured (Jones et al. 1997). Automatic methods such as statistics or graphical displays can help in making the vast amount of data more readable. Gonçalves et al. (2002) suggest an XML standard which could allow the development of generic processing tools. Graphical tools, which show for example time lines, are available. Dynamic displays showing progression are particularly beneficial. Once the data is in an ordered form, patterns may be identified, which can be done automatically (Borgman et al. 1996). Despite these processing possibilities, it is not necessarily advantageous to log all interactions, as these data can distract from the meaningful interactions, especially when they take place by mistake, for example clicking on un-clickable application areas. The evaluator can note this kind of behaviour during an observation and take the necessary actions if it should prove to be an issue that appears to require further investigation. Data can be especially complex if an action triggers several processes (Borgman et al. 1996). On the other hand, system responses, such as the records retrieved, are sometimes not included in the file and this can make it awkward to understand the searchers' next steps.

Problem areas can thus be highlighted by inspecting log files, but more obtrusive methods are required to identify the reasoning behind the interactions (Borgman et al. 1996; Hill et al. 2000; Jones et al. 1997). Log files alone can lead to false identifications of problems because the interpretation is subjective. Robertson and Hancock-Beaulieu (1992) suggest replaying the logs to users and interviewing them or administering questionnaires. Care needs to be taken not to remove individual transactions from the context of a session, which influenced user behaviour (Borgman et al. 1996) and may contain the necessary explanations. Synchronised audio recording can furnish useful information in these situations.

Logs are convenient to conduct continuing evaluations, for example by capturing data over a period of time, for example in a library OPAC system, but it can be awkward to identify sessions in systems which do not require the users to log on (Robertson and Hancock-Beaulieu 1992). Ways around this can be to consider a pause of more than for example five minutes as the end of a session. However, the user might be reflecting on the search or looking up a search-related question. In a busy environment or at peak times (for example, a University library just after reading lists, essay titles or exam questions have been distributed), the system might not be unoccupied for the specified time period. Another tactic is to consider a topic change as the end point of the session. Even if the next search is conducted by the same user, it can be considered separately from the previous searches.

Screen capture

Many web-based applications and those aimed at end-users are primarily mouse-driven. This makes it more difficult to log interactions. Although a screen capture tool essentially records the same interactions as the log, it is much more illustrative of the search behaviour that took place than the log file, no matter how well the latter can be represented graphically through post-processing (unless of course, it is possible to reconstitute the interactions that took place). Screen recordings can be played back to the participants in a post-search interview in order to elicit more information on the session (see also Time-line interviews in section 4.4.1, which fulfil a similar purpose). Screen recording can also be very useful in cases where no extensive log is available. The preparation of the files for analysis is more extensive than if a log was recorded directly, but at least much of the information is available and can be studied in detail (e.g. Shiri 2001).

However, screen recordings can require a considerable amount of CPU time and thus may slow down applications to a level that can be considered unusable. The generated files can also be large so that evaluators have to ensure that sufficient storage is available to avoid losing data.

Video recording

In the same way that audio recording is valuable as more precise analysis can be conducted based on the exact record of what was said, video recording can be useful in order to remember the users' behaviour during the sessions. Especially in real-life settings where users interact with other people or perform activities alongside searching, this method can capture aspects to which the evaluator may initially not pay attention. People sometimes feel self-conscious in front of a camera, and even if they agree to it, this can influence their behaviour. However, over time, participants can get accustomed to the presence of the camera (as is illustrated by various reality TV programs).

4.4.4 Scenarios

System developers often create scenarios especially for the purpose of creating a good interface (Carroll 2000) or to guide walkthroughs. Although these are mainly aimed at assessing the need for and integration of different tools, scenarios can also be adapted for the evaluation of more complete interfaces because a certain insight into users' behaviour can be gained (see e.g. 6.4.3 and 8.3.4). A scenario or set task consists of instructions for a task which the participants are expected to carry out to the best of their ability. Although it is understood that collecting data in real-life situations is preferable to set tasks, it is not always possible to set up a study accordingly. One example is the formative evaluation of a prototype where no real-life users exist yet. A scenario should resemble as closely as possible standard tasks as they occur in the setting for which the system is designed and normally encompasses the description of a setting or perspective to specify the question to be answered. The main problems associated with set tasks is that users' motivation in seeing them through differs from real-life searching for information that they personally need.

4.5 Ethical issues

Whenever a study involves human participants, ethical issues need to be considered in the process. The following section will discuss some of these issues which were considered and how they influenced the research methodology.

4.5.1 Right to non-participation

Participants should always have the right to withdraw from a study if they wish to do so. Participants were informed at the beginning of each study and in advance of the sessions of the aims and nature of the study and the intended use of collected data, so that every person was able to judge whether or not they were willing to participate. During the research studies conducted here, no participant opted out. In the MSc study, students had the option to attend the lecture and the choice not to attend the practical session without any repercussions. In the other studies, all participants had volunteered to participate in the first place and also did not change their minds later on, although there were objections against audio and video recording. When requested by the participants, no recordings were made.

4.5.2 Confidentiality and use of data

Participants have to be informed about the purpose for which the data is collected and how it will be used. It is normally expected that it will not be possible to identify individuals when the data and conclusions are published. Participants were informed about this, and their identities were obscured in all reports on the studies although this made it difficult to report certain incidents of interest. It was also attempted to conduct the studies in a private setting, so that participants were not overheard by others. This was not always possible, but in those cases participants expressed that they were comfortable with the situation.

4.5.3 Avoiding harm and discomfort

An attempt to avoid discomfort was made by meeting participants in familiar environments. If during the studies any situations occurred that caused them discomfort, for example when they encountered problems, they were reassured. Tasks were prepared in a way that the participants would be able to handle them, for example they were not impossible to complete, and the author explained after difficult tasks why the problems had occurred. In the FACET studies (chapters 6, 8 and 9), it was pointed out that the interface was only a prototype, and that some problems they might encounter were linked to this. Assistance and additional information was provided during sessions so that all participants managed to complete tasks, especially during the FACET 2 study. Before more complex tasks, the author pointed out that they might potentially have some difficulties. The author was particularly careful not to leave participants with the feeling that their searching was inadequate and concluded the sessions on a positive note.

4.6 Summary

This chapter started with a brief overview of the development of studies in domains concerned with information retrieval and searching. Qualitative approaches of investigation were introduced as a means to conduct inductive, user-centred research in information searching which was identified as an important but complex area of investigation which can at present not be advanced by use of quantitative and statistical methods. Qualitative approaches have here been successfully applied by a number of researchers (e.g. Barry 1994; Bates et al. 1995; Beaulieu et al. 1997; Fidel 1991a-c; Kuhlthau 1988; Lykke Nielsen 2001).

Selected methods for data collection which are feasible for the research described in the following chapters have been discussed on a general level. These methods are concerned with the collection of verbal and observational data, which need to be prepared before analysis, for example to be transcribed and transformed into more legible formats. This can require a large effort, and during the studies, the author encountered some specific problems related to data collection and analysis which are discussed together with recommendations based on the experience gained in chapter 12.2.

Chapter 5 - Preliminary studies

5.1 Introduction

As a first step to investigate thesaurus use behaviour, the author conducted two preliminary studies in order to assess available data collection and analysis methods discussed in chapter 4 and to gain practical experience in using them. These studies both took place at the University of Glamorgan with two small groups of participants, one of them an MSc class on the Multimedia MSc in the School of Computing, the other a group of research assistants in the School of Care Science.

The participants' motivation was a concern. It is a common experience in studies that "substitute" users do not have the same attitude as one would expect from real-life users. The study with the MSc students was therefore conducted within their class-time as part of their course, hence they were not required to sacrifice any additional time to participate. A database with data that could be useful to them was selected, and the tasks were relevant to their course. The research assistants used databases for their work, so in return for their cooperation, the author developed an advanced online searching tutorial for them.

One of the restrictions imposed on the studies were that databases used had to be web-based in order to be accessible simultaneously to the whole MSc class from the computer laboratories and for each for the researchers, from their computer, without time restrictions. In order to observe thesaurus use, the data had to be indexed with a thesaurus, which was also available for navigation online. The ERIC (Educational Resources Information Center) database was chosen for the MSc group because of its online and computer-based learning content, which related to their course. Records are indexed with the ERIC thesaurus. The Care Science researchers used the Ovid Medline interface, which was very appropriate for the study as it uses the hierarchically structured Medical Subject Headings (MeSH) for indexing and searching. For more information on both interfaces refer to sections 5.3.1 and 5.4.1.

5.2 Research questions

These studies did not aim at evaluating the interface design and system functionality. The main objectives were assessing the methodology, data analysis techniques and thesaurus use. These studies were also meant to identify research questions which could be pursued with the available data collection and analysis methods.

A number of different techniques which all had advantages and drawbacks (chapter 4) were considered. One of the main issues was to find an efficient way of combining the different methods in order to reduce the effect of the drawbacks. Analysis of the data was thus necessary in order to assess to which extent this had been achieved, but it was not considered appropriate to analyse the data to the same level of detail as in the subsequent studies.

With regards to the thesaurus, navigation and the effect on searching behaviour of the use of thesaurus terms compared to free text terms were of particular interest, and data collection concentrated on these areas.

5.3 MSc group using ERIC

5.3.1 Study set-up

5.3.1.1 Participants

The participants in this study were students on an MSc course in Multimedia. They took one particular module through which they had been introduced to some material on indexing and retrieval techniques, including thesauri and controlled vocabularies. They also learned about online searching in a lecture with the subject librarian for computing. The study hence complemented this theoretical knowledge with some practical experience of how thesauri work. At the time of the study, most participants had probably searched for some literature, but their conscious experience with thesauri was limited. Although more people attended the introduction session, complete data is only available for seven participants.

5.3.1.2 Training

One week after the tutorial on online searching, the author gave an introduction to the ERIC interfaces and thesaurus as part of a lecture on the theoretical side of thesauri. The introduction was designed so that participants knew how to search the ERIC collection both with and without thesaurus. The interface is in fact very similar to a standard web browser interface, and given the level of computer literacy of the students, it was predictable that they could adapt to it easily. The participants were also informed of the purpose of the study.

The search session, during which the data was collected, took place after the lecture in a computer laboratory during the students' tutorial time. It is common practise that students are first introduced to concepts during the lecture and then apply them in the following practical session. This study was thus conducted under conditions the participants were familiar with.

Those students who chose not to participate had the opportunity to leave after the introduction. No register was taken either during the introduction nor during the search session, so that students did not feel under pressure to participate in the study.

5.3.1.3 Information on Eric

The ERIC data collection can be accessed through two different interfaces. The standard interface provides a text box for free text phrases and some search options which can be set by selecting from drop-down menus. The default setting is "exact match", and can be set to "relevance", but it is unclear how this affects the results. The user also has the option to change the query option for "Words" between "exact", "some" and "most" terms (figure 5.1). The wizard interface, which is linked with the thesaurus, contains the same options, but is more structured in that users can enter terms and phrases into three text boxes called "sets", representing three concepts which can be ANDed (set 1, 2 and 3) and ORed (for the third set)(figure 5.2). Below the query-related part of the screen, the user can enter a term to search the thesaurus by clicking "Look up". On the right hand side of the screen, in a separate frame, the exact match, or if none exists as with the example of "Multimedia", a list of suggestions, is displayed. The user can then click on one of these terms and see information about it, for example the scope note, Related (RTs), Narrower (NTs) and Broader Terms (BTs). Users select terms by checking boxes. Buttons below the thesaurus display allow adding of selected terms to the respective sets.

Although the ERIC thesaurus is used for indexing, the ERIC system does not distinguish between thesaurus and free text terms. People can thus use terms which are not in the thesaurus in their search. It is also not necessary to locate thesaurus terms in the thesaurus, they can simply be typed into the appropriate set. When matching records, query terms seem to be matched to any field of the record, for example if terms from the thesaurus only appear in the abstract, the record will still be retrieved.

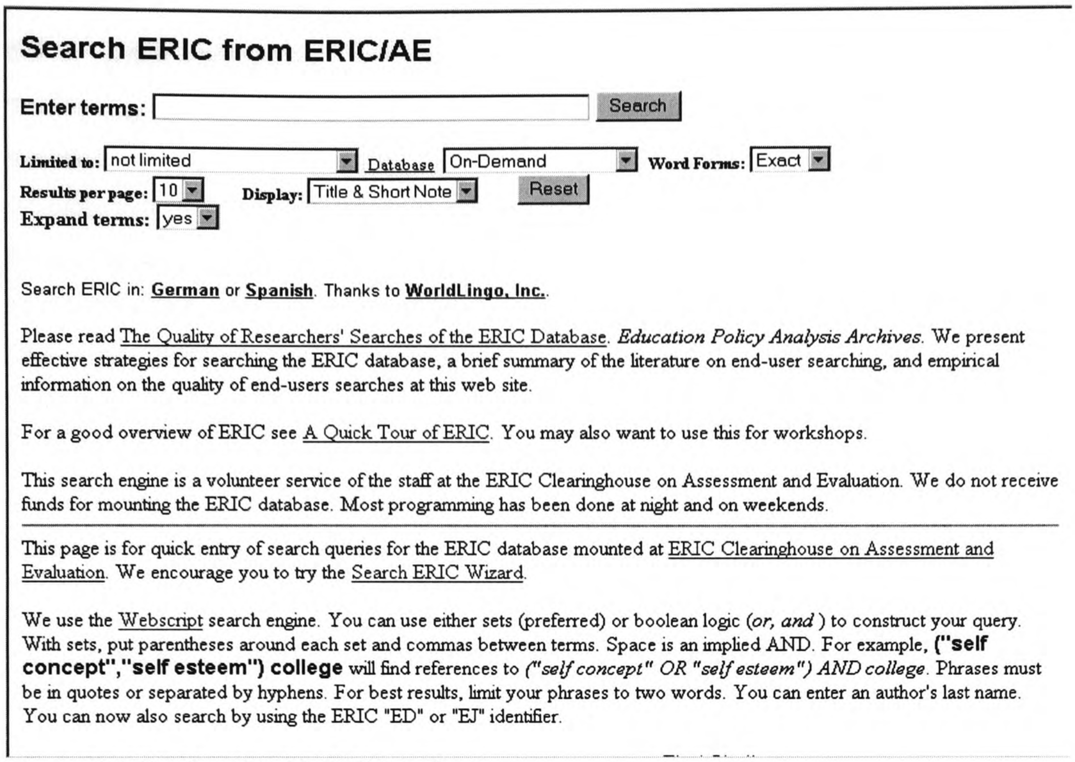


Figure 5.1 The Standard ERIC interface

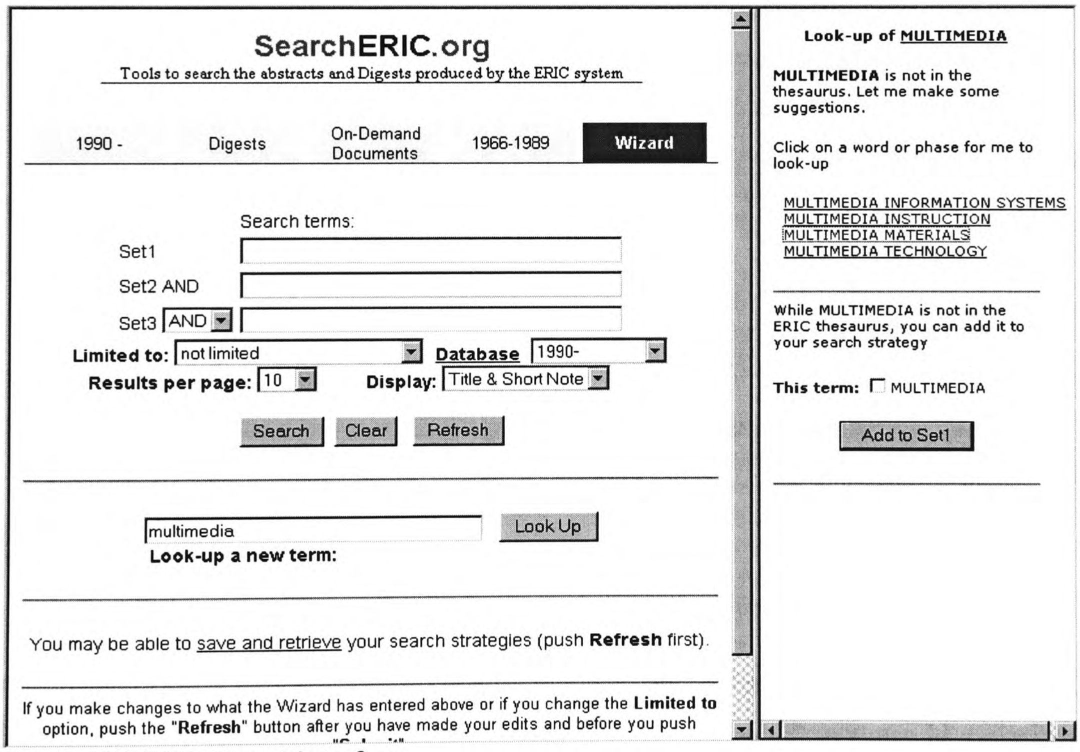


Figure 5.2 ERIC Wizard interface

5.3.2 *Data collection*

The setting of the study imposed numerous restrictions on the data collection. As the interface was part of a proprietary system accessed over the Internet, it was not possible to implement any logging. Video recording across the room would not have revealed any useful information due to the lack of information on users' interactions with the interface and at this time, no screen recording software was available. The laboratory where the study took place was large so that speech would probably not have been captured well and large amounts of background noise would have negatively affected the quality of an audio recording. Moreover, with several candidates completing tasks simultaneously, it would not have been possible to make much sense of their comments: Even if it had been technically possible to record all of them, this would still have been chaotic as participants worked individually. The author instead looked over participants' shoulders every now and then, as it is commonly done during tutorial sessions, to ensure no grave problems were occurring and noted participants' comments after the session. The task handout described below (also shown in appendix 5.1) was used as the main means of data collection. Participants were asked to select about five records from the thesaurus search results which they thought most relevant and save them in a globally accessible folder on the student network space. All participants found relevant records, but due to a problem with the folder, these files were saved on a floppy disk instead. This slightly awkward arrangement was chosen so that participants did not have to disclose their identity as task sheets were marked with Greek letters, and result files were to be saved with the same letters as file names.

5.3.2.1 Task sheet

The task sheets presented them with the scenario below and instructed participants how to progress through the session.

> In work, you are involved in a project group looking at developing multimedia materials for distance learning. These are targeted at adults who wish to acquire new knowledge and skills for a wide range of reasons. You and your colleagues are now interested in making some learning material available on the web. Without considering issues like regulating access and charging users, imagine that it is your task to write a report to discuss with your group and project management. In this tutorial, you are to create a bibliography of about 5 references, which seem to provide a good coverage of the issues involved.

The scenario had been designed for this study under consideration of the ERIC thesaurus. Some keywords from the scenario did not exist in the thesaurus in order to encourage navigation.

Participants were asked to first execute a free text search in the standard ERIC interface. The same search was then to be conducted in the ERIC wizard interface which provides access to the thesaurus. Participants were asked to use it and to write down the following information on the task sheet:

- Terms and statements used in free text search
- Terms and statements used in the thesaurus search
- Brief descriptions of how thesaurus terms were selected

One student submitted a slightly more detailed description of their search process.

At the end of the thesaurus search, participants were asked to save about five records which they deemed most relevant. These files were later used in the analysis in order to identify how many thesaurus terms resulted in the retrieval of each record. The terms and description of the thesaurus term selection were used to establish how participants had navigated the thesaurus.

5.3.2.2 Result record files

The ERIC interface gave users the option to save selected records to file. The query submitted is saved together with the number of marked documents, the document identifier for the ERIC database, bibliographic information on the records (author, year of publication etc.) and the descriptors.

5.3.3 Data analysis and description

5.3.3.1 Progress of the sessions

Participants were instructed to first execute a search using free text terms and briefly evaluate the results in order for them to become aware of the influence of the thesaurus on the retrieved results. They were then asked to complete the same search using the wizard interface and thesaurus terms. No detailed data is available on how exactly each person progressed, but seven task sheets with data were returned together with selected results, one without results file. These documents show the initial query and the thesaurus query, and also contain some information on considerations participants made when selecting terms from the thesaurus. They were allowed to reformulate their queries so that some sheets contain information on several queries which the participants had submitted.

Possible learning effects were welcome, but expected to be limited as participant interaction with the first set of records was superficial. Although analysis focused on comparison between thesaurus and free text term used, the primary objective was to assess the possibilities of data collection and analysis.

During the sessions, participants commented on not being able to find certain terms, like "distance learning", in the thesaurus. No other problems were reported. As no individual observations took place, it is impossible to say whether and with which feature or part of the task participants struggled.

The seven files[4] saved contained a total of 52 records, with a range from 5 to 12 and a median of 6 (see Appendix 5.2, "documents retrieved" row for individual values.) One person executed their own scenario in addition to the set task, but this is not considered further in the analysis due to lack of comparability with the other searches. Table 5.1 shows the query terms used in the free text and thesaurus task.

User ID	Non thesaurus search statement (Q = Query)	Thesaurus search statement (S = Set – up to 3 sets in one query)
Beta	Q1:Multimedia distance learning Q2: Distance learning AND Multimedia developing Q3: Multimedia ? for distance learning	S1: "computer mediated communication", "open universities" S2: Multimedia
Gamma	Q1: Multimedia materials, distance learning, Q2: Multimedia, distance learning Q3: Multimedia, distance learning, development	S1: distance learning, Multimedia
Delta	Q1: Multimedia AND distance learning Q2: Multimedia AND distance learning development	S1: Multimedia AND distance learning development S2: "Multimedia instruction", "educational media"
Epsilon	(distance learning, homestudy) AND learning material, study material, Multimedia material)	S1: "distance learning", "continuing education", "home study", "independent study", "lifelong learning", "non traditional education", "open university", "part time students" S2: "Multimedia material", "learning material"
Eta	Multimedia distance learning web	S1: hypermedia, "Multimedia instruction" S2: web S3: distance learning
Iota	"Multimedia materials" AND "distance learning"	S1: "Distance education" S2: "Multimedia instruction"
Kappa	("distance learning", "computer based learning") Multimedia	S1: "Distance education", "computer assisted instruction" S2: "Multimedia materials" (S3? : "adults")
Theta	Q1. Learning web Q2. Learning web delivery	S1: world wide web, online systems S2: adult education, education environment

Legend for Table 5.1

? – Information unclear on the handed-in form.

Table 5.1 Query terms used in free text and thesaurus searches

[4] Participant Gamma did not submit results files, where appropriate, the returned handout is considered though.

5.3.3.2 Use of thesaurus versus free text terms

From the returned task sheets and saved records, the query terms and other terms participants had considered during the search were extracted and entered into spreadsheets for further analysis (Appendix 5.2). In the first step, thesaurus and free text terms were distinguished and the amount of query terms used in each search was compared. Some terms used in the first, free text query actually corresponded to thesaurus terms. Systems like INSPEC distinguish between a term as a free text term and a thesaurus term, and results would have differed, but in ERIC, they have the same effect. This does not affect this study, as the results retrieved did not play a major role for the analysis.

Compared to the free text search in the standard interface, two people used less, four more and two the same number of terms with the thesaurus interface. Possible reasons for using fewer terms are that they could not find suitable terms and the participants' statements support this. Some dropped free text terms when using the thesaurus. The wording of the instructions might have led them to believe that they could not mix free text and thesaurus terms. When thesaurus terms had accidentally been used in the free text query, they were sometimes not used in the actual thesaurus search. Users might have seen the definitions and realised that some terms were inappropriate. For example the term "development" is used in a very general way rather than to mean "development of a system". Considering that the same task was executed twice, a certain amount of learning probably also took place. When they used more terms, they might have found terms in the thesaurus they did not think of before or seen them in records retrieved.

Pooled together, roughly the same amount of terms were used in thesaurus and free text searches. In the free text search, the ratio of thesaurus to free text terms was roughly 1:2 (8 terms which were also in the thesaurus versus 19 free text terms); in the thesaurus search, this relation was reversed to about 2:1 (24:10) (see appendix 5.3, "Thesaurus versus free text terms"). However, one person used one third (8) of these thesaurus terms in a single search statement (see column "Epsilon (thesaurus)" in appendix 5.2 – "Query terms and records retrieved"). Disregarding this case results in 16 thesaurus terms versus 8 free text terms, which is still a ratio of 2:1.

5.3.3.3 Diversity of terms

With the thesaurus search, more different terms are used (24 compared to 11 in the free text search), although most of these (18) were only used by one user each. The author had expected more inter-user consistency. As discussed later, users actually looked at different parts of the thesaurus, which explains the diversity. Examination of the most commonly chosen terms revealed that words from the scenario were used most often:

Distance learning	(by 8 -all- participants in 11 queries)
Multimedia	(by 5 participants in 8 queries)
Multimedia materials	(by 4 participants in 5 queries)
Development	(by 3 participants in 4 queries)
Web	(by 3 participants in 4 queries)

No information on why participants selected these terms is available. We can speculate that these terms might be the most obvious ones to use in this context, as the scenario description is the only real source of information and sticking close to it might have been considered the best approach in solving the task. Participants might not have considered it necessary to use other terms, they might not have realised that other terms were relevant or were not genuinely trying to find appropriate terms to represent concepts relevant to the scenario. However, the question why they still included so many terms with so little agreement despite these explanations, which would favour high consistency, cannot be answered with the available data. It could be due to slightly different interpretations of the scenario, and by more available options presented through the thesaurus. This situation highlights the needs for more qualitative data which provides information on individual sessions.

5.3.3.4 Thesaurus traversals and navigation

In order to make statements on thesaurus navigation, the data need to be formatted in a way that data from individual participants could be compared. After identifying thesaurus and free text terms, the structure of the thesaurus around the terms used was established in order to assess the traversals (see Appendix 5.4). For this, thesaurus relationships were identified (i.e. through NT, RT and BT relationships).

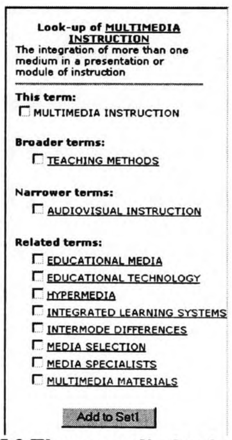


Fig 5.3 Thesaurus display for "Multimedia instruction"

The display of term details allows users to select associated terms (see Broader, Narrower and Related Terms in figure 5.3). Users can also select terms directly from the list of terms suggested by the system. A traversal was defined as the move from a term to either a Narrower, Broader or Related term. Moving from a free text term to a thesaurus term was not considered as a traversal.

According to the descriptions provided by the participants, they did not seem to navigate much in order to find terms. Participants often chose to display one of the suggested terms and selected NTs, BTs or RTs from here without viewing detailed information on the term first. Appendix 5.4 shows the five branches of the thesaurus from which terms were selected. Out of eight people, only two people (Theta and Kappa) seemed to really have navigated the thesaurus because their query terms came from a lower level than the one visible when entering the terms they claimed to have entered into the wizard (see Appendix 5.4, cluster 4 and 5). Kappa's navigation might have been influenced by the nature of one of the links though. The first thesaurus term was deprecated, i.e. no longer used for indexing, and directed the user to the currently used one (see cluster 3). All other participants navigated only across one traversal, from a suggested term to the record of that term where they were able to select NTs, BTs and RTs. One person seemed to give up straight away after not finding the right term, which might have been due to the fact that motivation can be low in a study like this. There might have been more navigation, but if so, participants must have decided that terms on the initial level suited their requirements better and did not provide the evidence in the account of their interactions. In subsequent studies this data would need to be captured in detail as this area is of particular interest.

5.3.3.5 Concepts

When the 19 thesaurus terms used are grouped by thesaurus branch, it becomes obvious that they come from 5 different parts of the thesaurus, namely, "distance education", "multimedia instruction", "computer-based learning", "Internet" and "learning". The fact that somebody used many terms does not mean they actually looked at many different concepts or navigated much. Three participants only navigated a single facet of the thesaurus, three others two and one person three. This is also the participant who made the most traversals (Theta). The amount of thesaurus terms in the query however seems to be related neither to the number of branches nor traversals.

Superficially, the concepts used might appear very similar. Four of them refer to education. However, ERIC is an educational database with particular focus on new media, which makes it necessary to distinguish between these different types in order to allow meaningful retrieval of documents. For the casual user, the differences might appear negligible.

5.3.3.6 Relevance of the result records

A record is potentially more relevant than others if it has been selected by several users (Ingwersen 1996). However, out of the 52 records selected from the result sets of the thesaurus searches, there were only two duplicates. Maybe this had to be expected considering the wide range of different terms and concepts used in combination with the exact match function. An investigation of the descriptors revealed that most records (30) were retrieved by only one thesaurus term each. 20 were retrieved by two thesaurus terms, and two by three terms each. Again, this is not too surprising, considering that overall, an average of 2.29 descriptors[5] or 3.42 query terms were used and the probability to retrieve a record with two or more terms must be limited if only two or three terms were used in the first place. Moreover, these records are only a subset of the retrieved result sets. It would be interesting to compare the number of terms retrieving selected and non-selected items – some of the later ones might have been retrieved by more terms. An indication would have been the position the record had in the complete results list, but the saved files do not contain this information.

A more reliable, but more time-intensive, way of assessing the relevance of the selected records in relation to the scenario is reading through the abstracts. The records were sorted by their ERIC identifier to avoid a bias towards abstracts retrieved by the same participant. Values between 1 and 5 were assigned based on the abstracts, 1 representing "not really relevant" and 5 "very relevant". The interpretation of the scenario seems to have influenced the participants' judgement of records. The author expected more general overviews or instructional papers while several of the abstracts were about interesting case studies. Collections of papers such as conference proceedings, where the record listed the individual paper titles, were judged based on the titles. Some items did not even contain this information and seven records were thus ignored for the relevance assessment. Relevance for records retrieved was then analysed for each participant. The chart in figure 5.4 shows the average relevance for each participant and table 5.2 the amount of abstracts of different degrees of relevance found by each of them.

[5] Ignoring the exceptional case where a person used 10 query terms.

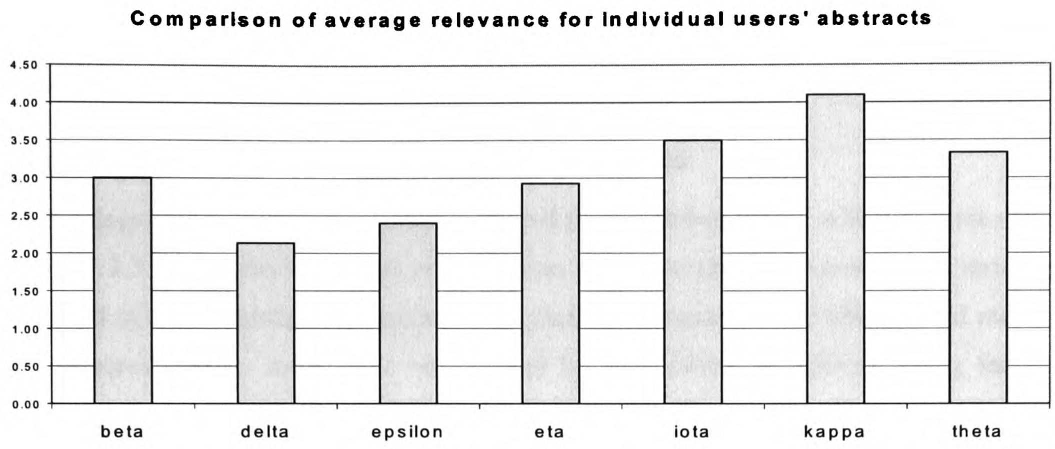


Figure 5.4 Average relevance of abstracts for each participant by degree of relevance

71% (32 out of 45) of abstracts are rated at 3 and above, i.e. they were at least reasonably relevant. Next, the average relevance per document was calculated for each participant, and four people out of seven reached the average 3 or above out of 5. The average relevance of abstracts for the remaining two participants reached 2.28, 2.4 and 2.93, which explain how the average still worked out as 3. Only one person achieved an average of more than 4, i.e. the suggested records were highly relevant.

Assigned relevance	1	1.5	2	2.5	3	3.5	4	4.5	5	Number of abstracts	"Total" relevance score	Average relevance
Delta	3	-	2	-	2	1	1	-	-	9	20.5	**2.28**
Epsilon	1	-	1	-	3	-	-	-	-	5	12	**2.40**
Eta	-	2	-	-	2	1	2	-	-	7	20.5	**2.93**
Beta	1	-	-	-	1	-	-	-	1	3	9	**3.00**
Theta	1	1	1	-	1	1	6	-	1	12	40	**3.33**
Iota	-	-	-	-	2	-	2	-	-	4	14	**3.50**
Kappa	-	-	-	-	-	1	2	2	-	5	20.5	**4.10**
Totals	**6**	**3**	**4**	**0**	**11**	**4**	**13**	**2**	**2**	**45**	**136.5**	**3.03**

Table 5.2 Relevance of abstracts retrieved, judged on a scale from 1 to 5 using half points.

In the next step, results for the relevance of documents and the counts of free text and thesaurus terms were investigated for patterns. The number of both thesaurus and free text terms does not seem to be related to the relevance ratings. However, Kappa made three transitions across links between terms, and Iota, Kappa and Theta navigated more than one branch of the thesaurus. These are also the participants who achieved the highest average relevance. Epsilon navigated two branches and employed the most thesaurus terms, which seem to be reasonable activities to retrieve relevant results, but the average is relatively low. The number of direct hits of query terms in descriptors also seems to play a role. It appears to be important to select documents which have been retrieved by several thesaurus terms, and use thesaurus terms representing several concepts in order to achieve more relevant results. This is logical considering that the scenario required documents to provide an overview. These results cannot be generalised due to the limitations of the study such as the number of participants.

5.3.4 Findings and potential research questions

The findings in this study are primarily related to the methodology, which is reviewed in section 5.3.5. Questions for future investigation were also raised. Overall, some terms are contained in the scenario, and starting from these, participants were able to find more in the thesaurus. Others must have been found by association, effectively using familiar synonyms, for example moving from "web" to "Internet" when "web" was not found in the thesaurus. The process of moving from free text terms to controlled terms is one area of interest. There is for example a discrepancy between these two sets of terms. Potentially, participants might have found terms they considered more suitable, or their free text term did not exist in the thesaurus. It would be interesting to know how they dealt with this situation. An associated question would be why there is such a diversity of terms, and why this increased with thesaurus use. The thesaurus is after all considered as a tool which can help to converge on a descriptor for a concept. The effect it has here is that of a generator for terms, which can also be beneficial.

The diversity of records selected raises issues on why there is so little overlap. Despite the controlled set-up of the study, only one record was selected by two participants, all other were selected by only one person each. The diversity of terms is likely to be partially responsible for this, but it appears that other factors, for example the way in which participants judge the relevance of records also comes into play.

5.3.5 Review of methodology

From the discussion of findings and research questions, it becomes clear that the data collection methods employed here were not appropriate to conduct in-depth analysis of thesaurus use behaviour. Nevertheless, the data yielded opportunities for several different approaches for analysis. The data revealed which parts of the thesaurus participants used, how relevant their selected records were and permitted investigation of how the use of the thesaurus influenced their query construction in terms of selection of free text and thesaurus terms.

5.3.5.1 Task sheets and qualitative data

The data were limited in the way that the task sheets did not collect qualitative information from the participants. The task sheet did not even include the "Do you have any comments?"- question, commonly asked at the end of questionnaires. The assumption was that participants were unlikely to report their own shortcomings. Even if problems were reported though, it would still be difficult to judge to what extent and how they affected participants' abilities to continue their search. This emphasises the need for more detailed data collection methods, especially on how users interact with the system. Measures for the effort made or the level of interaction with the thesaurus could be interesting in respect to assessing occurring problems and usability. Time limitation, i.e. the length of the tutorial (one hour), had required short task sheets, which in turn limited the data which could be collected.

Any investigation on reasoning for selection of thesaurus terms and for example attempts to explain the diversity of terms selected made in the context of this study are based on speculation. Qualitative information is thus of particular interest for the analysis and future studies would need to be set up accordingly. Some interesting answers can still be found by analysing the data though, for example whether or not users navigated the thesaurus and how many links they traversed. This however does not allow many conclusions on the participants' motivation, and thus does not advance our understanding of thesaurus use behaviour.

5.3.5.2 Practical aspects

It might seem trivial but it appears to be absolutely essential to tell participants to read the task sheets. Overall, the introduction and instruction on the interface seem to have been appropriate to the group of participants as reported problems refer to finding terms in the thesaurus rather than the use of the interface or thesaurus.

It was worth considering the wording for the scenario in respect to the terms used to describe the topic. If only words which correspond to thesaurus terms had been used, subjects would probably have used those in the first query already and the difference between the two query types would have been smaller. The results retrieved would potentially have been the same, as thesaurus terms are used for indexing, but are not distinguished from free text terms by the retrieval mechanism. In other systems, for example Ovid INSPEC or Medline, thesaurus terms are marked as such, and here, result sets would differ. Research questions have to be formulated bearing this in mind.

When employing user-reporting data collection methods, it might be helpful to agree some shorthand rules for participants to describe the search process. Their statements were sometimes open to interpretation. It might be best to change the layout of the task sheet accordingly, for example include a table for participants to enter terms and to prompt them to do this.

5.4 School of Care Science using Medline

This study was conducted with the help of four research assistants of the School of Care Science (SOCS) at the University of Glamorgan. After conducting a study with a group of MSc students who were given a set task to perform in a system they did not normally use (see section 5.3), it seemed beneficial to work with users in a more natural situation. The Medical subject headings (MeSH) are available in different interfaces, staff from the School of Care Science were contacted, four of whom kindly agreed to participate in this study. At the time of the study, they were all employed as research assistants. The primary aim of the study was to assess additional data collection methods and to investigate issues in the evaluation of information searching behaviour with a particular focus on thesaurus navigation.

5.4.1 Study set-up

The study was conducted at the work place of the participants and some meetings and sessions took place in more open environments, which resulted in a certain amount of disturbance.

Five meetings were scheduled with every participant over a period of two and a half months, some in groups and some on an individual basis. In an initial meeting, the purpose and process of the study were explained and interfaces and data collections the participants were using were identified. An initial search session took place to gain a first impression of the users' interactions with the interface. Participants also related some information about their experience with the different interfaces. Afterwards, a thesaurus tutorial was held on the basics of thesaurus use. The brevity of the study made a comparison of search behaviour before and after training inappropriate. The main search session was the primary data collection session. The discourse between the author and participants was audio taped and a record sheet was used to trace the participants' activities in the interface. These sessions were conducted individually over a period of two weeks. In the meantime, the author analysed the data and decided to introduce a set task alongside the participants' own tasks. These latter sessions are referred to as Set B, those in which participants only completed their own task as Set A. Use of a set task proved to be useful for the data analysis because it made it easier to compare sessions.

5.4.1.1 Participants' background

At the time of the study, all four participants were research assistants at the School of Care Science. Through their work, all of them had experience with computers and online searching. It is the nature of their work that they monitor the literature in their field, so that most of their searching took place to this effect. The searching background questionnaire (Appendix 5.6) provided some information on the participants' searching behaviour and the search training they had undergone. They all searched for information at least on a monthly basis, two searched on a weekly basis. Their major sources were several specific nursing databases, but two participants stated they also used references in papers. The Internet was not an important source of information, although it was mentioned by one person. Other sources include general resources such as newspapers, and more subject-specific materials, such as government reports. Three out of four of the participants had taken part in search training sessions before this study.

At the start of the main search session, participants completed a second questionnaire on the topic and background of the search. This questionnaire revealed that the participants were at various stages of their projects, and 3 out of the 4 were interested in a broad search, two for general information gathering, and one for a literature survey. The fourth person also wanted to gather general information, but only required a few references. Two participants mentioned quality of the publication as a requirement for their search for any items found. One person specified that the items should not be older than 5 years and in English.

	User 1 - Set A	**User 2 - Set B**	**User 3 - Set A**	**User 4 - Set B**
Project time	2 months	2 years	Varies	4 years
Searching	Weekly	Monthly	Monthly	Weekly
Databases used	Medline Cinahl Indiv. Journals	Cinahl Nursing collection Medline Embase	Cinahl Embase Medline Nursing collection Health promis	Medline Cinahl Nursing collection Nesli
Other sources	References from supervisor	Reports (to less extent newspapers, conferences, govt. sources)	Internet: - Journals - Ask jeeves Internal publications Libraries (e.g. HPW)	References in reports, papers and conferences
Searching workshop?	No	Library introduction about 2 years before	About 2 years ago	Several years before whilst studying 1 year before with a consultant

Table 5.4 Background data collected through questionnaires

5.4.1.2 Training of participants

The training took place in two stages. Through this arrangement, it was attempted to ensure that participants had at least a certain level of knowledge about the thesaurus without confusing them too much with additional information on online searching.

Pre-search training

The first search session was very brief and provided an opportunity to see the interface in use and to establish the level of participants' knowledge of online searching. The pre-search training was then designed accordingly. It aimed at very basic search strategies such as Boolean AND and OR as used by the Ovid interface. It also provided a basic introduction to thesauri in general and the MeSH subject headings in the Ovid Medline interface in particular. Since it was difficult to co-ordinate a meeting time, the training was repeated three times. The handout was however very detailed, so that differences between training sessions were minimal.

Post- search training session (together with feedback)

All participants were present for this session. This training session was conducted in return for the participants' time and effort they had invested in the study. It included some more advanced searching, for example explanations of different search strategies such as block building[6], successive fraction[7] and pearl growing[8]. Information on advanced operators such as wildcards and adjacency operators was also included. Detailed handouts were written specifically for the participants of this study for them to take away. At the time, the analysis had not been completed, but initial outcomes could be reported.

5.4.1.3 Information on the Ovid Medline interface

The Ovid Medline interface is one of many that access the U.S. National Library of Medicine's abstracts collection which covers a wide range of medical and health care topics. The Ovid interface uses the Medical Subject Headings, which is a medical classification system developed by the U.S. National Library of Medicine. This classification is widely used, as is the Medline database. One of the differences to most hierarchical classifications is that some terms can be qualified by using subheadings which are displayed where applicable. These effectively represent additional facets which allow a more precise definition of the scope of the required concept. A term can for example be searched for in the context of "genetics" or "drug effects". For example the term "Blood" has subheadings such as "diagnostic use" or "Blood supply", which describe two different perspectives on blood (fig. 5.5, example from the Ovid Medline Help). In order to demonstrate the difference between Narrower Terms and subheadings, figure 5.6 shows the term "Blood" in its hierarchical context. The term has only two Narrower Terms, "Fetal blood" and "Plasma".

---

[6] Queries are first executed for each concept individually whereas synonyms and variants are ORed. The queries are then combined using Boolean logic.
[7] The set of results is reduced by adding more and more concepts to the query.
[8] Information from a relevant document (e.g. author names, keywords) is used for a new query.

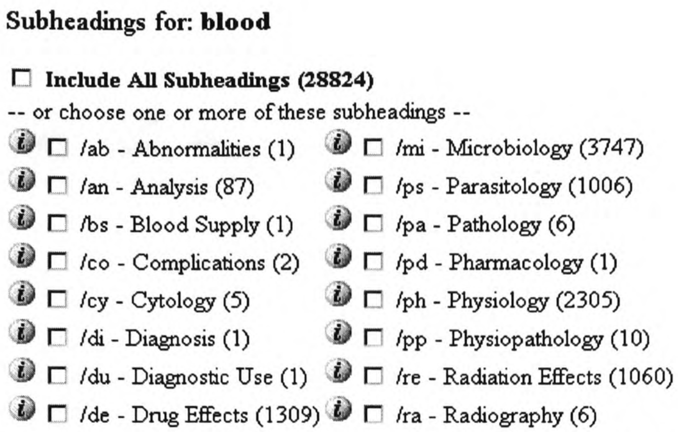


Figure 5.5 Subheadings for the subject heading "Blood"

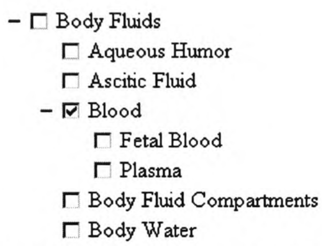


Figure 5.6 The subject heading "Blood" in its hierarchical context

The main search page on Ovid (fig. 5.7) contains a simple text box into which users type their search terms. By default, the check box to "Map term to Subject Heading" is selected, so the system searches the Subject Headings for appropriate terms that correspond to the free text term the user entered. These potential matches are in the form of hyperlinks and the users can browse the terms in their local context in the hierarchy of the controlled vocabulary. By checking boxes, they can select terms and run the query by pressing the "Continue" button. The search mechanism uses exact match Boolean logic. The interface contains a number of options, for example it is possible to combine queries using a specially designed page or typing in a Boolean statement using the query numbers into the search text box. All options to proceed through the interface are outlined in a flow chart in appendix 5.7.

Some of the tools of particular interest to this study are "Focus" and "Explode" (figure 5.8). When the user checks the "Explode" box for a term, this simply means that all Narrower Terms are also included in the search. If "Focus" is selected, the term concerned should be a main topic in the retrieved documents; documents are indexed accordingly.

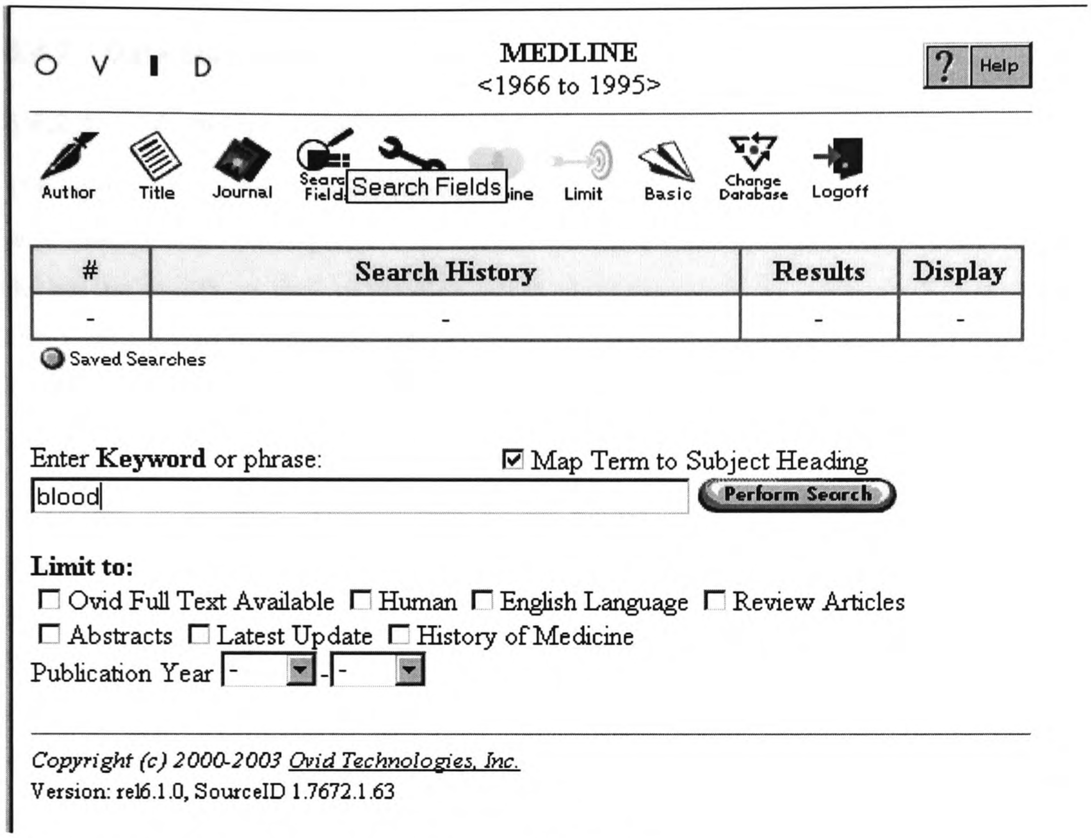


Figure 5.7 The Ovid interface to the Medline collection

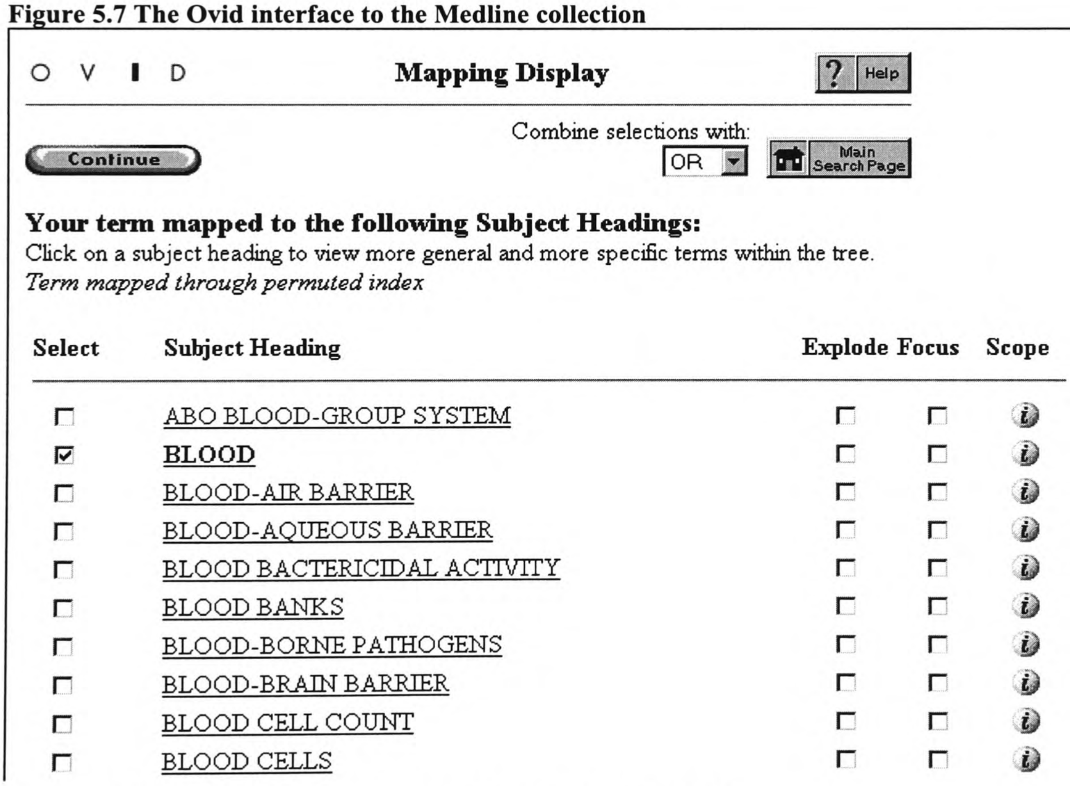


Figure 5.8 Excerpt of potential matches for the phrase "Blood"

5.4.2 Data collection

5.4.2.1 Audio recording

The conversations between participants and the author were recorded on audio tape and later transcribed. Participants were asked to voice their thoughts and considerations during the search so their motivation while searching could be understood. The author also prompted them and asked specific questions during the search to clarify information and to encourage participants to think aloud.

5.4.2.2 Prompts

A number of prompts and questions to ask before, during and after the main search session were prepared. These questions were related to thesaurus terms, abstracts participants viewed and their relevance. The prompts also covered interactions and incidents of particular interest during the search session, such as turning points in the search. These prompts were meant to remind the author of important aspects during the session, and also to introduce a higher consistency in the data collected.

5.4.2.3 Search questions forms

The author decided against extensive use of questionnaires, but prepared a set of questions regarding background information on the participants' projects, for example the stage they were at, use of information searched for, for example for publications or further research, and limitations, e.g. in particular languages or the publication date (see Appendices 5.5 and 5.6). These handout were based on examples of search request forms presented in the literature (e.g. Shaw 1986). One was handed out during the initial search sessions, the other one was completed immediately before the main search session. The data collected provides a backdrop to the participants' searches and clarifies the requirements and constraints on the main search session. The responses are presented in table 5.5.

	User 1 - Set A	**User 2 - Set B**	**User 3 - Set A**	**User 4 - Set B**
Project stage	Information gathering for grant application	Orientation Monitoring	Clarifying areas within a defined project	About to come to a conclusion
Use of results	Information gathering for grant application	General interest	Research paper	Speech, research paper and thesis
Purpose of search	Information gathering	Information gathering	Detailed literature survey	Information gathering
Search requirements	Comprehensive search	Comprehensive search	Comprehensive search	Narrow search
Search limitation	5 years old max. English language	-	Quality of papers, i.e. peer-reviewed journals	Quality of papers

Table 5.5 Responses to the pre-search questionnaire

5.4.2.4 Record sheet

As navigation through the interface and the thesaurus in particular were of interest, it was felt that they should be recorded as exactly as possible. No appropriate tool was available to record interactions, so this had to be done manually. After studying possible paths through the interface and drawing up a flow chart (Appendix 5.7), a table of A4 size was constructed (excerpt in Appendix 5.8). The rows represented different pages, buttons (shaded grey) and other interaction options. They were ordered in a fashion that would make it easier to quickly tick off the appropriate cells from the right to the left while the participants moved through the interface. For the same purpose, the descriptions of items are repeated. The columns represented the interactions sequence. It was unfortunately not possible to record the time that passed between interactions, but it was hoped that this information might be extracted to some extent from the key log text generated by the Keylogger (see below). Thesaurus navigation was to be recorded on a separate sheet by writing down the terms. Relationships between them would have been established after the sessions. However, this manual logging method proved too slow and was abandoned. Participants did not navigate the thesaurus much in any case.

5.4.2.5 Keylogger

The Keylogger is a small, public domain Visual Basic program which does not require installation. At the beginning of the sessions, it was copied onto participants' computers and started after they logged into the database. It wrote key strokes and mouse-clicks with time stamps into a text file which had to be copied back to floppy disk. The program was deleted at the end of each session. This tool allowed the author to concentrate better on the events on the screen, as it was not necessary to write down search terms and queries. Figure 5.9 shows an excerpt from the Keylogger log demonstrating input of the term "substance abuse". As the interface is mainly mouse-driven, large parts of the log only contain mouse-clicks with time stamps.

```
14:55:45
su
14:55:46
bst
14:55:47
ance a
14:55:48
bus
14:55:49
e
14:55:50
{Left Mouse}
```

Figure 5.9 Excerpt from Keylogger log

5.4.2.6 Set tasks

After the interactions for the first two search sessions which were conducted on the same day proved to be very similar to those observed during the initial search sessions, it was decided to give participants two set tasks for choice. This was because participants executed searches they were familiar with, so that the sessions did not add to the data already observed in the initial sessions. In particular, participants did not need to navigate the thesaurus to locate the most suitable terms. As this behaviour was of particular interest, the set tasks of general interest were designed to potentially require some navigation of the Subject Headings. Both participants selected the same task which asked them to imagine they were a parent looking up information on the influence of the social environment on substance abuse in adolescents. The first two sessions are referred to as A (2 searches), those which consist of the participants' task and the set task as B (4 searches).

5.4.3 Discussion of search sessions

5.4.3.1 Initial search sessions

All participants retrieved records in these sessions which they would consider as relevant to their topic. This was to be expected as they were demonstrating how they normally searched. Below, their search behaviour is described in combination with concerns they expressed.

User 1 (Set A)

This participant normally used a different interface without subject headings browser. Restrictions were easier to impose and the search overall was quicker to complete. Presumably, the data collection is not identical, or the matching algorithm works differently from that in the Ovid system as the participant states that references found in that particular system were not retrieved when using the Ovid interface with the Medline collection. When re-running the search, it was discovered that "exploding" terms, i.e. including Narrower Terms in the search, retrieved a larger amount of documents, but it was not possible to judge the quality of the documents in this result set or to make an informed comparison between the sets retrieved by the two systems. In the initial search session, User 1 only selected the required terms, and did not navigate subheadings.

User 2 (Set B)

This participant normally used the Ovid interface to the Medline collection. The general nature of her subject area lead to a large amount of records being retrieved which caused her problems. The participant was very concerned not to loose relevant items, and thus did not set any search limitations other than language. She used neither the "Focus" option nor the subheadings although they were available for her terms. During the initial search session, she did not navigate the subject headings. The participant noticed that documents which appeared in other databases were not found in this collection.

User 3 (Set A)

This participant used mainly the Ovid interface, but with a different database, the Nursing Collection. She tried her queries also on different databases. Within the Ovid interface, queries can be re-executed automatically, but as she normally started with the Nursing Collection, which does not use the MeSH, she would not use the thesaurus for Medline either. She also found subject headings were not understandable enough to use and that she normally ended up with too many results.

User 4 (Set B)

This participant normally searched the Medline collection through the Ovid interface. She was unsure about some functionalities, such as "Explode" and "Focus" as well as Boolean operators. During the initial search session, she selected specific subject subheadings for her search, which she had already used previously and thus knew about. She had arranged her search terms into different categories, which is interesting because it could suggest an awareness of facets or more systematic information searching (e.g. the construction of hedges, that is combining synonyms and variants with Boolean OR).

5.4.3.2 Main search sessions

Due to the interface design, all participants started their searches in the same way, entering a term which was then by default mapped to the subject headings. They then selected a term and executed the query. From this point onwards, the searches differed. Some participants scrolled the list of results and looked at abstracts, others first searched for more terms and combined queries. One person also applied limitations to the document language, another to the year of publication. Search histories revealed that combining queries, i.e. use of Boolean AND or OR, which contain only one term is the main search activity. According to three out of four search histories, participants submitted around 13 queries each, in the fourth, only five. The record sheets for two other tasks indicate few "combine" interactions. One possible reason for this is that these were repeated queries, where participants already knew exactly which terms to use, how to combine them and which limitations to apply. Transcripts revealed the importance of search outcome for reformulation. With large results sets, they might reformulate, but if the set is quite manageable, they might just look through all results despite of the effort involved. Overall, participants did not appear to be interested in refining query terms, and also felt that subheadings were awkward or inappropriate to use. Thus, when they wanted to limit a results set, they were likely to do this by limiting the year of publication, not by modification of concepts.

5.4.4 Data analysis

5.4.4.1 Graphs of search patterns

For three sessions, data on the participants' moves through the interface were recorded using record sheets, but the author got distracted by the users' interactions in the fourth session so that no record sheet exists here. Unfortunately, this limits the options to compare the two searches on the set task.

The ticks on the records sheets (example Appendix 8.5) were converted into numerical values. Values 1 to 5 represented auxiliary tasks (such as logging on and changing the database), 6 to 17 stand for searching activities (such as clicking the basic search button or selecting an option), 18 to 26 represent choice of tools (such as the thesaurus or permutated index), and 27 to 35 refer to dealing with results (for example displaying a full record). From these values, graphs were constructed and interactions by the users can be compared. A simple line graph seemed to be the display the easiest to read, even though the software did not allow for the y-axis values to be replaced by text explaining the assigned values.

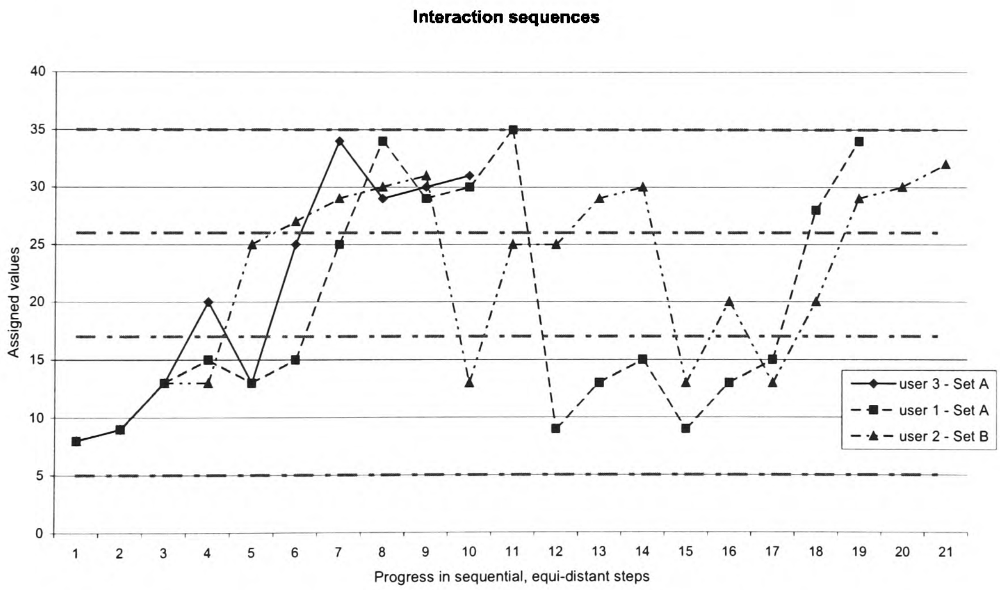


Figure 5.10 Search progress for users 1, 3 (own task) and 2 (set task)

The graphs move from "Search" via "Tools" to "Results". Other studies (e.g. Manglano et al. 1998; Spink and Saracevic 1997) looked at search cycles, going either until each print/display command, or up to a point when the user steps back in the sequence of search stages. Expressed in these terms, participants from set A in this study only executed one cycle for each question, ending in the "results" section[9]. Some details, for example where several limitations have been applied to the query simultaneously, such as the year of publication and the language, have not actually been captured in the record sheets. The distinction between set-up of a new query and a refinement of a previous one can only be made under consideration of the query history and does not become evident in the graphs. Effectively, a query is refined by selecting previous queries from the query history and combining them using Boolean operators with each other or new terms, or to impose limitations. To illustrate this, figure 5.11 shows an excerpt from one of the search histories. Four separate queries are submitted to the system, which together form one search on a particular topic. The value on the right hand side represents the number of records retrieved. Query 8 contains the term "Health education". Query 9 contains the term "Internet". In query 10, the searcher combines queries 8 and 9 using Boolean AND, i.e. searching for "Health education AND Internet". In query 11, the user then limits the results from query 10 to the years 1998 to 2000. Coincidentally, all records from query 10 are from this period, so that the number of retrieved results does not change.

Search	Strategy:	
8	Health education/	3170
9	Internet/	3007
10	8 and 9	64
11	limit 10 to yr=1998-2000	64

Figure 5.11 Excerpt from a query history

Although queries in this study were limited in regard to location and year (e.g. user 1), this is in fact not a modification of the previous query in the sense that this analysis is interested in. This participant's instructions actually specified a search per year of publication, so this modification was prescribed from the start. It did not arise from factors within the search session, or from changes in the searcher's Information Need. In another query, searching for individual countries of the UK after trying "UK", would have been a modification, but the participant entered them as part of the same search cycle (user 3). The modification does not become obvious in the graph because keywords are entered one after the next. In the second set of sessions (B), modification of queries for the set tasks took place (see Figure 5.10, user 2: steps 10, 15 and 17 indicate new terms being added, steps 11 and 12 combinations of queries after adding terms).

[9] User 1 limits the query to different years (steps 12-14; 15-17). User 3, step 4: the query is limited by year and language using drop down menus.

5.4.4.2 Sequence analysis

Sequence analysis can reveal patterns of behaviour. Here, it could be conducted with the values generated from the record sheets, but too few data are available to make this type of analysis worthwhile. First of all, only few sessions were conducted, and secondly, not much interaction took place in some of them as discussed above. Common sequences can however be identified from the graphs. Patterns which appear in all three search sessions are selecting the database (values 8 and 9, steps 1 and 2 in the graphs), the first search execution and the examining of results. Some variation can occur during search execution, but options are limited (see Flowchart in appendix 5.7). Sometimes, sequences occur naturally, for example clicking a particular button brings up a particular page. The sequences related to viewing results seem to be of this type. Differences that can be identified are related to the point of time when the "Combine" option (x=25) is selected, for example before and after interacting with results.

Searches in set B are more complex and, as mentioned above, it is possible to observe cycles where the searcher moves back to the main search page to enter more terms after viewing some abstracts. In a study on a larger scale, this might be of interest, but with the limited possibility for comparison, no particular conclusions can be drawn.

5.4.4.3 Content analysis of the transcript of tapes

The audiotapes of the sessions were transcribed and approximate information on the time scale included. Problems and comments on difficulties were extracted and the statements were then categorised using open coding, which highlighted a number of problem areas. Many of these were related to the database and database indexing rather than to thesaurus and interface use, but to the participants, who seemed to look at the use of the system as a whole, the origins of problems would not necessarily have a large importance.

5.4.5 Findings

One of the interests of this study was how the thesaurus informed searching. However, as participants executed familiar queries, the possibility to observe this behaviour was limited. Through the users' comments it was possible though to get a good impression of the influences on query formulation processes in general. This section thus discusses (after some general problems participants expressed) some of these influences and the query (re)formulation process as well as the assessment of records. The last section deals with use of free text versus thesaurus terms.

5.4.5.1 General problems

Overall, the Ovid interface to the Medline database is relatively easy to use especially if the person is familiar with the Internet and consequently with hyperlinks and buttons. However, participants found it awkward to set limitations for their search, for example the year of publication. They were also a little confused about some of the database-specific features such as "Explode" and "Focus". After executing a large number of queries, one participant got confused about combinations she had already tried. It was helpful to see the history of executed searches, but it got too long to be manageable. Although queries can be deleted, this did not seem to be the best option as the participant already re-executed previously run queries.

5.4.5.2 Influences on the search

The participants' role in their work place seemed to have a major influence on the way they searched. As research assistants, they were conducting literature searches to monitor the development of a field for other members of staff. Not "owning" their search also explains their recall orientation (rather than precision) and the little flexibility there seemed to be in query terms. The differences in search behaviour to the more open tasks, i.e. a search conducted for one of the participant's own projects and the set task (for sessions B), were striking. More reformulation took place on a conceptual level and participants tried out more options. The novelty of the set task also resulted in more thesaurus navigation as participants first had to identify the most appropriate terms, a process which had already taken place for the other tasks.

5.4.5.3 Query reformulation

The participants were willing to put a large amount of effort into the assessment of the results and thus were less interested in refining their query. They were more likely to refine their query using operational rather than conceptual moves, i.e. limiting by year of publication rather than modifying the concepts both to increase and decrease the set size. This together with the apparently limited flexibility regarding query terms could potentially mean that they miss out on opportunities to discover different terms and material. There seems to be some confusion as far as subheadings and Narrower Terms are concerned. The experience that subheadings were overall either not relevant or difficult to choose potentially led participants not even to consider Narrower Terms.

5.4.5.4 Relevance assessment of records

Despite the fact that two participants indicated the quality of a publication source as a criterion for retrieved documents, relevance judgements were based first on country of publication, and secondly on the content of the abstract. The abstract was more important for the relevance judgements than the subject headings used to index the records. If no abstract was available, participants judged items by title and by subject headings and stated that if in doubt, they would try to find the document or order it (User 2's session). Due to the nature of their work, participants investigated for every single abstract whether it related to the UK, before confirming the topic. With large sets, looking at every single article is a painstaking task, and not everybody might be willing to complete it conscientiously, but for the participants, this appeared to be the most reliable way of judging the records.

In the more open tasks executed in set B, participants skipped abstracts for which the abbreviated display already seemed to indicate irrelevance. This is to say, in the first sessions (set A), there seemed to be more hard and fast rules for relevance, while they were less well defined in the additional tasks for set B. In set A, it seems that all records retrieved can be considered relevant just because they were retrieved by a query which contains specific terms and is limited in a certain way, not because the abstract indicates the treatment of relevant research. In User 1's case this seemed to be particularly clear as she actually counted the documents retrieved for each year. Again, the purpose of the search clearly influenced behaviour: Because both searches were comprehensive and intended for review of the existing literature, even documents previously seen or read would have to be included in the final set, which was prepared for somebody else.

Some evaluation of documents normally takes place off-line, for example User 1 said she would save results and delete irrelevant ones from the file at a later point. This kind of behaviour might have consequences though for the design of interfaces used for successive searching, for example in terms of items which are considered to be relevant to the participant. Only a subset of those previously selected as relevant would actually be relevant. One observation during these studies also showed that a previously viewed item is not necessarily less relevant to the searchers as they might forget over time which items they have and have not viewed.

5.4.5.5 Free text term versus thesaurus term use

It was not possible to investigate thesaurus use in detail, as the data collection methods were simply not suitable to capture this behaviour in sufficient detail. Because participants executed mainly queries they were familiar with, thesaurus navigation was also limited as identification of the most suitable terms had taken place during earlier sessions. Participants were very recall-oriented and stated their suspicion towards the controlled vocabulary because use of thesaurus terms resulted in fewer records to be retrieved compared to queries with free text terms, probably because false hits were avoided. One user also stated that records seemed to be indexed incorrectly. The fact that users were willing to wade through even very large amounts of records rather than refining their queries illustrates an unwillingness to trust the database indexing.

It appears that at the beginning of the search, users selected relatively broad terms. The results then have a certain influence on the progress of the search. One participant explained that she used the "Explode" option if she felt that not enough records had been retrieved, another person stated that she avoided it because the terms she used resulted already in a large amount of results.

5.4.6 Review of methodology

5.4.6.1 Information Need and search context

Because the participants' own queries had already been executed in real-life situations previous to the search session, their Information Need did not change. They had already seen the information they required, which defied the idea that that people would be more motivated to execute a real task than a set one. The fact that the participants were in a study situation might also had an influence in that they might have tried to fulfil the author's expectations by demonstrating how they searched without realising that they were expected to actively search. Similarly, they referred to the selection of thesaurus terms in a way that implied navigation, although they did not actually navigate the thesaurus. The author noticed the discrepancies between the participants' expression of their understanding of what they are doing and their actions, but in other cases, this might be overlooked more easily.

Due to the nature of their searching, which required in the first place monitoring the literature, queries normally constitute re-running an "existing" query (even though it is not saved and reloaded). From this perspective, the repeating of the queries during the search sessions was realistic, although real searching should better be observed as and when it occurs. The analysis could have been conducted from the perspective of monitoring, but this was not a long-term research interest.

For the set task, some changes in Information Need had been anticipated. However, it seemed that, as the task was left very open, the participants were more interested in trying to define it better by asking the author. This seems the most reasonable thing to do, and might be natural for somebody who mainly searches for others, but is a potential disadvantage with set topics. One participant selected some subtopics for this subject, the other participant selected appropriate results without doing this. Defining tasks better would be a first step towards more independent searching and topic refinement on the participants' part. Generating an interest which would give them impetus to search is exactly the difficulty with set tasks, and as seen here, with the repetition of existing queries.

In subsequent studies, it might be most useful to prepare a scenario as well as observing the users' own queries. A query they are unfamiliar with requires them to actively think about thesaurus navigation, terms and query settings. It might be particularly useful to have a topic that they are unfamiliar with, so that they do not know the relevant facet of the thesaurus well. Due to the time span between the two sets of sessions, it was possible to develop a related topic of general interest which Users 2 and 4 completed. As a result, more thesaurus use could be observed than in the first sessions. Comparing the participants' own task to the set task, the set task was definitely helpful in encouraging thesaurus use. In these sessions, users even compared their interactions to their normal behaviour and stated that they did not use the thesaurus if they were familiar with terms, thus the approach revealed some useful information.

5.4.6.2 Logging

For the first set of sessions (set A), the Keylogger did not work because a DLL was missing on the machines. However, the log was not particularly useful for the data analysis anyway. An attempt was made to use the time stamps to get an idea of the pauses between interactions, but it was impossible to identify which mouse-click corresponded to which tick on the record sheet. All mouse-clicks were recorded, even if they did not trigger a process in the system and no coordinates could be recorded so it was impossible to tell where participants clicked. The only information that could be extracted were the terms participants had typed into the search text box, but this information was negligible for the analysis of the tasks.

The analysis conducted shows that a detailed record of users' interactions with the interface is required in order to conduct in-depth analysis not only of the types of interactions taken, but also in order to gain a better understanding of how users navigate the thesaurus and to allow for the discrepancy between the users' description of their behaviour and their actual interactions. For example in the case where participants claimed to use the thesaurus but did not actually navigate it, their view on what they did, did not correspond with how the author would describe their behaviour. From this, incorrect conclusions might be drawn if there is no way of verifying participants' statements.

5.4.6.3 Record sheet

Using the record sheet to capture participants' interactions was difficult because interactions followed quickly so that some might easily be missed, especially as the author was simultaneously following the logic of the search and sometimes talked with the participants. Thesaurus interactions are particularly difficult to record because terms were mainly selected, not even typed, and this process can potentially be very fast. Thus, a more reliable method, such as more complete logging, is required to record events during sessions, also because this would allow the author to concentrate better on the observation and discourse.

5.4.6.4 Questions and prompts

Although the list of prompts was available during the sessions, it was difficult to refer to them alongside observing the participants and filling in the record sheet, so some questions remained unasked. Some were not relevant because the specific context, for example thesaurus navigation, did not occur. On the other hand, the author also encountered situations which required prompts, especially as far as thinking aloud was concerned, but could not find the appropriate words. In some cases, participants described their actions rather than their reasoning behind them. Sometimes, unexpected incidents occurred, which were more important to follow up than the questions prepared and a researcher should be flexible enough to follow this development as it might reveal new areas of research.

Regarding selection of thesaurus terms, participants could have been probed more for information on how they had selected terms when they first started searching on a topic. However, in some cases, this had been long before the study, and participants probably had not deliberately reflected on this process so that they were not likely to remember it.

It might be advantageous to discuss specific incidents regarding thesaurus use directly after the sessions with the participants. A discussion was attempted in the last training session, but turned out to be difficult.

5.4.6.5 Search question form

The search question forms provided some uniform data on the participants' project context. Some common problems with questionnaires were encountered in that users interpreted certain formulations, for example regarding broad and narrow topics differently, and that the forms did not correspond sufficiently to their experience of information searching, which appears not to be expressed in quite the same terms as common models showing the stages of Information Need.

5.4.6.6 Initial search session

Although not deliberately intended to achieve this, the repeated meetings and initial search session might have functioned as an ice breaker so that participants might have been more comfortable with the author and might have been less concerned about the study.

5.4.6.7 Audio recording

Despite some issues regarding background noise and the quality of the audio recordings, they are a good source of information, and quite an exact record of how the search progressed. As the conversation related to what happened on screen, it also helped the author to remember difficult situations the participants had encountered. It was possible to distinguish between records in the recording, because the author deliberately referred to them as "the first", "the second" etc., with this in mind. From the search history, it is then possible to identify records, even if only the selected ones were retained. However, updates in the database mean that other records will be retrieved when re-running a query too long after the search session.

5.5 Conclusions on the methodology from both studies

The studies revealed weaknesses and limitations of some of the available methods. Information on records retrieved by a particular search was not persistent, i.e. because databases are updated periodically, so any necessary information should be captured as soon as possible after the sessions. One of the main issues though was that data was in parts not detailed enough. The Keylogger only recorded keyboard input and mouse-clicks without any information other than the time stamp. This tool is useful for a command-line interface, but not for a mouse-driven one. However, the potential in automated logging becomes apparent when comparing it with manual recording techniques. These are simply too slow and unreliable and they distracted the author too much from the general progress of the session. When participants are to record their steps themselves, it is impossible to know how conscientious they are about it.

Planning the data analysis beforehand can ensure that data is collected at the right level of detail for the intended analysis. Prepared prompts and questions were useful, even if it is impossible to be prepared for every eventuality. Having a general idea of the issues of interest can help the author to keep an open mind but also assess the significance of unexpected events and potentially follow them up on the spot.

Some statistical methods were investigated, as they are generally accepted and applied analysis techniques. However, the author found it difficult to extract meaning from the results. The qualitative data collection methods yielded more information. They also made it possible to identify the motivation behind interactions which would otherwise have remained hidden. Participants could also comment on issues that they have encountered in previous searches. However, some points need to be verified with quantitative data, such as a record of interactions, for example where participants claimed to use the thesaurus but did not actually navigate it. Therefore, it is important not to rely on one single data collection technique. Conversing with participants during the sessions also allowed the author to clarify some issues, for example their understanding of the short questionnaires or set tasks. This also provides the opportunity to ensure that users do what they have been asked to.

It was observed that both the scenario and tasks with real Information Need had advantages and drawbacks. For some participants completing the set task, motivation might have been low and they might have found it difficult to reformulate as the framework within which they made their judgements did not support this well. The scenario or the instructions might not have been detailed enough. However, the participants in a real information searching situation did not need to navigate the thesaurus as they had already identified the most suitable terms.

5.6 Conclusions

The two studies reported in this chapter were conducted as exercises in research methodology and data analysis. Not only have they revealed a number of practical and theoretical issues related to quantitative and qualitative research methods, but the data has also opened a window on the opportunities for research when looking at thesaurus use in information searching.

The SOCS study resulted in more findings than the ERIC study due to the qualitative data collected. Some findings correspond to those reported in the literature, for example attitudes to free text searching. Users' misconceptions about functionalities they used were revealed, and these appear to influence the way in which they use them, for example use of subheadings. In the SOCS study, the influence of task context on search behaviour became very clear due to the imbalance of effort spent on query reformulation compared to the assessment of records. The most noticeable influence here was the position of the person executing the search. If they work for or closely with somebody else, they were less flexible in modifying the query according to the results, because their task description can be very specific. Their area of research might also impose specific restrictions, for example in research where policies are important, only articles relating to a specific country are relevant. On the other hand, when somebody searches out of general interest or maybe to explore a new avenue of research, they are more flexible in modifying concepts and trying out different query formulations. Other influences on searching behaviour are the recall-orientation linked to the task setting, which resulted in reluctance to use narrower, more specific terms and a certain mistrust in the indexing, the thesaurus and the retrieval mechanism which lead people to prefer free text query terms. Further investigation of these and similar influences might allow the construction of a more complete understanding of searching behaviour.

In the ERIC study, an influence of the thesaurus on query terms has been detected. With the thesaurus, participants used more diverse query terms than in the non-thesaurus search. This variance possibly resulted in a larger diversity of records retrieved so that results submitted showed almost no overlap. This highlights the thesaurus role as a tool for generating new terms, although it is also thought to introduce consistency into retrieval. Further investigations might focus on these different roles of the thesaurus and its general influence on searching. This requires a slightly broader approach which does not simply concentrate on interactions with the thesaurus, but to investigate its role in information searching behaviour, which has been taken up in the studies described in the following chapters.

Findings related to the methodology have been discussed in the previous section. Alongside these, the author also gained valuable experience in setting up studies and especially in the interaction with participants which was beneficial for the following studies.

5.7 Urls for chapter 5

ERIC: http://searcheric.org/

ERIC wizard: http://searcheric.org/scripts/ewiz/amain5.asp

OVID for Medline (registered users only): http://gateway2.uk.ovid.com/

Chapter 6 - FACET 1 study

6.1 Introduction

The study discussed in this chapter moves from working with proprietary online interfaces to FACET, an experimental system using semantic expansion and faceted query matching, developed at the University of Glamorgan. This system was designed specifically to explore the use of thesauri as tools in information searching and participants were drawn from available volunteers at the University and collaborating museums. The version of FACET used for this study is version 1, and the study resulted in significant changes in the interface (see chapter 7). The data collected was also used to draw up a preliminary framework for the model of information searching in controlled vocabulary enhanced systems discussed in chapter 10 and 11.

6.2 Objectives

During the previous studies, the focus shifted from the investigation of thesaurus navigation to use of the thesaurus as a tool. In the FACET 1 interface, the thesaurus has been integrated in two specific ways. Firstly, the interface provides easy access to the thesaurus, and secondly, it uses query expansion which has a number of functions, but particularly helps to avoid "zero hits" by including additional terms in the query based on the semantic relationships of the thesaurus. The FACET interface requires queries to be constructed in a faceted structure, i.e. concepts have to be broken down into separate terms which might stem from different parts of the thesaurus, which can potentially lead to difficulties and requires skills quite different from those needed in many non-specialist interfaces, for example on the web. The study was interested in how users deal with a thesaurus in this context and assessing potential implications for this type of system.

6.3 FACET system and interface

The EPSRC-funded FACET (Faceted Access to Cultural hEritage Terminology) is an experimental system developed at the University of Glamorgan in collaboration with the National Museum of Science and Industry (NMSI) which includes the National Railway Museum. It follows on from early work applying a semantic database approach to hypermedia and information retrieval (Beynon-Davies et al. 1994). The FACET project investigates the possibilities of semantic expansion in faceted thesauri based on measures of semantic closeness. With a view to making their collections more accessible to the public, NMSI is indexing parts of its collections database using the Art and Architecture Thesaurus (AAT) (see section 2.2). An export of these partially-indexed records (approximately 400,000 records) is used as the underlying data set for the system.

Initially, the interface accessed the Railways and Timepieces collections, with other collections added for the next study. The system is implemented in C++ with a Visual Basic interface and an SQL Server database. An in-memory representation of the network of thesaurus relationships permits an efficient semantic expansion algorithm (refer to Tudhope et al. 2002 for details on the system architecture).

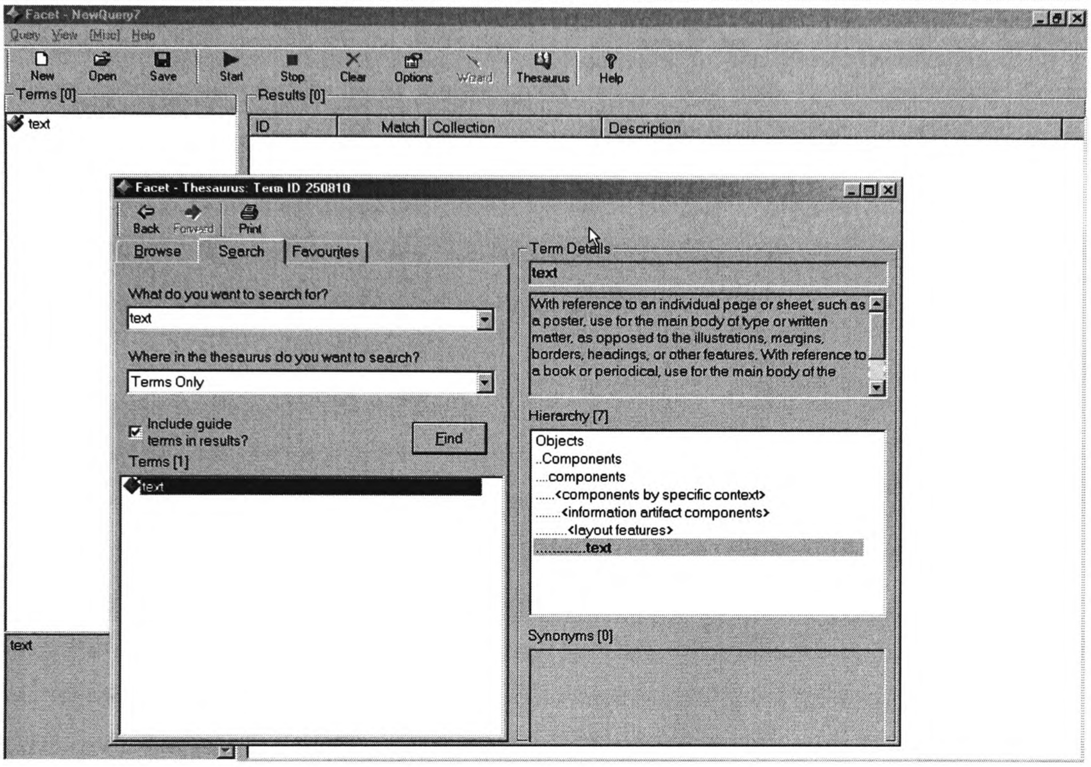


Figure 6.1 FACET 1.1 Thesaurus window showing Term Finder

Figures 6.1 and 6.2 illustrate version 1.1 of the FACET interface. Figure 6.1 shows in the background a query with one query term on the left ("text") and on the right the area where results will be displayed. In the foreground is the window which allows users to access the thesaurus. This figure shows the string search facility called Term Finder that attempts to map an initial free text string to controlled vocabulary terms, which are required to construct a query. This window is brought up by clicking the thesaurus button on the menu bar on the main form. On the "Search" tab, users can enter terms into a text box and select whether they want to search thesaurus terms, scope note or both. By un-checking the checkbox below, the user can select not to return candidate terms which are guide terms (for guide terms, refer to section 2.3.1, towards the end). When the user clicks "Find", the system searches the thesaurus according to the criteria specified and displays the candidate terms. In the figure, one term, "text", has been retrieved, here shown highlighted.

Clicking a candidate term displays detailed information on the right-hand side. From the top, panes include information on the scope note, the direct ancestors of the term (under the "Hierarchy" heading), synonyms and related terms.

Double-clicking a term takes the user to the Browse tab where it displays the term in its local context in the thesaurus hierarchy (figure 6.2, window on the left hand side in the foreground). The details of the term are still displayed on the right, and are updated whenever the user double-clicks another term, which will also expand Narrower Terms below the term concerned. Numbers in square brackets following the terms indicate the number of Narrower Terms available.

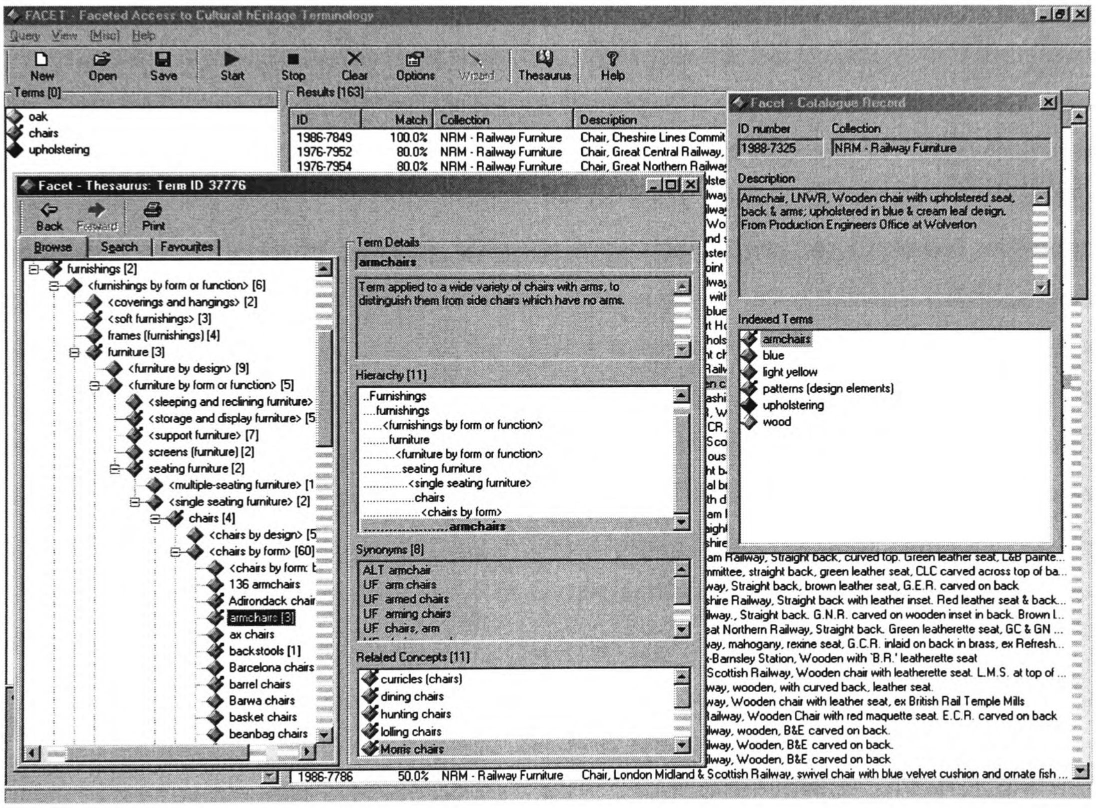


Figure 6.2 FACET 1.1 showing Thesaurus Browser with result record and a list of results in the background

From both the Browse and Search tab, terms can be dragged into the query. Clicking the "Run" button on the main toolbar starts a search for matching records. A ranked list of summaries of records appears in the results pane in query screen (fig. 6.2 in the background on the right hand side) and details can be viewed by double-clicking. The record opens in a new window in the foreground and shows the full description and the indexing terms (Fig. 6.2, right hand side in the foreground). Double-clicking on any term will open the Thesaurus Browser, centred on that term. This has happened in the example in figure 6.2 with the term "armchairs".

6.3.1 *Retrieval of candidate terms from the thesaurus*

In order to construct a query, the users' terminology must be matched to controlled terms from the thesaurus. The system performs a search on the thesaurus preferred terms and alternate terms in order to identify candidate terms. The wildcard character "*" retrieves partial matches. Users also have the option to search the term definitions, i.e. the scope notes. This can be useful when the user does not know the term to describe a concept, although they still have to identify the correct suggestion from the list of candidate terms. For example, a user could search for "fish" in the scope note. "Aquarium" is retrieved amongst other terms which also contain the word "fish" in the scope note.

6.3.2 *Semantic closeness expansion*

The semantic relationships between thesaurus terms provide the system with information on how similar two terms are by calculating the traversal cost from the amount and type of traversals necessary to move from one term to the other (Rada et al. 1989; Tudhope and Taylor 1997). Thus, it is not necessary for the users to construct a query that exactly matches a record's set of indexing terms because query terms are expanded and these terms included in the query. The system calculates a match value for each indexing term depending on the traversal cost to the closest query term, which must not exceed a threshold. The term match value feeds into a function that produces ranked results for a multi-term query. Different from other systems such as Ovid (see "Explode" function in section 5.4.1), terms are not only expanded along Narrower Terms, but also across Related and Broader Terms. Traversing different relationships can occur different costs in order to reflect the fact that BTs and RTs can be considered less close to a term than its NTs.

6.4 Study set-up

6.4.1 *Participants*

Museum professionals from a number of collaborating institutions, as well as two members of staff at the University of Glamorgan participated in the study. The participants at Glamorgan were a library professional and a lecturer in Human Computer Interaction. The librarian's previous work had included general information searching and at the time of the study, she was involved in a project on metadata. The lecturer also had a little information searching experience. Through their museums work, the other participants were to some extent familiar with the collection accessed by FACET and had some experience searching for records using MultiMimsy, a museums collection management system. The latter participants in particular correspond to the anticipated target group of a system like the FACET system. It was hoped that their comments would therefore be useful to the further development of system functionality, the interface and individual components within it. Participants were aware of the fact that this study was conducted as part of an ongoing development process and that the interface and its functionality might be improved.

6.4.2 Set-up of sessions

During the whole session, the system logged interactions. The conversation between participants and the author as the main evaluator was recorded with a tape recorder using a conference microphone. The screen capture was switched on after the training (refer to Appendix 6.1 for settings). During the pilot session, the screen capture was not continuous as it was unclear how large the files would be and the author did not want to lose data should a technical problem occur.

The set-up was not exactly the same for each session as slight changes had to be made to the data collection methods. Travelling to different locations meant the environment varied, it was for example not always possible to conduct the study session in a separate room and sometimes, other evaluators observed the session. The participants also had different levels of knowledge of the project and it was not possible to rectify this. However, the differences were recorded and considered in the data analysis.

Before the search sessions, participants were given a demonstration of the interface including an explanation of the semantic closeness expansion and in certain cases, shown features of the interface which might be developed for a later version, for example alternative thesaurus displays. The author then sat with each participant at a table so she could observe the events on the screen of the laptop. In one case, two participants worked together. The participants were given a handout of a training scenario with step-by-step instructions, which asked them to construct and modify a query for armchairs. It covered all aspects of the interface functionality needed to complete subsequent tasks. Participants were informed that they could ask the author for help with the interface at any time during the session and they were allowed to refer back to the training sheet. Most of the functionality had already been seen during the demonstration, so two participants decided not to go through the training.

6.4.3 Tasks

The task-based method described in 4.4.4 was adapted for the study to provoke specific, real-life behaviour, for example improving an initial choice of (thesaurus) terms, thesaurus browsing and reformulation using indexing terms from records. Tasks were constructed based on the collection and thesaurus in an attempt to imitate enquiries to museum's staff as no examples were available to the author.

The participants were given a handout outlining four tasks (see Appendix 6.2 for details), starting with a focused warm-up task that required participants to search for a record similar to one presented to them on paper. They could thus simply search for terms from this record in the thesaurus and drag them into the query. The second task required participants to find objects decorated with text. The main challenge for this task was to find suitable thesaurus terms when terms from the task question were not the best choice. The third question was more open. Participants were given a picture of a chair, and asked to find a matching record in the collection. No images of collection items were available within the database or otherwise, so the chair in the picture was not from the collection. However, there were matching records in the collection, and participants were successful in their searches. The last open task asked participants to explore the interface, to try out functionalities or construct their own queries. Not all participants completed this task due to time restrictions or lack of ideas.

6.5 Data collection methods

6.5.1 Audio recording

Already in the previous study (SOCS, section 5.4), audio recording was the primary source of qualitative data. Some of the participants' comments were very useful as they explained motivation behind interactions and revealed misunderstandings which would otherwise not become apparent. A separate interview element was considered in addition to prompts and questions during the search sessions, but was omitted in this study in order to shorten the sessions as participants were often under enough time pressure. For practical reasons, it was more appropriate to use a tape recorder rather than video equipment which would have been too bulky to carry around and set up in the locations visited. On one occasion, no recordings were made because a participant felt uncomfortable with the idea, but additional evaluators were present who took detailed notes.

6.5.2 Log files

Experience in the SOCS study (section 5.4.6) showed that information obtained from the tape alone was patchy. Although the taped conversation itself can normally be followed without problems, it is not always obvious which incident or precise interface element the speakers refer to. For this study, it was therefore considered necessary to have a comprehensive history of interactions to complement the audio data. The FACET system records specific user actions and system reactions into a log, based on recommendations from the literature (section 4.4.3) and considerations of the analysis conducted in the preliminary studies. An excerpt of raw log data is shown in figure 6.3.

```
20010905 12:55:45.012     BRQ440J default     frmThesaurusHelp.cmdFindNow_Click
        <DETAIL><FINDWHAT>upholstered</FINDWHAT><FINDWHERE>Terms
Only</FINDWHERE><GUIDETERMS>TRUE</GUIDETERMS></DETAIL>
20010905 12:55:45.012     BRQ440J default     frmThesaurusHelp.FindSingle
        <DETAIL><TERM>upholstered</TERM></DETAIL>
20010905 12:55:51.109     BRQ440J default     frmThesaurusHelp.lvw_DblClick
        <DETAIL></DETAIL>
20010905 12:55:53.359     BRQ440J default     frmThesaurusHelp.GetData
        <DETAIL><ID>137717</ID></DETAIL>
20010905 12:56:05.000     BRQ440J default     frmThesaurusHelp.lvw_DblClick
        <DETAIL></DETAIL>
20010905 12:56:06.051     BRQ440J default     frmThesaurusHelp.GetData
        <DETAIL><ID>231992</ID></DETAIL>
20010905 12:56:25.988     BRQ440J default     frmQuerySDI.tvwExpression_DragDrop
        <DETAIL></DETAIL>
20010905 12:56:28.898     BRQ440J default     frmThesaurusHelp.Form_Deactivate
        <DETAIL></DETAIL>
20010905 12:56:28.949     BRQ440J default     frmQuerySDI.Form_Activate <DETAIL></DETAIL>
20010905 12:56:31.809     BRQ440J default     frmQuerySDI.tbrToolbar_ButtonClick
        <DETAIL><BUTTON>Start</BUTTON></DETAIL>
20010905 12:56:31.859     BRQ440J default     frmQuerySDI.QueryStart     <DETAIL><?xml
version="1.0"?><FacetQuery/></DETAIL>
20010905 12:57:44.691     BRQ440J default     frmQuerySDI.QueryStart
        <DETAIL><RESULTS>175</RESULTS></DETAIL>
20010905 12:58:02.270     BRQ440J default     frmQuerySDI.lvwResults_DblClick
        <DETAIL><OBJECT>217477</OBJECT></DETAIL>
```

Figure 6.3 Excerpt from the raw log file

The user input logged included the terms submitted in the Term Finder including selected options, terms clicked, double-clicked, dragged or deleted from the query. Other logged interactions included pressed buttons or selected menu options and activated windows. Each log entry included a reference to the container or window in which it occurred, so that it was possible to determine with which window the user was interacting. Each term is uniquely identified by a number, which was logged rather than the English description as this was quicker and identifiers could easily be replaced by descriptions during the post-processing of the log files (section 6.6.1).

Logging the complete system response, for example all candidate terms or records retrieved, was considered to be too much data. The underlying collection was stable so that responses could be reconstructed, should related analysis be necessary. Mouse clicks on un-clickable objects and resizing of windows were not logged either. The data was not considered to be important enough in the context of information searching and would have created a large amount of noise in the log. If required, these interactions could also be identified using the screen capture files.

Although the raw log file can be difficult to understand, it can be prepared for analysis with a suitable post-processing program without further human intervention as opposed to screen capture files which have to be transcribed or annotated.

6.5.3 Screen capture

A screen capture tool proved useful to remind the author of the events of the sessions. She used Camtasia by TechSmith Corp.[1] which creates avi files[2] of a reasonably small size. These data might seem redundant as information on the search process is already captured by the log. However, the avi files revealed more information about terms entered, queries executed, results returned and exact interactions. They proved to be a good memory aid when no tape recording was available and were very useful for illustrating interesting incidents by cutting the related frames into separate files or by creating screen dumps.

6.5.4 Evaluator notes

The author took notes after the sessions to record any information not otherwise captured, for example if she had felt uncertain about how to react to a situation or observed a particularly notable incident to analyse. During some sessions, several evaluators were presents and multiple sets of notes exist.

6.6 Data analysis

Most of the methods used, especially audio recording and screen capture, resulted in data which required further processing before analysis, even though qualitative methods were to be used. Tape and screen records were first transcribed. In order to reduce the amount of work, the author first prepared rough transcripts, from which incidents of interest were identified and the relevant sections were transcribed in detail. The incidents were grouped and coded using open coding so that problem areas became apparent. The author also summarised the main issues from sessions and compared how participants completed tasks across tasks and between participants to identify the different approaches taken. Additional information about session progress was extracted and used to create graphs to facilitate this task. For particularly significant incidents, all available data from the different sources was collated to obtain a full picture. Comparisons with other participants' approaches highlighted potential reasons for difficulties and allowed some generalisation of issues.

[1] For more information see http://www.camtasia.com/products/camtasia/camtasia.asp

[2] Video file format - Audio Video Interleave – which stores video and audio data consecutively.

6.6.1 Log processing and analysis

An example of raw log data is shown in figure 6.3. It is not easy to understand and was therefore post-processed using Excel spreadsheets and a Visual Basic program written by the author. Some evaluators manually encode interaction log files which is a laborious task. In this research, an objective was to make the sequences of interaction more readable in order to analyse the process, but slight variations in the sequence (e.g. accidentally activating another form and going back to the previous one) had to be allowed for. Writing a program to perform these operations would have been unreasonable as these patterns were not the main focus of analysis.

Values from the log were converted to numbers and represented in graphs. This approach plays a more important role in the analysis of data from the FACET 2 study, and is discussed in more detail in section 8.5.2. Term identifiers were also replaced with term names (e.g. in column 3 in Figure 6.4), but later on, it was found that this information was largely unnecessary except for clarifying points for other analysis. The navigation of the thesaurus in itself was more important than the specific terms browsed.

12:55:45	FrmThesaurusHelp: Click cmdFindNow	- Details: Query:<DETAIL> <FINDWHAT>upholstered</FINDWHAT> <FINDWHERE>Terms Only</FINDWHERE> <GUIDETERMS>TRUE</GUIDETERMS> </DETAIL>
12:55:45	FrmThesaurusHelp: FindSingle	- Details: upholstered
12:55:51	FrmThesaurusHelp: DblClick lvw	
12:55:53	FrmThesaurusHelp: GetData	- Details: Thes term: {<upholstered seats>}(ID #137717)
12:56:05	FrmThesaurusHelp: DblClick lvw	
12:56:06	FrmThesaurusHelp: GetData	- Details: Thes term: {upholstering}(ID #231992)
12:56:26	FrmQuerySDI: DragDrop tvwExpression	
12:56:29	FrmQuerySDI: Activate ...	
12:56:32	FrmQuerySDI: tbrToolbar	- Details: Start
12:56:32	FrmQuerySDI: QueryStart	- Details: <FacetQuery/>
12:57:45	FrmQuerySDI: QueryStart	- Details: Results: 175
12:58:02	FrmQuerySDI: DblClick lvwResults	- Details: 217477
12:58:02	FrmCatalogue: GetData	- Details: Thes term: {Fail to find ID 217477 Could be record.}(ID #217477)

Figure 6.4 Excerpt from an early post-processed log file

6.6.2 Transcribing tape and screen recordings

Altogether, 8 hours of log files, about 7 hours of tape recordings and 4.5 hours of avi files were captured. For each session, the author made an index of events based on the avi files. These overviews contained brief information about Term Finder searches, thesaurus navigation, queries and results viewed. The visual nature of the avi file helped in the reflection on the process of the sessions in a way that the detail of the log files was unsuitable for. Not every form activation or single click actually represents a deliberate action by the user and this was easier to recognise in the avi file. As mentioned above, audio recordings were also transcribed.

6.6.3 Analysis of individual problematic situations

Each session was summarised and analysed under different aspects. For each task completed, it was assessed whether the task had been completed in a satisfactory manner from two perspectives. Firstly, it was judged whether they had fulfilled the tasks described in the handout (see Table 6.1), and secondly, whether it was possible to observe the behaviour the task had been designed for. The former aimed at identifying participants' troubles in completing tasks, the latter at improving tasks for future studies.

Initially, the audio transcripts were used to identify problematic situations contained either explicitly or implicitly in these data. Other incidents analysed were those the author noticed during the sessions and when looking at the avi files. These included incidents such as unsuccessful reformulation and unexpected interactions with the interface, for example one participant tried to drag a complete record rather than individual terms into the query.

For each important or problematic incident, data from different sources were collated (for more information refer to section 6.6.4 and figure 6.5 in particular), analysed in the context of the incident and compared with other participants' interactions for the same task to investigate why other participants did not have the same difficulties. This helped to explain what might have led to a particular situation and although influences such as the number of evaluators, skipping the training and time constraints were considered in particular, this approach left room to discover other factors, for example misinterpretation of specific features or the participants' habitual search approaches. Although these individual incidents as such are not representative, the author attempted to draw conclusions for consequences for information searching interfaces on a more general level. For example the two different approaches to browsing (section 6.7.7) showed that the same activity can under different circumstances be useful or inappropriate.

The incidents were also categorised using open coding in order to obtain a more general picture of the types of problems participants encountered. This part of the analysis is described in section 6.6.5. HCI issues at an interface level were identified while conducting these analyses and noted separately in order to modify and improve the interface.

6.6.4 Collating data from different sources

For incidents of particular interest, the data from the different sources was collated so that the participants' comments can be seen in the context of their interactions and retrieved result records. This resulted in a rich description of each incident. The collated data was coded using different fonts in order to distinguish the source:

Descriptive summary of screen capture files
LOG FILE (POST-PROCESSED)
`Transcript of the audio recording`
Post-hoc author notes

A good example for an incident described using data from all sources can be seen in figure 6.5. However, not every incident analysed later on in detail is actually documented equally well. Some incidents took only a minute or two, so that the participant might not have made any comments within that time. Often, screen capture and log file map so closely that only one or the other is used to describe the incident. Still, combining data collection methods and collating data from different sources allows analysts to see as much information in the data as possible. Refer to section 6.7.9 for the discussion of this incident.

THE USER SEARCHES THE THESAURUS FOR "TEXT":
52:08.1 THESAURUS FORM: CLICK "FIND NOW" TEXT

The user then adds "text" to the query. After this, he looks at the related terms *(probably prompted by the evaluator telling him that the related terms are hidden because the Thesaurus Browser window is not at its full size)*, **and drags the (only) related term ("words") also into the query.** THE USER EXECUTES THE QUERY.
53:06.4 QUERY FORM: QUERYSTART START QUERY: TEXT, WORDS
53:12.0 QUERY FORM: RESULTS: 3

Three records come up as a result:

ID	Match	Collection	Description
1999-7588	23.2%	NRM Railway Furniture	Station Seat: ex Market Weighton Passenger Station Cast Iron frame...
1986-7908	23.2%	NRM Railway Furniture	Station seat, GER, Slatted bench on cast-iron supports decorated with...
1986-7948	23.2%	NRM Railway Furniture	Station Seat, Great Central Railway, Wooden bench on log-design cas...

The user looks at the first record. The indexing terms are: cast iron, embossing, inlays, lettering (layout features), seating, wood.
53:57.0 CATALOGUE RECORD: ACTIVATE WINDOW

`Participant: Right, okay, so you've got the words "embossing". Ah, okay. That's just a different tense, isn't it?`

He then drags "lettering" from the record into the query.

`Participant: So I can now try that and see if it will get me anything more.`

54:40.9 QUERY FORM: TOOLBAR START QUERY: TEXT, WORDS, LETTERING
54:46.5 QUERY FORM: RESULTS: 3

The same three results with a better match come up.

(Note: All records are indexed with the term "Lettering" and this direct match improves the match value. The results were retrieved in the first place due to semantic expansion because "lettering" is a sibling of "text" and "words".)

The same results as above now have a 50.0% match.

`P: Right, so we still got only three. So I could try deleting those two.`

The user deletes "text" and re-runs query.

55:02.6	QUERY FORM: TOOLBAR	START QUERY: WORDS, LETTERING
55:07.1	QUERY FORM:	RESULTS: 3

The same results as above are returned now with a 66.7% match.

(Note: As the non-exact matches are removed and the exact match, "lettering", remains, the match improves because relatively more query terms match the indexing of the results.)

`P: Still only got three. I think it's because I have taken out the word "text", it's still the same three objects. So... so we could try taking out ...`

The user deletes "words", re-runs the query which now only contains "Lettering (layout features)".

55:34.4	QUERY FORM: TOOLBAR	START QUERY: LETTERING (LAYOUT FEATURES)
55:39.0	QUERY FORM:	RESULTS: 3

The same results as above now have a 100% match.

`P: Yes. Still the same three benches. I wonder whether there is other words that I can use. It's because of the words that I'm picking, I'm only bringing up the three. Benches do have the station name on and the crest, it's the railway furniture. So, I'll try ... Which word ... [`*typing sounds*`]. You've not had that before, have you?`

THE USER OPENS THE THESAURUS BROWSER AND SEARCHES FOR "DECORATIONS":

56:29.7	QUERY FORM: TOOLBAR	THESAURUS
56:29.7	QUERY FORM: CLICK "VIEW THESAURUS"	
56:48.4	THESAURUS FORM: CLICK "FIND NOW"	DECORATION
56:55.1	QUERY FORM: DRAGDROP EXPRESSION	

He drags "decorations" into the query.

`P: Right, so I could try that ...`

He executes a query with "decorations" and "lettering".

57:17.5	QUERY FORM: TOOLBAR	START
57:31.1	QUERY FORM:	RESULTS: 3

The query returns the same three results as above, this time with a 66% match.

(Note: "Decorations" is not matched by any indexing terms even with semantic expansion, which decreases the match value.)

`P: And I've still got the same three benches. I think I'm just going to have to admit - three benches.`

Figure 6.5 Example for collated data. Refer to section 6.7.9 for discussion of this incident.

6.6.5 Categorising breakdown and incidents

Apart from identifying how participants approached their searches, the author was also interested in gaining an insight into the nature of the difficulties participants encountered. Therefore, problematic situations and other incidents were categorised using open coding (used e.g. by Iivonen and Sonnenwald 1998; Xie 2002, see section 4.4.2) to obtain a larger picture of problems during the search sessions. During an initial analysis, significant incidents extracted from the audio transcript and other sources already discussed in 6.6.3 were first grouped with similar incidents to identify categories and then categorised accordingly. The categories identified related primarily to problems with thesaurus terms, for example understanding them and mapping concepts to thesaurus terms. Participants also had problems with the facets. Other difficulties were related to the presentation of information on the screen and misunderstandings of functionality. Altogether, the more serious problems related to conceptual rather than design issues. Therefore, the author then re-categorised the incidents according to the components of the interface and subsequently, according to the stages/phases of the information seeking process as proposed by Marchionini (1995).

- Defining the problem
- Formulating the query, including finding or selecting thesaurus terms
- Query construction, for example changing query settings
- Executing the query
- Examining results, including attempts to understand how they have been retrieved
- Reformulating the query

The last stage, reformulation, is expressed in Marchionini's model through iteration of previous stages. Here, it was felt that there is a significant difference between the initial query formulation and reformulation due to the necessity to modify the query and incidents were thus grouped separately. Within these stages, the incidents were grouped further in order to identify specific problem areas.

6.7 Results

The analysis of problematic incidents (6.6.5) revealed that the more significant problems were related to conceptual issues rather than design problems. This is to say, many participants lacked an awareness of the information searching process, which caused them more difficulties in searching than the interface itself.

The following section discusses a number of incidents which demonstrate these conceptual problems. In the context of this research, these problems are more interesting than interface-related issues, which are discussed briefly in connection with the resulting modification of the FACET system in chapter 7 and in Blocks (2002).

6.7.1 Overview of the sessions

Table 6.1 details the tasks completed by each participant and how long each session took. The pilot session is the only one to include tasks X and Y which concerned low level browser display issues resolved in that session. This session took with 3:30 hours the longest, but the participant was also consulted in some detail on some specific features of the interface. The differences in length of screen capture and duration of the sessions extracted from the log files can be explained by the fact that the training was not recorded with the screen capture software, and additional discussion took place after the searching in some sessions.

Session	1	2	3	4	5	6	7
Number of evaluators	4	4	1	1	1	1	1
Time constraints	No	Yes	No	No	Yes	No	Yes
Task 1: Find a record of a table to match the record of a chair	2	X	2	2	2	2	2
Task x: Finding specific terms by browsing the thesaurus	1	N/a	N/a	N/a	N/a	N/a	N/a
Task 2: Find items decorated with text	2	4	2	1	4	2	1
Task y: Comparing single and 2-level thesaurus displays	5	N/a	N/a	N/a	N/a	N/a	N/a
Task 3: Finding a chair that matches the image provided	1	3	3	2	4	2	1
Task 4: Open search – exploring the interface	X	X	X	3	X	3	X
Training	Yes	No	No	Yes	Yes	Yes	Yes
Audio	Yes	No	Yes	Yes	Yes	Yes	Yes
Camtasia	T1,3,5	Yes	Yes	Yes	Yes	Yes	Yes
Interface version	1.0.0	1.1.0	1.1.1	1.1.2	1.1.2	1.1.2	1.1.2
Approximate length of session (log, in hours)	3:30	1:20	0:45	0:50	0:30	1:00	0:30
Approximate length of screen capture (in hours)	0:50	0:32	0:50	0:45	0:19	0:53	0:27

Table 6.1 Search session information

Out of tasks 1, 2 and 3, which were used in all sessions, only 4 out of 20 completed were unsuccessful (value "1" in the table). A further 2 were completed in a manner that they have to be considered as appropriate, but terminated by the author because participants expressed doubts that further modifications would improve the query ("3"). Three tasks were declared completed by the participants themselves in sessions where they were pressed for time ("4"). This was also one of the reasons why some participants did not complete the open search tasks, but others did not know what to search for and also felt that exploration of the interface might have been more interesting before actually conducting a search. Not attempted tasks marked with "X", tasks not presented to participants with "N/a". Value "2" indicates that the participant completed the task successfully; the author normally moved on with the session. Task Y was found to be inappropriate ("5"). The outcome of tasks reflects the increasing level of difficulty from task 1 to 3.

6.7.2 Search approaches

Although there are not enough participants to speak of definite patterns, different search approaches can be identified. One main distinction is the effort spent on term selection (in both initial query construction and reformulation) versus the effort spent on looking through records.

For the first approach, which relies more on query reformulation than browsing the results, participants selected several terms before executing a query. The participants then expected to retrieve a relatively small number of results, several of which they might look at. They selected further terms from the records that appeared suitable for the task, dragged them into the query and re-ran it. In some cases, they repeated this procedure. Sometimes, participants were uncertain about additional terms to include, and thus ran queries containing only one term, but still invested more effort into reformulation than browsing records.

The other type of participants, who reformulated less but spent more effort on browsing lists of results, executed queries consisting of only one term without much consideration for additional ones. These queries mostly retrieved many results, and participants read through larger parts of the list. They mainly selected records from the first set of results without reformulating. This type of searching seems to be linked to time constraints – it was observed primarily in the sessions where the participants were aiming at finishing quickly. Time thus appears to be a factor which heavily influences search behaviour. In real life, this could also map to a number of motivations, for example finding any results that match the query, getting a quick overview of the data available or other types of searches where precision is of less importance. In the FACET 2 study, it became for example apparent that curators sometimes extract information on the collection for potential borrowers so that more specific reformulation would be inappropriate (section 8.7.4).

Although participants adopted different approaches to each task, a person's approach seemed to be relatively consistent during the session in that they either tended to concentrate on reformulation or browsing of records. Even when browsing through results rather than reformulating, most queries were successful, but this might be linked to the fact that the collections accessed in this study were relatively limited (389 records) and it was feasible to use this approach because only a reasonably small number of records achieved a high match. If participants utilise only one common term in a large collection, there is a risk that the number of highly ranked records will be so large that it is difficult to manually identify those which are suitable, without specifying additional criteria in the query.

6.7.3 Inappropriate transfer of approaches from Boolean systems

As mentioned in the introduction of this chapter, the FACET system with query expansion and ranked retrieval requires particular skills. A number of participants made references to Boolean logic and SQL statements though. They seemed to be used to this type of system either through database interfaces they had used previously or through Internet search engines, and seemed to expect the FACET system to work in a similar way. This affected their search behaviour. Some participants tried to be very specific in selecting thesaurus terms. Most did not have sufficient knowledge to select the correct technical terms, and this process caused them thus unnecessary frustration. The introduction to the interface might have given them the impression that it was desirable to use the most specific query terms possible. However, the query expansion algorithm would normally include Narrower Terms without affecting match values too negatively, so finding the very best term has not the same importance as in for example a Boolean system.

6.7.4 Conceptual misunderstanding of the hierarchical structure of the thesaurus

One participant who was not familiar with thesauri to the same extent as most others, seemed to have some trouble to distinguish between searching for a term in the thesaurus and querying the database, which is demonstrated in the following quote:

> (looking at armchairs and Narrower Terms in the thesaurus during the training)
> Participant: `this window here, this window is all chairs, so everything in here is related to chair?`
> Evaluator: `yeah.`
> Participant: `so at some point, and so at some point,` **`I'm going to click something and this is going to be filtered down to … a smaller option, a smaller selection.`**

The participant seemed to have understood the concept of using thesauri in searching on the whole, but struggled with specific details. In the example, he expected that moving to Narrower Terms in the thesaurus means filtering out records from the collections, as it might be encountered in subject classifications on the internet, for example in Yahoo!.

6.7.5 Breaking tasks down into components

One general recommendation in information searching is to identify the different concepts which form part of the question. Ideally, each concept is represented using one or more query terms. Breaking the task down into its components is particularly important when searching with a faceted thesaurus, but seemed to be for the most part difficult.

One participant thought at first that facets represent obligatory fields to be filled in for the query, then seemed to distinguish specifically between Objects and the other facets. This might have been influenced by the fact that the collection consists of objects (as opposed to e.g. documents or images). In another case, the participant seemed to interpret the facets as fields in a relational database. Although for example the "Materials" facet could represent a field of a record, the indexing terms do not necessarily have to occur in different fields. Each term is unique. "Tables" for example are distinguished as "Tables (Support furniture)" and "Tables (Documents)". Thus, for example "wood" will always be understood as a material, and not as a collection of trees.

Guide terms, which are used in the AAT to categorise long lists of Narrower Terms according to different perspectives (section 2.3.1), led some participants to expect to find other concepts represented in a similar way. Guide terms are for example "tables by form", "by function", "by location and context" and "by design". Several participants explained (illustrated with the words of one of them):

> The first thing about this table is that it's made of oak, so "tables by … material"… . It does not seem to be here.

This is an example where confusion regarding use of terms from several facets versus compound terms arises as well.

An example for breaking down a question can be found in task 3. Participants were asked to find a record of a chair that matched an image of what they referred to as an "upholstered chair". They ought to construct this concept from two separate terms, "upholstered" and "chairs", but considered it as one single concept. The fact that the thesaurus actually contains a term "upholstered seat" re-enforced this idea. The existence of this compound term appears to be an inconsistency in the AAT as it normally takes a faceted approach. However, very common multi-concept terms are entered as compound terms, which might be the case here. This term refers to the seat of a chair that corresponds to that in the image, but the records are not indexed accordingly. It is unfortunate that participants encountered this problem which is linked to the type of terms required in this specific task, but it is an opportunity to observe a problematic situation which might otherwise not have revealed itself. Sometimes, searchers will select the wrong term under the impression that their interpretation of its meaning corresponds to that of the thesaurus. This lead to unsuitable results and few possibilities to improve the query. Reasons in this particular case might be that this phrase just conforms too much to our expectation. One participant even said **before** looking the term up in Term Finder, without knowing what he would find in the thesaurus:

> I'd probably look for wood in the thesaurus, and copy that across, and **upholstered seat,** and copy that across and then search all three straight away. Because otherwise, [[3]there might be too many results and it is not practical]. So I'll go back to the thesaurus and do that now.

He was obviously expecting to find this exact term in the thesaurus. Another participants pointed out that the upholstered seat was one of the most noticeable features about the chair. Participants just did not realise that in fact the use of this term in the collection differed from their personal use. One person explained that "upholstering" is an activity, rather than an attribute, which might have lead them to select "upholstered seat" rather than the term "upholstering", which is actually used for indexing. All in all, this raises issues related to making the participant aware of the indexing techniques employed and associated implications for searching.

6.7.6 Selecting thesaurus terms

Selection of terms occurs at two different stages in the search process, once when formulating an initial query and again when modifying it. The data collected indicates that a number of issues contribute to the decision of which terms are considered appropriate. When the term was an activity rather than an attribute, participants felt that the term was not appropriate because of its grammatical form. This influenced for example the selection of "upholstered seat" above "upholstering". The participants' understanding of the term also influenced selection. If participants found a term that corresponded to their expectation, they often selected it without checking its meaning in the AAT, and problems occurred as described in figure 6.5 (p. 95) and section 6.7.9 with the term "decorations". Finally, the list from which terms have to be selected also affected selection as going through a long list of terms (for example retrieved when searching scope notes) can be laborious and frustrating.

---

[3] Square brackets within quotes indicate an approximate transcription.

On several occasions, participants were frustrated by the result records the system returned and felt unable to improve their query to a satisfactory level. One of the main reasons for this seems to be that they were unable to locate more appropriate thesaurus terms. Participants explicitly commented on this and several of them asked for advanced options to search the thesaurus. From their comments, this might require the option to use Boolean operators or to search for compound terms rather than one single word in order to be able to express more complex ideas.

> `[if I had been able to] pick out more than one word I could have put more ideas into what I was trying to describe.`

and

> **Participant:** `Okay, so. I'll start with chair. I could try with the definitions as well, see if that brings up anything more. Blimey. Okay.`
> **Evaluator:** `[what did you expect?]`
> **Participant:** `I didn't think I would get quite that many. [`*`clicking sounds`*`] Can you use more than one word?`

In the last case, the participant then attempted to browse the thesaurus from high levels of the hierarchies. This in itself is rather fruitless behaviour as can be seen in the detailed analysis of browsing in the next section.

6.7.7 Browsing the thesaurus

Participants showed different browsing behaviours, some more successful than others, which are illustrated by two incidents, analysed in detail.

In the first case, the participant started at the top level facet, "materials", in an attempt to find the term "wood" while searching for a chair matching the picture (task 3). No transcript is available for this session and it is unclear precisely what motivated him to opt for this approach rather than initially searching for "wood". He had just added a term from the Thesaurus Browser to the query so that the Term Finder pane, which might have led him to search, was hidden. The participant was unable to locate the term because he was unsure which terms to navigate through. This might be due to the structure and size of the AAT (over 28,000 preferred terms). The closer to the root or top level the terms, the more abstract they are. This makes it difficult for users, especially those who are not familiar with the

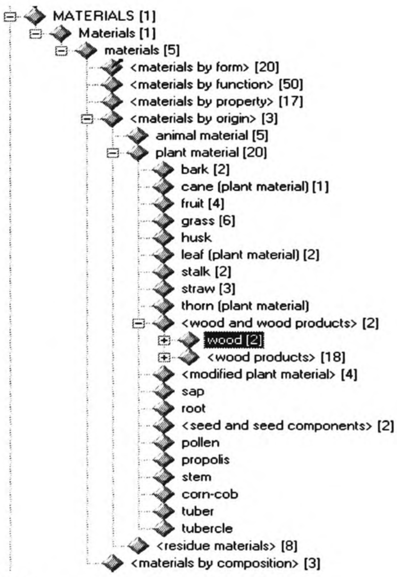


Figure 6.6 "Wood" in the thesaurus hierarchy

AAT, to distinguish between them. Figure 6.6 shows the path (six levels deep) he should have taken in the Thesaurus Browser.

Figure 6.7 shows a log excerpt of this incident. The first three clicks are actually on the same term. The participant initially clicked twice so that the hierarchy contracted and had to be expanded again. The participant did not get close to the hierarchy of the term "wood" and even decided to browse a second facet ("Physical attributes"). After another sequence of unsuccessful browsing, the participant thought of searching the thesaurus with Term Finder and instantly found "wood".

07:04.6	Thesaurus form: Double click Term	Thesaurus term: {Materials}
07:09.8	Thesaurus form: Double click Term	Thesaurus term: {Materials}
07:17.3	Thesaurus form: Double click Term	Thesaurus term: {Materials}
07:19.0	Thesaurus form: Double click Term	Thesaurus term: {Materials}
07:27.0	Thesaurus form: Double click Term	Thesaurus term: {materials}
07:34.7	Thesaurus form: Double click Term	Thesaurus term: {<materials by composition>}
07:50.6	Thesaurus form: Double click Term	Thesaurus term: {organic material}
08:03.3	Thesaurus form: Double click Term	Thesaurus term: {<materials by form>}
08:19.9	Thesaurus form: Double click Term	Thesaurus term: {<materials by function>}
08:47.7	Thesaurus form: Double click Term	Thesaurus term: {Physical attributes}
08:56.5	Thesaurus form: Double click Term	Thesaurus term: {Design Elements}
08:58.9	Thesaurus form: Double click Term	Thesaurus term: {<design elements>}

Figure 6.7 Log excerpt showing browsing for "wood"

In the successful other case, the participant entered the term "train" into the Thesaurus Finder. From the two results, the participant double-clicks "trains (vehicle groupings)" and browsed along its narrower term, "passenger trains" to "high speed trains". The participant then chose "passenger trains" as a query term. The main difference between these two incidents is that the entry point into the thesaurus is much lower in the second case. Within three clicks, the participant reached the bottom of the hierarchy. The term "trains" is six levels away from the top level facet – as is "wood" and not everybody would naturally identify "Object groupings and systems" (on the level below "Objects") as the facet to contain "trains", so even finding this seemingly obvious term by browsing down from the top of the thesaurus would be difficult.

6.7.8 Semantic closeness expansion

Semantic expansion has the advantage of including terms which are semantically close to query terms when searching for records (see section 6.3.2). Its effect might need to be explained in more detail in the training or help and be more clearly presented in the interface, so that users can benefit from it, unlike the participant in the following example.

Figure 6.8 shows a participant browsing the thesaurus in order to finds a term which best describes the chair in the picture (task 3). The participant starts by entering the text word "chair" into the Term Finder. He then double-clicks the candidate term "chairs" and opens the Thesaurus Browser. This sequence of interactions finishes by the participant dragging "fauteuils" into the query and executing it. However, later on in the session, three more, very similar interaction sequences occur, as the participant still felt that the terms were not appropriate. Altogether, the browsing took almost six minutes. Certainly, browsing is sensible and encouraged by providing the Thesaurus Browser window, but in this case, the benefit is limited since expansion is likely to include such terms, the participant could not identify a suitable term and the procedure was time-consuming.

11:14:11	frmThesaurusFind: Click cmdFindNow	
11:14:12	frmThesaurusFind: FindSingle	- Details: chair
11:14:23	frmThesaurusFind: Double click term	Thesaurus term: {chairs}(ID #37772)
11:14:26	Thesaurus Browser: Activate ...	
11:14:40	Thesaurus Browser: Double click term	Thesaurus term: {chairs}(ID #37772)
11:15:07	Thesaurus Browser: Double click term	Thesaurus term: {<chairs by design>}(ID #38257)
11:15:13	Thesaurus Browser: Double click term	Thesaurus term: {<chairs by form>}(ID #37775)
11:15:28	Thesaurus Browser: Double click term	Thesaurus term: {chair-tables}(ID #38148)
11:16:13	Thesaurus Browser: Double click term	Thesaurus term: {armchairs}(ID #37776)
11:16:32	Thesaurus Browser: Double click term	Thesaurus term: {elbow chairs}(ID #165321)
11:16:35	Thesaurus Browser: Double click term	Thesaurus term: {fauteuils}(ID #37856)
11:16:43	Thesaurus Browser: Double click term	Thesaurus term: {French elbow chairs}(ID #167602)
11:17:27	Thesaurus Browser: Double click Related Term	Thesaurus term: {fauteuils en cabriolet}(ID #37948)
11:18:21	Thesaurus Browser: Double click Related Term	Thesaurus term: {fauteuils}(ID #37856)
11:18:26	Thesaurus Browser: Double click term	Thesaurus term: {French elbow chairs}(ID #167602)
11:18:59	Thesaurus Browser: Double click term	Thesaurus term: {fauteuils}(ID #37856)
11:19:04	Thesaurus Browser: Double click term	Thesaurus term: {fauteuils}(ID #37856)
11:19:20	FrmMainMDI: Click ...	
11:19:20	frmQuery: Activate ...	
11:19:22	Thesaurus Browser: Activate ...	
11:19:32	Thesaurus Browser: Select	Thesaurus term: {French elbow chairs}(ID #167602)
11:19:37	Thesaurus Browser: Select	Thesaurus term: {fauteuils}(ID #37856)
11:19:39	Thesaurus Browser: Select	Thesaurus term: {bergeres}(ID #37801)
11:19:46	Thesaurus Browser: Select	Thesaurus term: {elbow chairs}(ID #165321)
11:20:00	Thesaurus Browser: Double click term	Thesaurus term: {French elbow chairs}(ID #167602)
11:20:03	Thesaurus Browser: Select	Thesaurus term: {fauteuils}(ID #37856)
11:20:06	frmQuery: DragDrop tvwExpression	

Figure 6.8 – Log excerpt showing navigation through the thesaurus in an attempt to find a suitable term to describe the chair in task 3

When completing task 2, where participants were asked to find objects inscribed with text, a participant had entered "embossing" and "text" into the query, and realised that even the results which were rated relatively highly did not necessarily contain these exact terms.

> Surface marking techniques… different types of embossing… metal work… That has come up with lots of different ways of marking things, but not necessarily…

> these… it can be embossed without it being text. So if I search on "embossing" and "text", it will give me things with embossed text. [The task is finding things decorated with text, not necessarily embossing], so I might be excluding some items.

The semantic closeness expansion algorithm lets the system retrieve records indexed with terms semantically close to "embossing" and "text". This allows the participant to become aware of other potential terms which are semantically close to his query terms. In this case, the participant would rather not use a too specific term as he wanted to include any records that could be relevant and he realised that many of the related thesaurus terms describe relevant techniques. The solution for reformulating the query would have been to add a broader term, for example "surface marking techniques", to the query, so that the expansion would include all these terms. Although the participant demonstrated an understanding of the semantic expansion, it appears that he does not fully see how to apply it to his task.

The participant's comment also touches on the issue that many systems do not deal with pre-coordination of indexing terms. The two terms can appear together as indexing terms in a record, but do not necessarily represent the concept of "embossed text". More advanced indexing techniques which create relationships between the indexing terms would be required for this type of retrieval. It is unclear why the participant assumed that the two terms would necessarily represent this concept as the same issue would exist in most systems. Even in a Boolean system, if these terms were connected using AND, writing could be painted onto an object and a pattern embossed into it, so that the two terms do not represent one concept. Potentially, this might be a general misconception on the part of the participant which goes beyond using FACET.

6.7.9 *Struggling with reformulation*

Although query reformulation resembles in principle query construction and is often modelled accordingly (e.g. in Marchionini's model and in that discussed in chapter 10), it takes place under different conditions as searchers may have gained knowledge about how their initial query attempted relates to the collection.

One participant had a particularly persistent difficulty with task 2 (finding items decorated with text) in that he reformulated the query repeatedly, but the same results were returned each time. For full details on the incident, refer to figure 6.5. It is not completely clear why reformulation of this query was unsuccessful, although the terms employed play a role. The challenge in this task had been to move from the terms in the task description to more suitable ones. The participant used two closely related terms ("text" and "words") which cover each other through the query expansion algorithm, i.e. they need not both be included in the query. To use a term from an initial result is generally speaking a good approach, and would have retrieved more results if this term ("lettering") had not been used to index only the three records already retrieved. It is unclear why the participant rejected the term "embossing". Additional records with more potentially suitable terms would have been retrieved, which could have broken the impasse.

The term "decorations" did not improve the query due to a misunderstanding. The participant believed that "decorations" refers to additions for aesthetic purposes. As he did not check the scope note or the local context but directly dragged the term into the query (56:55.1), he remained unaware that this term actually refers to medals or badges. This incident demonstrates the risk that when a term's meaning is supposedly known, the user may not take the time to double-check the local context. Training might not be a solution because this participant demonstrated awareness in other situations. Providing more context in Term Finder, for example the Broader Term, might have helped the searchers realise the ambiguity. However, this information would have to be visible without any further action by the user, which raises issues of space and information overload in the Term Finder display.

Overall, the participant employed reasonable, advisable strategies to modify the query. It was not the strategies themselves that led to the results, which would indeed have fulfilled the task requirements. The task was considered unsuccessful by the user because he **wanted** to find more results which were not necessarily benches, and was unable to do so, even though different strategies would have retrieved other records.

This and other incidents around this task demonstrates a particularly serious issue as participants actually followed common sense approaches to searching and reformulating. As far as it can be judged, the main reason behind the difficulties was an unfortunate choice of terms, for example selecting an indexing term which was only used for exactly those few records retrieved by the other query terms. Other terms would have retrieved new records with more possibilities to use indexing terms and ended the vicious circle. It would be particularly important to provide some form of support to searchers in these situations where the reasons for non-success are difficult to detect. It would not be necessary for the system to assess what caused the problem, it might simply be enough to make suggestions for alternative approaches they could follow. Potentially, they might not find better results, but at least they might persist slightly longer, knowing they have another option to approach their task, rather than giving up once their initial attempts seem to have failed.

6.7.10 Assessment of records

Once results have been returned, searchers normally assess whether they match their expectations. How exactly they do this seems to vary from person to person, but some of the influences are the match value, the free text description and, similar to dealing candidate terms, possibly the position of the item in the list.

Although participants mentioned that one of their criteria was that all query terms should be included in the indexing terms, they did not necessarily inspect all records with the same match value, which indicated an equal amount of match between query and indexing terms. Some participants also looked at records with lower match values. This seemed to be primarily to judge records overall, but could be influenced by previous experience, for example with Internet searching where the highest ranked matches are not necessarily the most appropriate.

Some users' choices seemed to be more influenced by the value of the match than others. Uncertain about how exactly it is calculated, they might have preferred to refer to other information. There also seemed to be some discrepancy of what can be judged to be a good match, possibly influenced by previous online searching experience. While others would look at records with much lower values, one participant commented that "54.5 %, that's actually quite a way from the....(query)". However, considering that guide terms will increase the distance between two terms and thus negatively affect the match value, this match value is relatively good, unless the user insists on a very close match.

The textual description which is part of the simple display also helped users make an informed relevance judgement on which records correspond to their expectations. Record descriptions often contain terms in some way related to the query term, for example when searching for the upholstered chair, users clicked on chairs described as "...padded drylon seat...". For task 1, users specifically clicked on records where the description contained the words "Royal Saloon".

6.8 Evaluation of methodology

The following section talks about the main lessons learnt from conducting this study. Overall, the data collection was successful and the necessary analysis was possible.

Collating data from different sources proved to be very useful in order to examine incidents in detail and to contrast behaviour in similar situations, which resulted in more general conclusions. However, sometimes incidents were only about a minute or two long, and there is little or no information on the participant's thought process during that time. This might be due to them thinking and not being able to reflect at the same time, or executing actions in an "automatic" manner, so that they do not consciously think about them. In some cases, people did not speak while they were thinking, or they did not actually voice their reflections but only their conclusions (also see Branch 2000). A discussion after the incident or the search session might help to clarify these situations. It was not always easy to prompt participants to speak without creating the impression that something might be wrong (Tamler 1998hh). Consequently, data for certain investigations, for example precise reasons for selecting thesaurus terms or records, would have to be collected specifically for this purpose in a way that ensures the necessary detail.

6.8.1 *Introducing the interface*

When introducing participants to a system with which they are unfamiliar, it is important to explain especially the exceptional features clearly. In this study, the semantic closeness expansion was probably not explained in enough detail which resulted in participants trying unnecessarily to find the most specific term instead of making use of the expansion. Input on features under development should be collected after the search session in order to avoid confusion about the relevance of these features for the tasks. Similarly, only one option to perform actions (menus, buttons or shortcuts) should be introduced in a short study. The author also found that an introduction to searching in itself would probably benefit those participants for whom this is only a small aspect of their work, unless this defeats the research topic.

Although the evaluator should be prepared for potential differences between users, it is probably helpful to take a general approach in the training and deal with specific issues as and when they arise, which can be done easily in a one-to-one session. Instructions contained information on how to drag and drop a term, which resulted in confusion because a participant was familiar with the concept but found the description unclear.

6.8.2 *User background and additional information*

It is hardly ever possible to control all aspects of a study, but it is important to record the differences. In this study, the demonstration of the system was dependent on the different audiences. Some participants had more knowledge about the project than others and this had to be taken into account. It was also necessary to fulfil their different expectations in return for their time. The demonstrations were not recorded, so that it is not possible in hindsight to verify the assumption that certain issues mentioned might have influence behaviour. In future studies, when it is not possible to ensure that all participants receive exactly the same information, audio recording might ensure a means of at least assessing the influence of the inevitable differences.

Other differences between participants were familiarity with the AAT, with the collection, and the librarian's professional experience in online searching. The sample was small and their work situation as well as their search proficiency were known to some extent, so that this information was taken into account when analysing the data. Two visits took place where two participants completed tasks independently at the same location. These pairs of sessions each showed particular characteristics which were not identified in the sessions conducted in other locations. In one case, both participants mentioned that they had forgotten that they could search the thesaurus for terms other than object names. One of them explained that they were involved in work with a thesaurus containing solely object names, and that this might have been the reason. In the other case, the participants had trouble getting started. They were familiar with the MultiMimsy interface to the collection, which presents participants with a dialog window for a fielded search. This suggests that either the participants' backgrounds through working in the same environment, or the demonstration/training and session set-up resulted in particular conditions reflected thus in the data. In other studies with larger groups or with anonymous participants, demographic data and information on their level of experience with computers and searching should be collected in order to be aware of factors that could have caused differences in behaviour.

6.8.3 Scenario construction

Overall, the tasks proved to be acceptable to conduct a near-life (i.e. as realistic as possible) study of the interface and discussion of the scenarios with the participants showed that most of them were indeed the kind of requests a museum would receive. The warm-up task required participants to search for a record similar to one presented to them on paper. This actually corresponded to an activity when putting together a new display of exhibits in a museum. The second task, finding objects decorated with text, was rather unrealistic as museums professionals were in the first place interested in the objects as such, rather than techniques used in producing or decorating them. However, other user groups such as artists or artisans might approach a collection from this angle. For the third question, participants searched the database for a record that would describe the provided picture of a chair. Apparently, this is a common type of request the public brings to the museum, for example to find out whether furniture they owned was of any value.

The tasks were meant to lead to some specific interactions, for example thesaurus navigation starting from a term found in the thesaurus or in a record. As participants sometimes took different approaches than anticipated, this did not always happen and results were sometimes retrieved without much interaction. In some sessions, the participants were under time pressure so that they aimed at completing tasks as fast as possible. This resulted in a superficial approach, and it would probably have been better to give them only one task for which they would have had a reasonable amount of time.

The intention behind constructing a task is not to impose a certain approach over users' individual styles, but to investigate the situations where troubles occur because in a real-life situation, even searchers who are willing to browse large amounts of records might encounter a situation where this approach does not result in satisfactory records and where they thus need to fall back on alternatives.

It can be a complex task to develop scenarios which make participants utilise a specific functionality, especially when the study demands a realistic setting. At most search stages, searchers have choices. It would reduce the value of the data to use the tasks to restrict these, and might lead to overlooking problematic situations that might occur beyond that point in the search. Some paths might be too complex for participants to succeed, which might demoralise them. Therefore, an evaluator would want to construct the task in a way that will not lead participants to take this particular approach or which might result in them searching for certain types of terms, for example those which are ambiguous or too general. Participants need a certain amount of contextual information so that the task is not too vague as they found it difficult to see what they were expected to do and also had the option to easily accept records as suitable. On the other hand, a task that is closed from the start might be solved too easily so that no data relevant to research questions regarding problematic situations can be collected.

The collection normally imposes certain limitations – some thesaurus branches might be particularly suitable for observing browsing, but they might not actually have been used to index the collection. One solution to this dilemma might be an evolving scenario where the participants receive a general overview over the task, but start only with a few initial instructions. As they move through the search, the evaluator gives additional information. In the case that the search progresses well, these could be set-backs, for example "This record is not relevant because of this property." If the participant is struggling, comments like "This property would actually be relevant as well." would give them additional information to work with. In some way, the participant will function as an intermediary trying to fulfil the needs of an end-user. This might seem more removed from a real-life situation, but would have been realistic for many of the participants in this study. If they are searching for the author though, they would not have the option to just terminate the search either. Judging from participants' comments on the tasks, this approach would also give them the necessary direction in a search they do not really own.

6.8.4 Interacting with participants

In this study, the author used an informal interview technique by discussing the searches as the participants went along. Communicating in this situation is more difficult than it might appear. In order to avoid putting words into the participant's mouth, the author permanently had to be aware not to follow standard rules of communication, for example completing a sentence for the other person. On top of that, non-verbal communication could not express surprise or other reactions which might discourage the participants. Even though in this study, discourse and interactions were recorded automatically, the author still had to pay attention to numerous aspects such as the time taken for the task and the interactions completed in order to prompt appropriately for more information and attempting to ask the prepared questions to ensure a certain uniformity of information collected across sessions and to deal with unexpected situations. Participants encountered various situations where they got frustrated and the author had to reassure them accordingly. In other situations, participants were silent, and it is very difficult to prompt users here without giving them the impression that something is wrong as many prompts could be understood as reproaches, for example "Why did you do that now?". As Tamler (1998) reports, this is often understood to imply criticism and he suggests rephrasing the sentence to "When you did that, what did you expect to happen?". The question is also more precise.

6.9 Conclusions on the methodology

The methodology has reached a level where techniques can still be improved – often related to the author's experience, for example interacting with participants, but overall resulted in detailed, valid data. Looking at the different data sources, it was very valuable to use them, despite the fact that for example transcribing the tape and analysing the screen recording files was time consuming. They provide a rich basis for analysis and triangulation of data proved to be very useful obtaining a fuller picture of participant's reasoning and interactions.

Certain aspects in the methodology still need to be resolved. Not all participants are comfortable with audio recording and an alternative method of collecting this rich, varied data might be considered in future work. The observer prompts also need to be developed further, especially comments to encourage thinking aloud would be helpful. One option might also be to conduct sessions with two collaborating users, which would require them to confer on their steps. However, this set-up might not work with every person or be difficult due to their work relationships. During the sessions, it also became clear that the author needs to reassure users and give them appropriate pointers when they are for example unsure where to start or how to proceed. At the same time, it is not possible to be prepared for all eventualities, as users do take some unexpected routes.

The tasks given to participants to complete needed revising in order to really ensure that specific interactions can be observed. Evolving scenarios might be a solution to this problem, as participants would have enough information to successfully continue searching, but ought also encounter some problems. The latter is of particular importance to collecting necessary data, and possibilities for intervention could be assessed at the same time. The participants might be more likely to complete tasks with a feeling of achievement, which is desirable in this type of study. Evolving scenarios might also make it easier to accommodate time restrictions, as shorter and longer versions of the tasks could be envisaged.

6.10 Conclusions

Most troubles people seemed to encounter, apart from small low-level issues, are conceptual. These troubles seemed to cover most aspects of the IR process, some are specific to using the thesaurus. One issue was that it is not evident how to resolve these issues sufficiently through changes in the interface.

The interface provides good interaction mechanisms such as the drag-and-drop from the thesaurus display and records which are easy to learn and use and thus enable participants to easily construct and reformulate their queries. The display of results seems to provide enough feedback for participants to decide which records are relevant and to have an indication on some possibilities for reformulation. At the same time however, participants need more **conceptual** support in constructing and reformulating their queries when they do not fulfil the users' expectation. Making good use of the semantic expansion was a particular challenge, as participants seemed to have difficulty in understanding how it worked in the first place.

Looking at the problems that occurred during the sessions, it becomes apparent that despite some minor issues, the interface itself is easy enough to use for a person with little knowledge about it. However, constructing a good query that returns satisfactory results seems to be more challenging. It is during the search process itself that most users become frustrated and do not know how to proceed best. This ranges from breaking down the query into concepts that match thesaurus terms to improving results through repeated reformulation. It might help users to have a better concept of the stages of a search. Some guidance throughout the system might re-enforce this. From the data collected and information searching models in the literature, it was possible to construct a preliminary model that demonstrates points where support is required for both general information searching and thesaurus use in particular, which was extended using data from the FACET 2 study (chapters 8 and 9) and could eventually be integrated into interfaces. Thus, the focus has now shifted from examining thesaurus use within information searching to how the thesaurus can inform the information searching process as a whole.

One of the major findings was that problems seemed to be of a conceptual nature and related to knowledge of the search process, rather than the actual design of the interface. In this study, an attempt was made to find means to reconcile the powerful technology of semantic expansion and thesauri with the users' experience of them. A better understanding of the search process including reformulation and relevance assessment of results in a thesaurus environment will assist this development as the overall search process and the users' mental model of it seem to be particularly important aspects.

Chapter 7 – Modifications to FACET 1

7.1 Overview of the chapter

The study conducted with the FACET 1 interface has been described in detail in chapter 6. The main objective of the study had been to investigate information searching behaviour, not to evaluate the interface, and the more serious problems which occurred during the information searching process were linked to conceptual, not interface issues. However, some findings related to the design of the interface and fed back into the further development of the system. These findings are discussed briefly in this chapter together with a description of the revised version of the interface, which was then used in the FACET 2 study described in chapter 8. One incident is discussed here in more detail (section 7.3) as it also demonstrates the limited support of the search process within the pilot version (1.0) of the interface, which was not been resolved completely in the version used in the FACET 1 study (version 1.1) either and which the FACET 2 version attempted to improve further.

According to the objectives of the FACET project, some modifications were also made to the underlying retrieval mechanisms.

7.2 Changes to version 1.1 of the FACET interface

Some modifications of the interface were made in order to facilitate interactions in general. The single term search was replaced by a multi-term search facility, allowing users to enter both single words and phrases. The thesaurus display was made simpler by omitting the number of child terms which had been interpreted for example as levels below the current term and caused some confusion when they indicated only one Narrower Term. Different views on the thesaurus were implemented in order to allow users to take different perspectives on the terms in the thesaurus. In addition to these changes, navigation of the thesaurus was made easier by including "Back" and "Forward" buttons (Figure 7.1 and 7.17) which let users jump back to previously viewed terms without actually having to locate them in the thesaurus display. A message box indicates that the query is executing to make this more obvious.

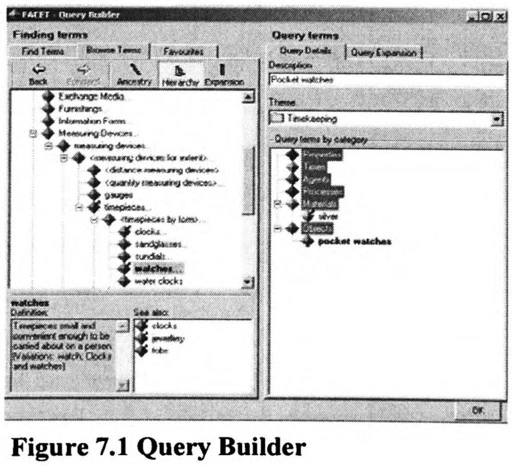


Figure 7.1 Query Builder

The background of the interface is now taken up by a list of queries and results (Figure 7.2 and 7.18). For each query, the collection, name/description and query terms are shown and the result records of the selected query are displayed in a list on the bottom of the screen. Indexing terms are shown in the list of records so that users can judge records without having to open them. This also facilitates the comparison of several queries and users need not deal with multiple windows.

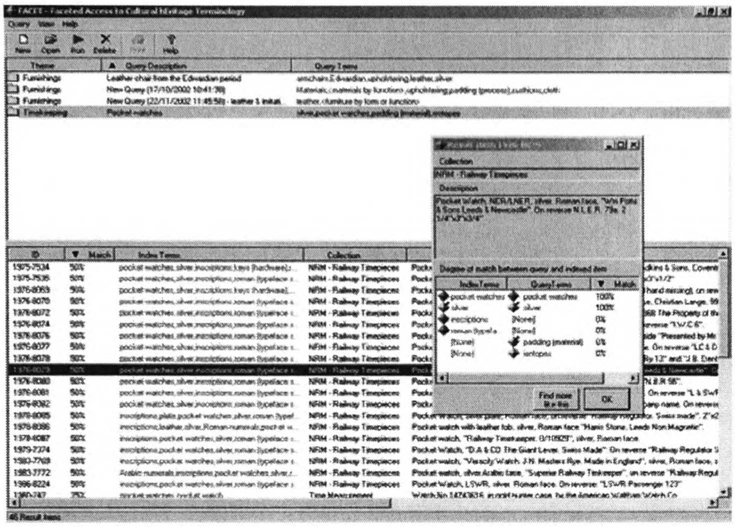


Figure 7.2 Impression of revised interface

When double-clicking a query, the Query Builder, described in section 7.4, opens with the query terms displayed on the right hand side. The display of individual records has been described in section 7.5 in connection with the retrieval mechanism and match value.

7.3 Window switching in FACET 1.0

This incident, which demonstrates the difficulties that can arise from multi-window displays, occurred during the pilot session of the FACET 1 study. The initial prototype comprised a number of individual windows for searching the thesaurus, browsing it, constructing the query and viewing the results (Figures 7.3 and 7.4). In the example shown, the participant is trying to find objects decorated with text (task 2). Screen dumps of the original screen capture file provide a visual representation of the events (Figures 7.5 – 7.10). Figure 7.11 shows the corresponding extract from the post-processed log. The indentation of the second column indicates the interface form in use. It can be seen that the participant interacted very little with one window before changing to another one. Note that the participant had to return to the background window or main form to execute the query (e.g. 11:04:40). On several occasions, the participant activated windows (e.g. 11:04:38) simply to move them. The following paragraph presents the narrative of the incident taken from the screen capture transcript.

The user has opened the Term Finder window and enters "text". He executes the search and uses the menu to open a new query (fig. 7.5). He then drags the result, "text" from the Term Finder into the query and goes back to the Term Finder to search for "text decoration" and then just "decoration". He drags the term "decoration" into the query. He moves the query window up on the screen (fig. 7.6) and executes the query by clicking the "Start button" on the main form. The results appear (fig. 7.7) and the user double-clicks one to open the record (fig. 7.8). On the record form, he then double-clicks the term "lettering (layout features)" which brings up the Thesaurus Browser (fig. 7.9). The user then has to move the record form in order to uncover the right hand side of the Thesaurus Browser which is obscured (fig. 7.10 and around 11:10:05).

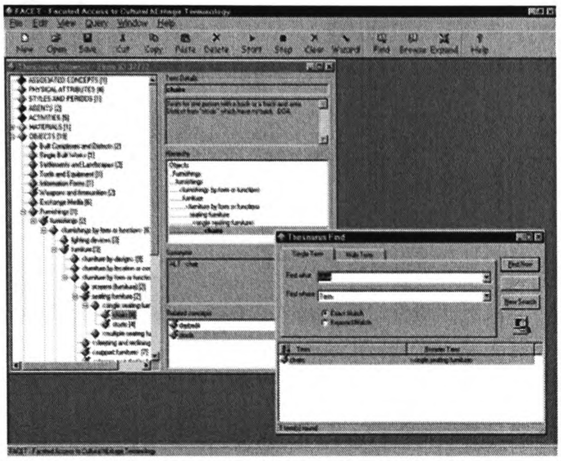


Figure 7.3 Thesaurus Browser and Finder

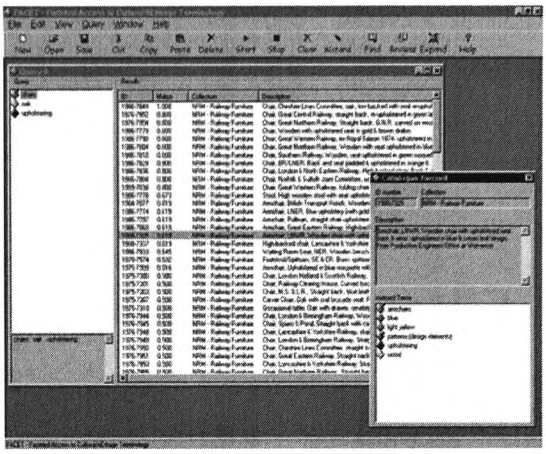


Figure 7.4 Query window and results record

11:02:44	Main form: Click (Open Term Finder)	
11:02:44	Term Finder: Activate window	
11:03:01	Term Finder: Click "Find now"	Text
11:03:13	Main form: Click (Open new Query form)	
11:03:13	Query form: Activate window	
11:03:17	Term Finder: Activate window	
11:03:39	Query form: DragDrop Expression [fig. 7.5]	Text
11:03:59	Term Finder: Click "Find now"	Text decoration
11:04:17	Term Finder: Click "Find now"	Decoration
11:04:34	Query form: DragDrop Expression	Decorations
11:04:38	Query form: Activate window [fig. 7.6]	
11:04:40	Main form: Click (Execute Query)	
11:04:40	Query form: QueryStart [fig. 7.7]	Text, Decorations
11:05:13	Query form: Double click Result record	217474
11:05:15	Catalogue record: Activate window [fig. 7.8]	
11:05:35	Catalogue record: Double click	lettering (layout features)
11:05:37	Thesaurus Browser: GetData	Thesaurus term: lettering (layout features)
11:05:38	Main form: Activate window	
11:05:38	Thes. Brows.: Activate window [fig. 7.9]	
11:10:05	Catalogue record: Activate window [fig. 7.10]	

Figure 7.11 Extract from the post-processed log showing windowing issue

As a consequence of this pilot session, the interface was modified as described in chapter 6 to reduce window context switching. Underlying these problems appears to be a lack of intrinsic order in the participant's progress through the search. He had to interact with a number of components all contained in separate windows which could be opened in any sequence and moved anywhere on the screen. The lack of sequence was further reinforced by the general toolbar which contained the menus and buttons referring to

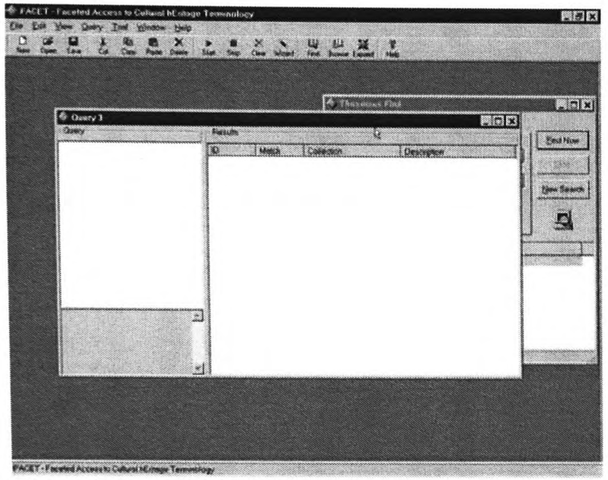


Figure 7.5 – at 11.03:39 in fig. 7.11

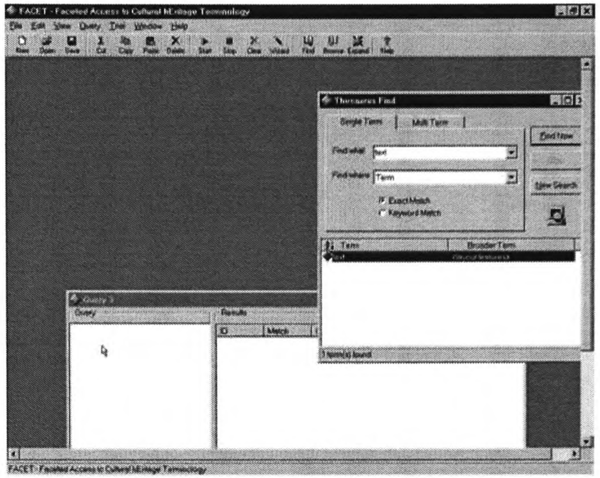


Figure 7.6 – at 11.04:38 in fig. 7.11

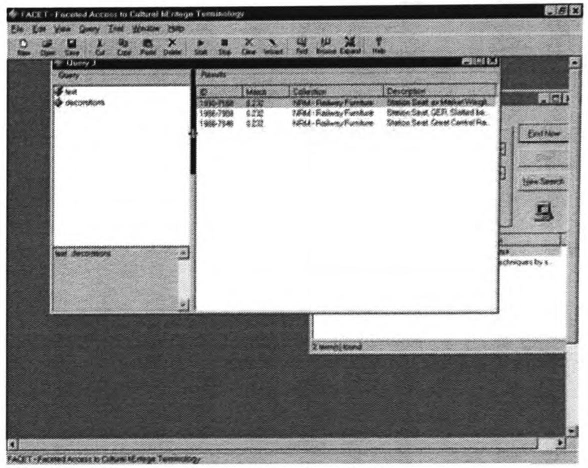


Figure 7.7 – at 11.04:40 in fig. 7.11

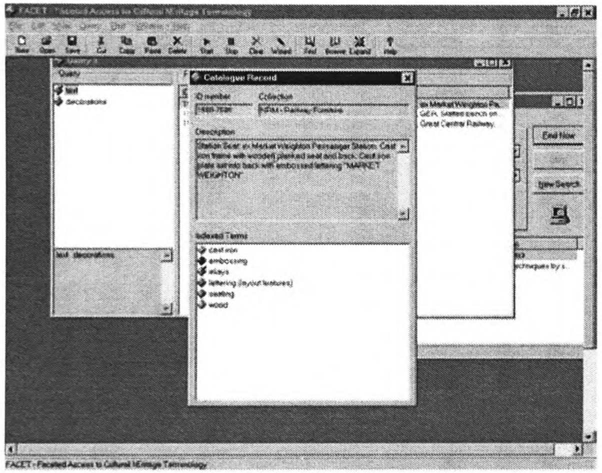


Figure 7.8 – at 11.05:15 in fig. 7.11

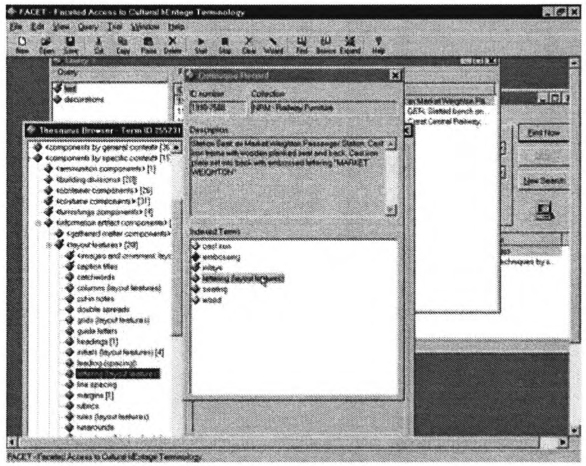


Figure 7.9 – at 11.05:38 in fig. 7.11

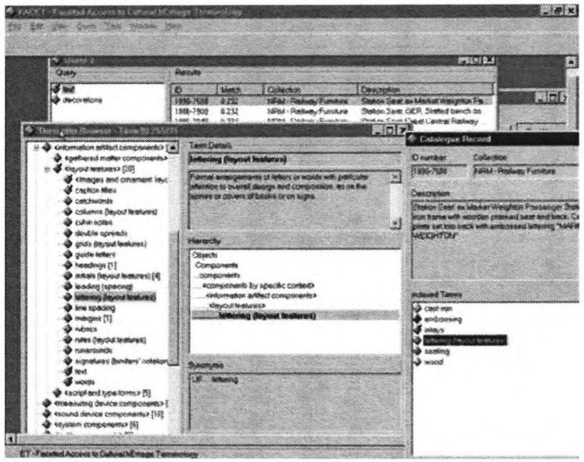


Figure 7.10 – at 11.10:05 in fig. 7.11

117

different windows. The searcher could not anticipate easily which windows will follow, continually having to move or adjust existing windows as new ones appear.

The new interface design was also meant to reinforce stages of information searching as identified for example by Marchinonini (Chapter 6.6.5), which were adapted to the FACET interface. The aim was to help resolve issues regarding confusion between searching for terms in the thesaurus and querying the database. The use of overlapping windows remained an issue though, for example when the Thesaurus Browser window was not minimised before executing the search, it covers the results. When participants minimised it and restored it again, it sometimes was not large enough to display the related terms, and some participants were unsure of how to restore this window.

Viewing records was also slightly awkward in that participants tended to close the window before double-clicking the next record to open it. As only one record can be displayed at the time, this action is redundant.

The study described in chapter 6 also showed that the stages of the search process are still not explicit enough. Further information on changes made from FACET 1 to FACET 2 because of this are discussed in the next section.

7.4 Support of the search process

Participants in the FACET 1 study seemed to primarily encounter problems in information searching because of their unawareness of the search process and the limited guidance that the interface provided here. Thus, the interface needed to be restructured. In order to support the search process better, the thesaurus search facility, Term Finder, and the query were integrated into the Query Builder (Figure 7.1). On the left hand side of the screen, the user has access to the thesaurus, on the right hand side, the query is constructed and options are set. The advantage of this arrangement is that the processes of term selection and query construction is more natural, taking place from left to right, and both windows are available so that when users open a query to modify it, they directly have access to the thesaurus. After searching the thesaurus for terms, they can also easily be included in the query.

On the "Search" tab, which is the default, the user enters a word or phrase which is mapped to potential thesaurus terms (Figure 7.12). Double-clicking a term jumps to the "Browse" tab where users can see the term in its local context and view other information such as the scope note and related terms. Three different views on the term are available, the local context within the hierarchical structure (Figure 7.17), the ancestry (Figure 7.13) which only shows direct broader terms of the selected term and the expansion view (Figure 7.14), which shows those terms in the thesaurus which are semantically close to the selected term, starting from the highest match. This display resembles that for modifications of the Query Expansion discussed in section 7.3, which is shown on the right in figure 7.14 as an alternative option for the query. It was assumed that these different display options would allow users to better understand the structure of the thesaurus and how individual terms relate to the whole.

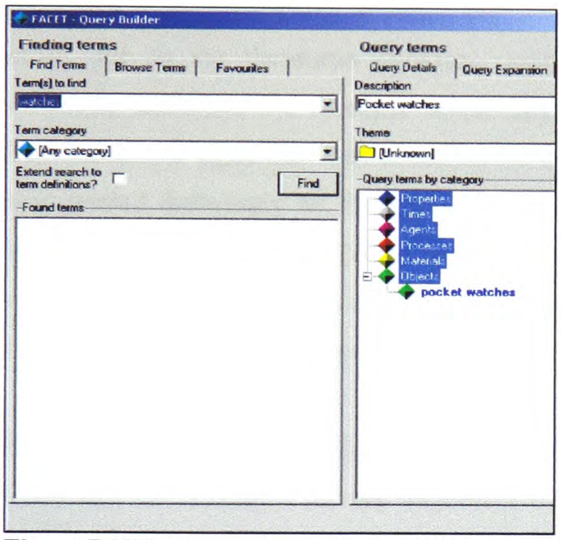


Figure 7.12 Query Builder - Search tab on the left, query on the right

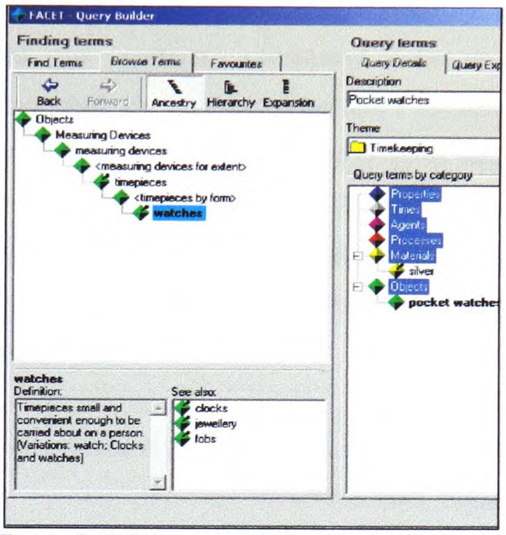
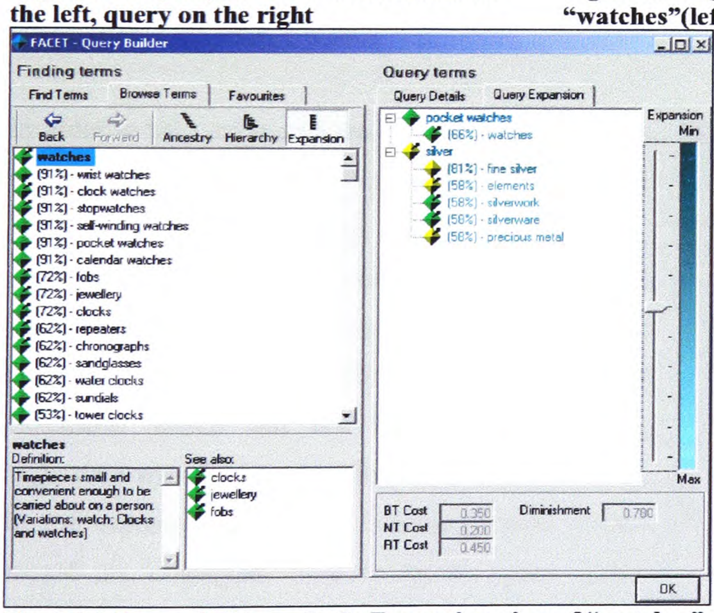


Figure 7.13 Query Builder – Ancestry of "watches"(left)



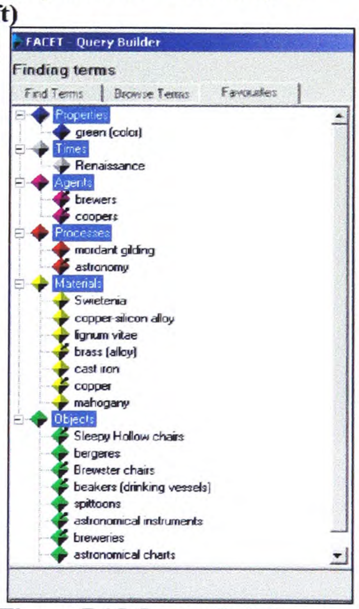


Fig. 7.14 Query Builder – Left: Expansion view of "watches" Right: Expansion of query terms ("silver", "pocket watches")

Figure 7.15 Query Builder – Favourites

A third tab, "Favourites" (Figure 7.15) allows users to "bookmark" terms. This functionality is considered especially useful for regular users because it saves them searching the thesaurus at the beginning of every session.

From any of the thesaurus displays on the left, terms can be dragged into the query on the right hand side. In order to increase users' awareness of the faceted nature of the thesaurus, the query terms are no longer simply listed, but appear under a heading, which represents their facet of origin (Figures 7.1/7.17, 7.13, 7.15).

7.5 Retrieval mechanism

During the FACET 1 study, it became clear that the indiscriminate expansion algorithm retrieved unsuitable results at times because certain terms were considered semantically close due to the thesaurus structure, whereas the searchers considered them to be distinctly different, for example when they searched for chairs, tables were retrieved. Therefore, the algorithm was modified so that users could select one focus term which was treated differently by the algorithm. This term was only expanded downwards, not across Broader or Related Terms to ensure that all records more specifically fulfilled this criterion. The focus term is set by selecting "Set focus" from the right-click menu on the term and it subsequently appears bold (e.g. figure 7.13).

In the second version of FACET, users also have more control over the query settings. They can for example set the expansion rate so that they can influence how closely records have to match the query. This is done by using a slider on the "Query expansion" tab (Figure 7.14, on the right).

In order to help users understand better how changes in the query expansion affected their results, more information on how records match the query are provided. Firstly, the match value of the records was rounded instead of being presented with a decimal point as this seemed to indicate a high degree of exactness, which was not intentional. Secondly, more data was displayed in the record (Figure 7.16). In the column on the left, indexing terms of the record are displayed. Next to it, the query terms which were instrumental in retrieving the record and their match to the corresponding indexing term are shown. This presentation allows users to quickly identify those indexing terms which are not represented by a query term and also see which query terms have not been matched by indexing terms of the record. The match value indicating the distance between the two

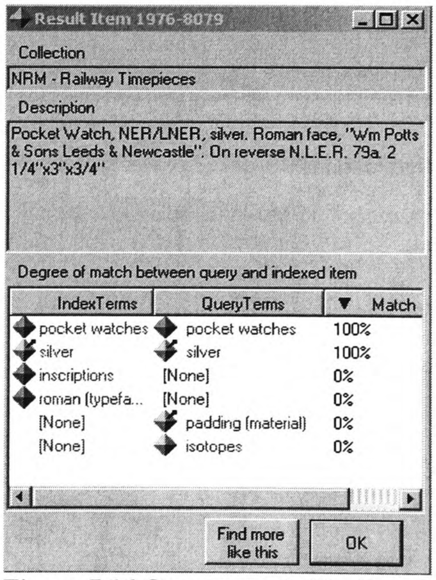


Figure 7.16 Open record. The list of queries and list of results for query "Pocket watches" in the main screen.

gives users an idea of how the total match for the record was calculated and an indication of how the terms contributed to its retrieval. Inspection of several records can thus aid query reformulation as it gives users an idea of query terms which should be modified and indexing terms which could be included in the query.

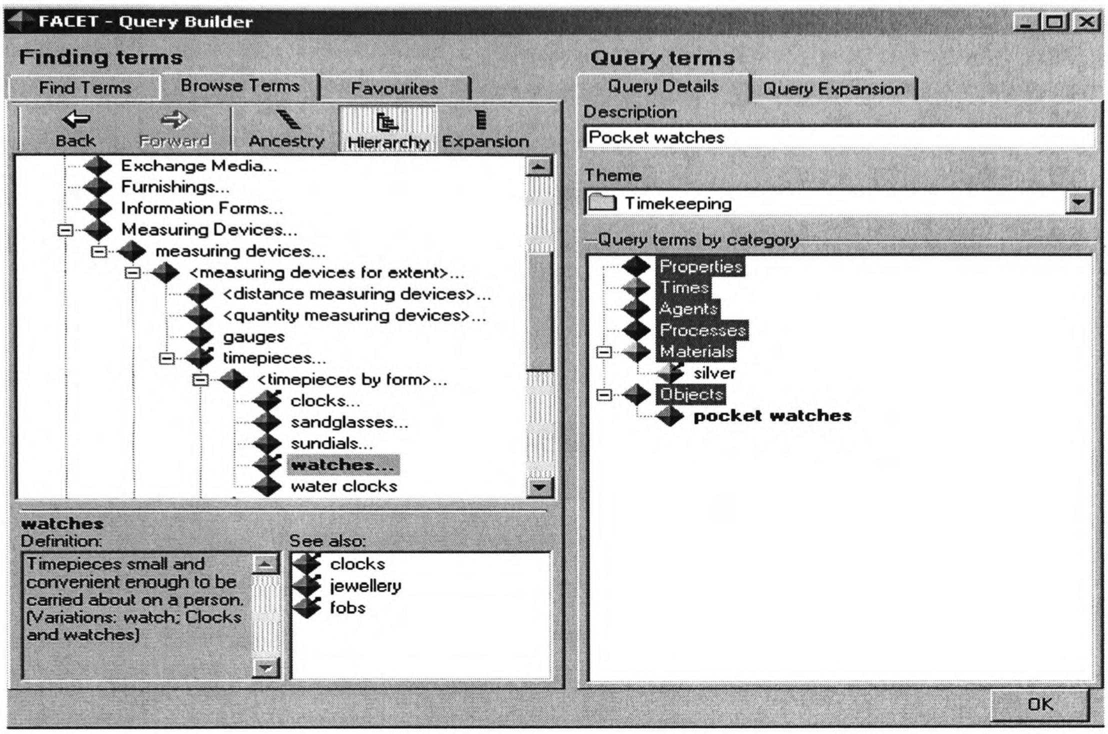


Figure 7.17 The Query Builder

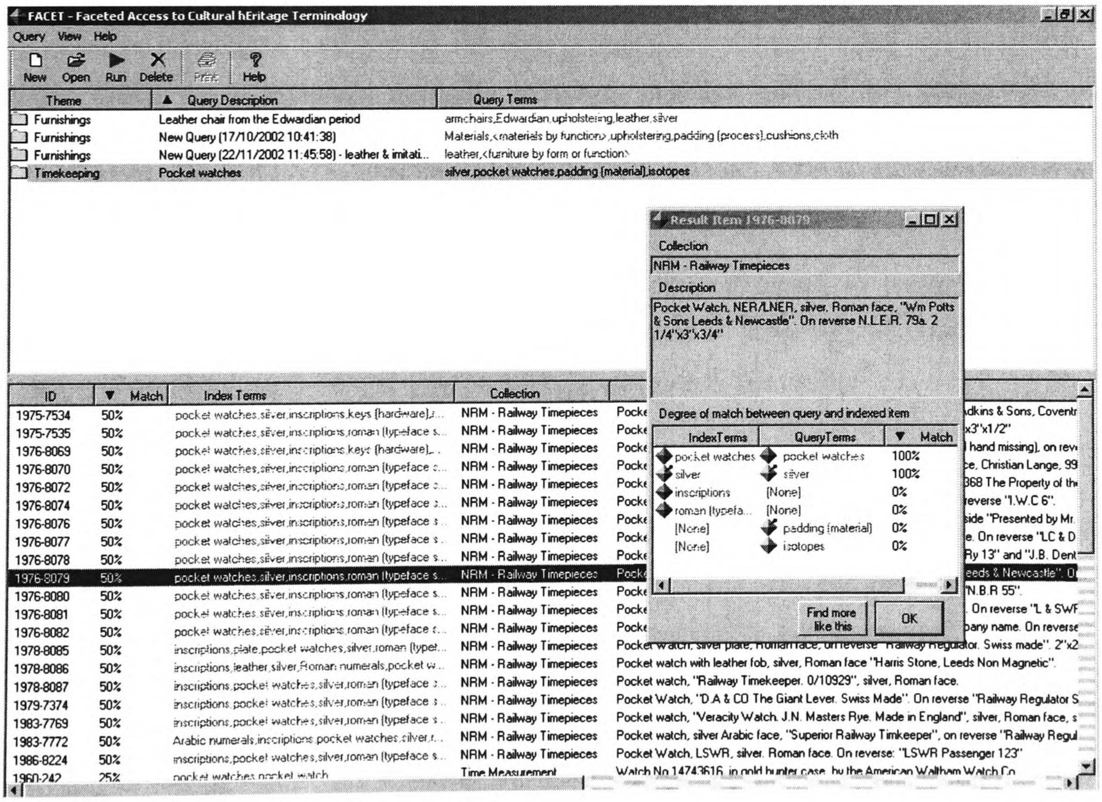


Figure 7.18 The revised FACET interface

Chapter 8 – FACET 2 study – set-up and data collection

8.1 Introduction

The following study was conducted with the second version of the FACET system as described in chapter 7. More participants were involved and the training was standardised compared to the study with FACET 1. Based on the experiences of that study, methods which had proved to be useful were modified, and additional methods, such as questionnaires and semi-directed interviews, were employed in order to collect data related to information searching behaviour outside of the search sessions. The study looked more on the information searching process as a whole and took a specific interest in how the thesaurus informs this behaviour.

8.2 Objectives

As in the FACET 1 study, the author was interested in investigating how a thesaurus informed the search process and in identifying problems associated with the different stages. The previous study suggested certain beneficial uses, such as local browsing, as well as approaches better to be avoided. However, there was still an interest in finding out more about how users approach the thesaurus and how else it could contribute to the search. These data were used to correct and extend the preliminary model of the search process in controlled vocabulary-based interfaces which was based on the data collected in the FACET 1 study and which is described in chapter 10 and 11. Additional data were collected to investigate connections between personal search approaches and conditions under which the participants normally searched. Another aim of this study was to evaluate the revised version of the FACET interface. Improvements made due to findings from the last study were to be assessed, for example whether the search process was now better supported and whether related problems were reduced. The interface here also included user-adjustable query expansion, which was anticipated to create additional issues. After the last study, evolving scenarios seemed to be a way of controlling the collection of useful data and this method was put to a test in this study.

8.3 Study set-up

This study was conducted under very similar conditions as the FACET 1 study described in chapter 6. The revised version of the FACET system was installed on a laptop which the author took to the participants, who again corresponded to the anticipated user group of a system like FACET.

8.3.1 Participants

Altogether, 13 people took part in this study, two of whom had previously used the system during the FACET 1 study. Four were associated with libraries, the remaining nine with museums, where some worked as curators and others were involved with museums documentation. Ten out of 14 participants had been in their current or a similar job for nine years or more. The remaining participants had been employed between ten months and seven years part-time. When they searched for information, this was often for other people, which made the study set-up as realistic as possible for the participants apart from the fact that they only had a short time to familiarise themselves with the interface. The library professionals were selected for this study because of their more formal experience with database systems and searching, while most of the museums professionals were familiar with the collection.

8.3.2 Set-up

During some of the sessions, a second evaluator was present, and during one session, two participants worked together. Six sessions were conducted within two days, and a general overview was given to this group rather than having individuals read through the information. The introduction was followed by individual hands-on training sessions. Instructions were presented in a PowerPoint file, demonstrating the search process in small steps and then requesting participants to repeat them in the interface. After the training, participants filled in a background questionnaire. They then completed two tasks which the author had prepared. At the end of the session, the author discussed the searches and interface with the individual participants, using semi-structured questionnaires for guidance.

Again, no time restrictions were imposed for the training or the tasks. The participants had been informed that the sessions would take about one hour and a half to two hours including the set-up. In some cases, time restrictions due to the participants' commitments applied. In these cases, one task was normally left out.

8.3.3 Training

During the FACET 1 study, participants had been given an overview of the system followed by a task with instructions to work through. For this study, both the introduction and the training were formalised using PowerPoint presentations. The author made as few additional comments as possible so that participants all received the same information. The first part of the instruction gave an overview of thesauri, indexing and the stages of the search process as identified especially from Marchionini's model (sections 3.2.1 and

6.6.5) and adapted to FACET. The stages of identifying keywords, searching for terms in the thesaurus, constructing and executing the query, judging the results and reformulation were then applied in the hands-on training to the FACET system. Step by step, the training presentation repeated information regarding the search process at that particular stage (example in figure 8.1) and showed appropriate interactions with the relevant components of the interface using images and video (example in figure 8.2). After a search stage had been demonstrated, the participants were asked to complete the same interactions in the interface and the author brought up the FACET interface. A handout, which could also be used for reference during the searches, reminded participants of the steps to take for each stage. After completing a search stage in the training, the author switched back to the presentation for information on the next stage. The training was structured like this in order to re-enforce the search process further. The handout and an outline of the training session can be found in appendices 8.1 and 8.2. Below in figures 8.1 and 8.2 are examples of screen dumps from the training presentation. As the presentations relied on animation and video, it was not feasible to include all slides in this thesis.

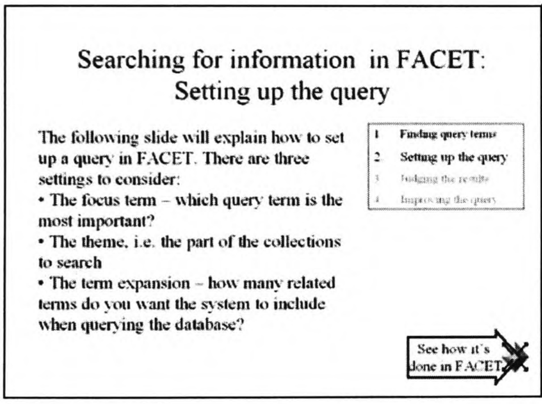


Figure 8.1 Information on setting up queries

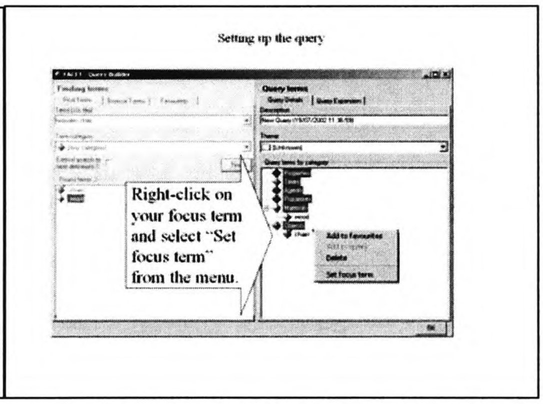


Figure 8.2 Setting Focus term in FACET 2

8.3.4 Tasks

Altogether, four different tasks were given to the participants (details in appendix 8.3):

- searching for chairs from the late 19th century
- searching for a record which corresponds to the image of a clock
- finding out about different materials used to upholster chairs
- alternative task for the curators to search their own collections

The tasks were based on experiences and comments from the FACET 1 study. Here, the image of an object had been given to participants. Overall, this task had worked well, but for the FACET 2 study, the image of a chair was replaced with a clock as it contained more easily identifiable features. This task became the standard task for all participants. The descriptive task about a 19th century chair was dropped after two sessions as it proved too complex to express the time period using the AAT and this issue was not of major interest to the investigation. The task was replaced by one asking participants to find out about different materials used to upholster chairs, which was designed with the system's query expansion feature in mind, and constituted the second task. The museum professionals who were responsible for their own collections did not complete this task, but were asked to execute a search of their choice within their collections as this might be interesting for them and could reveal information that related to the fact that they knew the collections well. In cases where this would not have been possible, they were to be given the "upholstered chair" task. Section 8.5.1 details tasks completed by each participant.

After the experience made in the FACET 1 study that participants sometimes completed tasks in ways that did not yield interesting data, the author had chosen to use evolving scenarios in this study (section 6.8.3). Participants were thus told that they might receive more information on the tasks as they went on.

8.4 Data collection methods

During the FACET 1 study, many of the data collection methods used proved to be useful in identifying problematic incidents and underlying reasons, so the author essentially employed the same methods in this study. Camtasia was used to capture events on the screen during search sessions, and after problems with the tape recorder, the software was also used to record the discourse between the author as the main evaluator and participants. The video and audio data was thus automatically synchronised which made them easier to work with. The author also made evaluator notes after sessions.

During the previous study, it had not been considered desirable to collect demographic data on the participants, partially because enough seemed to be known about them. However, the findings indicated that is was necessary to find out more about the participants' background with regards to information searching. A two-page self-administered questionnaire was designed for this purpose, containing a mixture of closed and open questions (appendix 8.4). The questions first tried to establish a general computer use, for example office packages, multimedia and search-related software, and then asked more specifically about search training and experience with thesauri.

A post-search discussion was prepared in order to clarify participants' answers to the questionnaire and to talk about the search session. The discussion was based on a number of open-ended questions which were adapted depending on the situations (appendix 8.5), as participants might already have answered a question, for example by expressing some difficulty with the interface. Participants were asked to assess the level of difficulty of using the system in general and the use of the adjustable semantic closeness expansion in particular. They were also specifically asked to compare their usual search activities to the tasks completed in the search session. The remaining questions aimed at probing information on additional functionality and support.

8.5 Data analysis

The data analysis conducted for this study is essentially the same as for FACET 1, described in 6.6. Audio and screen capture files were transcribed and the transcripts inspected for incidents, which were categorised using open coding. Incidents of particular interest were identified and data from different sources collated.

However, some of the methods used for the FACET 1 study were developed further, for example the graphical representation of search approaches and these were examined from different perspectives using a number of additional methods based on the post-processed log files. The log files contained more complete information on the users' interactions and could thus be used in more different ways, for example to extract some statistical data and to create a summary of the sessions.

8.5.1 Log processing and analysis

The examples in figures 6.3 and 8.3 show that log files were in themselves not easily understandable. Although changes in the interface resulted in different log entries, this was still the case. Some of the data were not necessary for certain types of analysis, so they were post-processed to be used more efficiently. For post-processing, the text files were imported into Excel and then processed using a Visual Basic 6 program the author wrote especially for this purpose. It mainly filtered out the interactions of interest which the author identified by inspecting original log files. The log files were processed to generate four different types of output data. The summary provided an overview over the interactions performed, the transcript aid converted log entries into English sentences, the statistical data provides for example the length of time a participant spent on a task, and the final processing type extracted free text and thesaurus terms from the log.

Summary

This type of output was similar to that used in the previous study with FACET. It only displays the most important interactions, and acts as a quick overview over the session. They effectively represent the sequence in which a user moved through the interface. This data is stored in Excel so that graphs can be created, which is explained in section 8.5.2.

Transcript aid

During the previous study, transcribing the screen capture files was very time-consuming and tedious. Therefore, the log file was post-processed so that it identified those interactions which were previously transcribed from the screen capture files and generated grammatically correct, readable sentences. Figure 8.3 shows an excerpt from the log files which is translated in figure 8.4:

```
09:54:24.039   ctlTermFinder.cmdFind_Click   <DETAIL><FINDWHAT>wooden
chair</FINDWHAT></DETAIL>
09:54:25.691   ctlTermFinder.FindTerms       <DETAIL><FINDWHAT>wooden
chair</FINDWHAT><FOUND>2</FOUND></DETAIL>
09:54:42.254   ctlTermFinder.tvw_MouseDown
        <DETAIL><CLASS_ID>11914</CLASS_ID><TERM>wood</TERM></DETAIL>
09:54:48.055   ctlQueryParameters.GetExpandedTerms       <DETAIL></DETAIL>
09:54:48.195   frmMain.m_QUERIES_Changed
        <DETAIL><QUERYID>20021017_095342.xml</QUERYID></DETAIL>
09:54:48.195   frmQueryBuilder.ctlDetails_DragDrop
        <DETAIL><QUERYID>20021017_095342.xml</QUERYID><CLASS_ID>11914<
/CLASS_ID><TERM>wood</TERM></DETAIL>
```

Figure 8.3 Excerpt from log file

09:54:24.039 THE USER SEARCHES THE THESAURUS FOR TERM 'WOODEN CHAIR'.
09:54:25.691 TERM FINDER RETURNS 2 CANDIDATE TERM(S) FOR TERM 'WOODEN CHAIR'.
09:54:42.254 THE USER CLICKS 'WOOD' IN THE LIST OF CANDIDATE TERMS.
09:54:48.195 THE USER DRAGS THE TERM 'WOOD' TO QUERY 20021017_095342.XML.

Figure 8.4 Excerpt from log file translated into grammatical sentences

The resulting file was then modified where necessary to include aspects from the screen recording which the log file did not contain. For the last two search sessions, these output data were available before the audio was transcribed. The audio transcription was thus also included in this file, so that the data are fully collated.

Statistics

The statistical data was stored in text files and recorded how often certain interactions, for example changing the expansion rate, adding or deleting query terms, occurred during a session. This data was then used in order to identify patterns (see sections 9.3 and 9.4).

Free text and query terms

In order to identify patterns related to the free text terms entered into the Term Finder and the query terms actually used, this post-processing option identified all free-text and thesaurus terms used during the sessions. This data was used in the analysis described in section 8.5.4 – Use of concepts and their origins.

8.5.2 Graphs of the search process

From the post-processed log files, graphs were generated in order to analyse the search process. This was done visually as an automatic analysis was deemed too complex to implement, as a certain amount of "fuzziness" has to be allowed for. Users could for example search for any number of thesaurus terms, or take any number of thesaurus-browsing steps, before moving on the next type of interaction. Activities such as searching Term Finder or browsing the thesaurus can easily be identified as such using the graphs.

In order to generate the graphs, interactions and components recorded in the log files were represented by numerical values. As several interactions can apply to each component, these are distinguished by using values with one decimal point, activating a form for example is 0.1 and double clicking a term 0.4. These values were added to the value allocated to the component in creating the final values. Based on these values, Excel can draw graphs representing the interactions which occurred in the search session.

Values between 1 and 11 were first allocated arbitrarily to components, but were then adjusted in order to generate clearer graphs. Essentially, the numbers thus present the search process from searching for a term in Term Finder via constructing and executing a query to inspecting individual results.

The following paragraph is a description of a search session with the corresponding graphical representation. This graph has been simplified in that interactions are numbered and these values rather than the actual time stamps used as x axis values. This means indications of pauses are lost, but interactions can be referenced more easily.

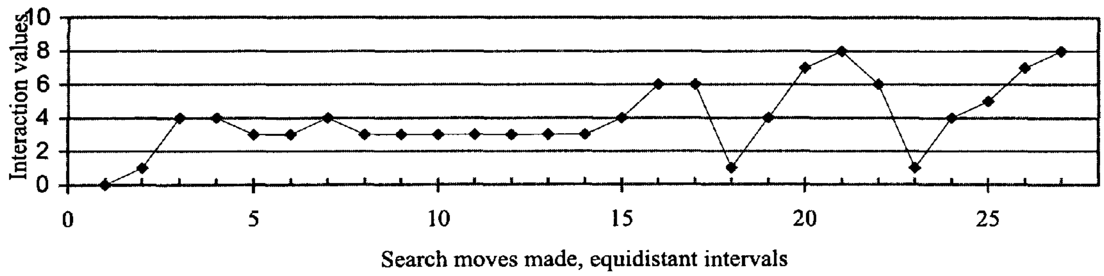
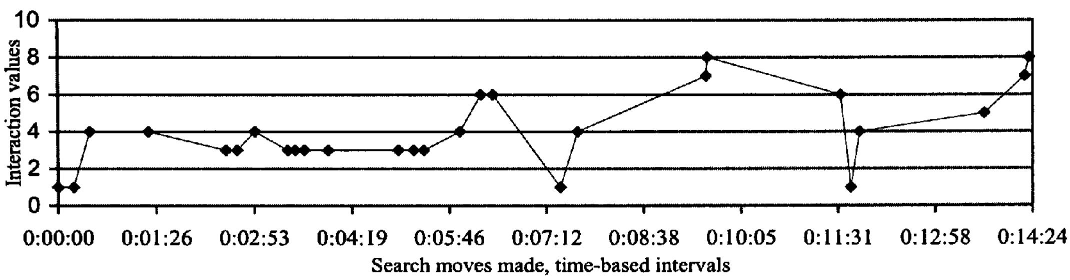


Figure 8.5 Example for sequence of interactions presented in equidistant intervals

Legend for interactions in figures 8.5 and 8.6

0 – creating new query
0.5 – "more like this option"
1 – Term Finder search
2 – double-clicking candidate term
3 – double-clicking thesaurus term
4 – dragging a term into query
4.2 – adding a term via menu
5 – changing the expansion
6 – deleting a term
7 – running the query
8 – results are returned
9 – opening a record
10 – displaying an indexing term in the thesaurus
11 – adding a term from a record to the current query

The searcher creates a new query (x=1) and searches the thesaurus for "moquettes" (2). She drags this term and "furniture" into the query (3 and 4) and browses the thesaurus starting from "furniture" (5&6, 8-14). She adds more terms to the query (7, 15), some of which she deletes again (16, 17). She then searches the thesaurus for "upholstering" (18) and drags this term into the query (19). She runs the query (20&21), and then deletes "moquettes" (22). She searches the thesaurus for "leather" (23) and adds the term to the query (24). She changes the expansion value to 70% (25) and runs the query again (26&27).



8.6 The same graph time-based intervals, starting at 0:00

Above is a graphical representation of the same session. This time, the temporal dimension has been maintained, showing faster and slower sequences of interactions. This type of graph was primarily used for the analysis.

Full descriptions of each search session can be found in Appendix 8.6, as it is not feasible to present all the data within this chapter.

Certain interaction, for example searching for a term in Term Finder, dragging it into the query and executing it show distinctive patterns which can easily be identified when inspecting the graphs (x=18 to 21 in figure 8.5, interactions between x= 0:07:12 and 0:10:05 in figure 8.6).

In the FACET 1 study, graphs had only been created in an attempt to assess the possibilities of data analysis, however they played a more important role in the analysis for this study. Based on these graphs, the author studied the users' steps before and after the first query run and also identified where changes in the query expansion took place. This information was then used to create a different type of graph described in connection with the analysis of the search process (section 8.5.4 – Types of searching behaviour).

8.5.3 Analysis of significant and problematic incidents

As in the FACET 1 study, audio and screen capture files were transcribed and the author identified incidents of relevance which were coded using open coding. The difference to the FACET 1 study is that this time, the author considered all incidents, rather than concentrating on breakdown and difficulties. The reason for this was that she wanted to ensure that all significant incidents were covered, and even though they might not individually be considered a real issue, several similar incidents might indicate potential problems.

8.5.4 Analysis of search approaches/processes

Results for the following analyses are described in sections 9.3 and 9.4.

Comparison of sessions

During the analysis of FACET 1 data, the author compared approaches to sessions for each participant and along tasks. In this analysis, sessions were again summarised and the author was especially interested to compare data from FACET 1 and 2 for the two participants who had been involved in both studies.

Types of searching behaviour

As already observed during the FACET 1 study, different participants had different approaches to searching. The comparison of sessions again revealed differences. To analyse these, the author examined the sequence of interactions in graphs based on post-processed log data. She distinguished for example between types of initial query construction and different reformulation behaviours.

Summary graphs of the search sessions were used to prepare data for the analysis of the search process on a more abstract level. The author coded the interactions according to Bates' (1979b) tactics and Fidel's (1985) moves which have been paraphrased in Appendix 8.7. First of all, applicability of the moves and tactics in a system using thesaurus expansion rather than Boolean and probabilistic retrieval was assessed because not all of them applied to the FACET system, for example those which intersect sets from different queries. Others refer specifically to the use of free text as opposed to thesaurus terms, for example Fidel's "Weight 3 – limiting a free text term to a pre-defined field". Some moves did not apply in their particular form, but in principle expressed similar moves to those which could be performed in this system. Fidel's "Include" is the equivalent of including the Narrower Terms of a term in the query. In the OVID interfaces, this would be done by checking "Explode" (see section 5.4.1. – Information on the Ovid Medline interface). In FACET, the term expansion would be increased, although this also includes, in addition to Narrower Terms, Related and Broader Terms. Considering that the general idea is to include semantically close terms, this move was used in the classification of interactions. Appendix 8.8 shows the frequency of moves and tactics identified during the search sessions. Table 8.1 shows the moves and tactics identified for some of the user sessions.

Task	Upholstery	Upholstery	Clock	Clock	Clock	Clock	Clock
Participant	A	B	C	D	E	A	F
Interaction 1	include	more like this	t2	s2	s2/add3	add3/s2	s5
Interaction 2	add3	s3		s2/add3	s5	include	t1
Interaction 3	add3	include		s2/add3	s2/add3	more like this	s2
Interaction 4	add3	s3		t9			[include]
Interaction 5	add3	add3		t2			s2
Interaction 6		add3		refine/t6			include

Table 8.1 Examples of participants' interactions, coded with abbreviations for moves and tactics paraphrased in appendix 8.7. E.g. "add3" indicates adding terms from a record, "include" increasing the expansion rate, and t-values indicate term replacements.

After the interactions had been classified, the author allocated values to them in order to create another set of graphs which solely present the query construction and reformulation. Several moves refer for example to replacing terms, so they were grouped into four categories, adding, replacing and deleting terms, and changes to query settings. Interactions were considered at this level so that for one, possible classification errors when using moves and tactics would not actually change the data, and secondly to indicate the similarity of these actions. Creating the graphs in this fashion allowed a quicker judgement of which of the main interaction options had been chosen and this also allows further generalisation regarding the interaction types. Figure 8.7 shows an example of such a graph used for the analysis.

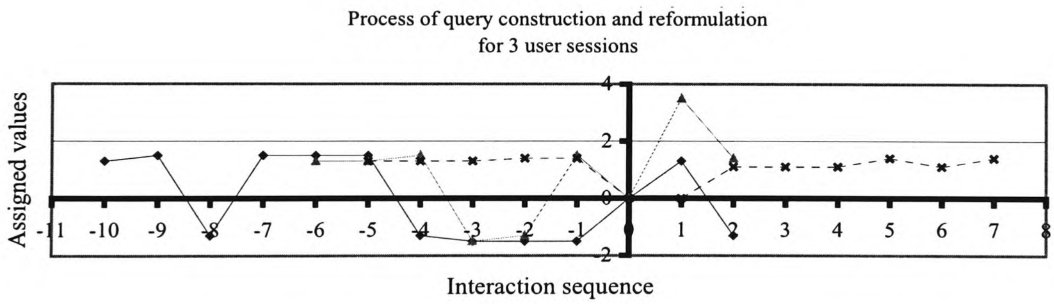


Figure 8.7 Example of graphs generated for 3 user sessions based on tables like table 8.1, to show whether terms were deleted, added or replaced, or a change made to the expansion.

Legend for assigned values in figure 8.7

-2 to -1 – Deleting terms
1 to 2 – Adding terms
0 – Changes to query settings
3 to 4 – Replacing terms

The exact moves represented by each value are detailed in appendix 8.8.

The first query is executed at x= 0, so interactions on the left of this point represent the construction of the first query and interactions on the right query reformulations.

The influence of various factors on the search process was also examined. These included the participants' professions, the success of their queries and their attitudes or frame of mind during the search. For each of these analyses, values were identified for each session, user data was regrouped accordingly and similarities examined for every group. However, no particular conclusions could be drawn from this analysis and the analysis itself is therefore not discussed in more detail either.

Use of concepts and their origins

Many studies of thesaurus-based systems examine the origin of query terms in order to assess thesaurus use and neglect. Other studies investigated the origin of query terms and their benefit in retrieving records (e.g. Spink and Saracevic 1997). In the context of a system where a query can only be constructed from thesaurus terms, and where relevance assessments only play a minor role, this investigation is of less interest, even though terms can originate from a number of sources within the interface (namely Thesaurus Finder, thesaurus browsing, records and existing queries). The origin of the concepts represented by query terms is more significant, as we are looking at the contribution of the thesaurus to the search process, and one of these roles is being a tool for the suggestion of terms or even concepts. The analysis is described below and section 9.4.2 will look at how concepts and free text terms are affected by having to select from the thesaurus in order to construct a query.

In order to identify the origins of concepts, the author first of all identified the first occurrence of each concept in audio transcripts and log files as well as the exact word used. The author also noted whether this term had been used for the first query, a later query or not at all. The author then allocated each of the first occurrences of a concept to a source. This information was also coded according to whether the concept was used in a query straight away, later on during the search or not at all. The author then assessed the number of concepts from each source and counted how many of them had been used in queries. She then also investigated to which extent participants exhausted the concepts because she noticed that some participants did not employ all the concepts they identified during their search sessions. The sessions were grouped accordingly and factors which might have influenced this behaviour examined. The data are discussed in section 9.4.1.

Free text and thesaurus terms

Usage of the thesaurus terms was of particular interest, thus they were analysed together with the free text terms extracted from the log files. Analysis was conducted of:

- free text terms entered into Term Finder by each participant for each task
- number of free text terms entered by each participant for each task
- frequency of free text terms used across participants
- thesaurus terms used in queries by each participant for each task
- number of thesaurus terms entered by each participant for each task

- frequency of thesaurus terms used across tasks

These values were considered in combination with the number of query runs and number of terms added and deleted from queries. Terms were also grouped by concept, i.e. if one participant used "wall clocks" and another one "clocks", these were considered to be one concept and their frequency across all participants was investigated. Due to factors specific to individual sessions, it was not deemed reliable to conduct automated statistical analysis on these data. For example, one participant who was experimenting with the interface added and deleted an exceptionally high number of terms; some queries were based on records ("More like this") and allowed participants to delete more terms than they actually added. The author felt that it was important to consider these factors during the analysis and was also not interested in obvious causal relationships such as "more thesaurus terms mean more query runs", as participants who modified their queries logically had to run more queries to evaluate the effects.

8.5.5 Evaluator interventions

The author extracted all interventions she had made during sessions with their context from the audio transcripts. These were categorised with the same approach used for the incidents in order to identify situations in which support might be required (see section 9.5.2) and fed into the model discussed in chapters 10 and 11.

8.6 Participants' searching background

This section and the following are intended to give a background of the participants' computer use, search experience and their day-to-day information searching tasks. The latter are presented in section 8.7 by contrasting them with the tasks from this study.

8.6.1 General computer use

	Daily	Weekly	Monthly	Occasionally	Never
Email	14	-	-	-	-
Word processing	11	1	-	2	-
Games	-	2	-	5	6
Multimedia	-	2	1	6	5

Table 8.2 – General computer use - replies per category

All participants stated in their questionnaires that they used email on a daily basis and word processing programs were also used frequently (table 8.2). They also indicated a number of systems used regularly for searching (table 8.3). Overall, they seem to have a good command of computers even though they might not all have been too keen on using them. One participant admitted a certain reluctance to them, but said she saw their usefulness for her work. Her attitude did not seem to have a detrimental effect on her searching in the session.

8.6.2 Search experience

Training

10 out of 14 participants had taken courses in online searching either whilst at university or as part of their professional development. Those who had not completed courses appear to do less searching, but in the post-search discussion, there was an indication that this is related to their work. Within this context, they might not require formal search training.

Regular searching

Apart from one person, who also indicated that she did not enjoy searching much, all participants searched on a daily or weekly basis. The museums professionals spent up to five hours/week searching, that is much less time than the librarians. Dealing with information being the main purpose of their profession, one subject librarian even searched 30 hours a week. The exact amount of time depended for everyone on the specific work requirements. Interestingly, the reasons for searching are similar for librarians and museum professionals (Appendix 8.4 – Self-administered questionnaire, question 3.1). The most common reason stated is searching on behalf of somebody else, mostly a client but sometimes a colleague (7 out of 14). Based on conversations between the author and participants, this number is actually probably higher (10 out of 14). Five participants searched for information as part of their work, and two also mentioned searching for personal use.

Applications used

On a daily and weekly basis, the Internet and local databases are the most commonly searched applications. Other commonly used applications include web databases and computerised Library Catalogues (also referred to as OPACs). The librarians primarily used the later ones, while the museum professionals mainly relied on local databases and the Internet as sources of information. This difference is related to their work domains.

	Daily	Weekly	Monthly	Occasionally	Never
OPACs	5	1	1	6	-
Local databases	11	1	0	2	-
Web databases	5	2	3	2	2
Internet for research	6	2	-	4	1
Internet	10	1	2	1	-

Table 8.3 Applications used for searching - replies per category

Experience with thesauri

Part of the questionnaire had asked about previous thesaurus use, but these questions seem to have been misunderstood by most participants and were thus not analysed. Only very few people seem to have used thesauri in a previous employment or work role, but most of them were currently using thesauri in searching for information. Out of nine replies, five participants stated that they used a thesaurus on a daily or weekly basis, and only one person said they did not use a thesaurus. The most-mentioned thesaurus was the AAT, which is integrated with the museum's collection management software. A number of specialist thesauri were also mentioned.

8.7 Sessions versus real-life searching situations

8.7.1 FACET compared to interfaces participants are familiar with

FACET differed in a number of ways from the systems participants normally searched. One of the main differences is use of Boolean operators in the participants' systems compared to semantic expansion and ranking of results. Another difference was constructing a query using fields rather than facets. Minor differences include use of wild cards and being able to do free text searches on the database. The latter points were only mentioned during the post-search discussion when the author enquired about the systems the participants had experience with, while the differences related to use of Boolean logic and fielded searching can actually be noticed when analysing searching behaviour and comments that reflect the participants' reasoning and expectations (see section 9.8.1).

8.7.2 Differences in thesauri

MultiMimsy, the system used by the museums professionals, allows users to type in a thesaurus term and see its local context. Some participants mentioned that they did not think the same types of relationships, for example associative ones, existed between terms. It is probably a matter of representation though, and the system also does not seem to make use of links between terms. The AAT is available on the web, where navigation of it is easier. The person who mentioned this preferred to use this version as it was also more up-to-date. The Waterways thesaurus is used in the museums in addition to the AAT and provides thesaurus terms in domains where the AAT was not detailed enough. This thesaurus only contains object names, and attributes are entered into records as free text terms. During the last study, participants' searching behaviour had been influenced by this approach (section 6.8.2).

Some of the interfaces used by the library professionals resembled FACET in that they mapped free text terms to thesaurus terms and made suggestions, similar to what is described for the Ovid interface to Medline in section 5.4.1.3.

8.7.3 Topic familiarity or domain knowledge

Most participants had no specialist knowledge of clocks or the upholstering of chairs. Some of the museum professionals were familiar with the relevant collections though. One of them commented that depending on the size of the collection, they need not even search because they know their items so well. A librarian highlighted the importance of domain knowledge by pointing out that over time, they learn about a domain, but when moving to a different one, they become almost like novice searchers again. It appeared that both in libraries and museums, professionals could be experts in the domain they work in or deal with a subject area they have no extensive knowledge of.

8.7.4 Normal search tasks and approaches

Through additional data collection methods such as questionnaires and post-search interviews which contained questions specifically about the participants' searching as part of their day-to-day work, it was possible to gain an insight into potential influences on searching during the sessions. When looking at participants' normal tasks related to searching, the author discovered that in both contexts, some people searched for information, while others were also involved in indexing. In the museum, the people involved with documentation for example specifically searched the thesaurus for terms to complete the indexing of records. Thus some might have more contact with thesauri and indexing or use thereof.

All participants in this study searched mainly for others, either clients or colleagues. Thus, set tasks correspond to their standard search tasks. Although the author had pointed out to participants that they could ask questions, some participants felt though that the set tasks differed from these in that they would normally clarify the enquiries in a dialog with the client. Evolving scenarios were partially designed to reflect this process, but proved difficult in practice (see Conclusions, section 12.2).

The kind of enquiries the museums professionals received tended to be along the lines of "I have this type of object, is it valuable?", "Where can I get this repaired?" or "Can you tell me more about this?". Subject experts are likely to be able to answer most of these questions based on their knowledge and without having to execute a search (see 8.7.3), especially as many systems are probably not even designed to answer them. In those cases where they have to execute a search, they normally enter a term describing an object, as within the context of a museum this tends to be the focus of enquiries. Other terms might be added to be more specific. These generally include the maker's name, rather than properties such as the material. A museum professional commented on the fact that text on the clock face which was probably the maker's name, was illegible in the print out and they were not told what it said. However, the FACET system does not support these types of searches.

In cases where clients wanted to borrow objects, curators run general queries to generate an overview of available items. From this list, the clients then judge the potential suitability of objects and inspect them to determine whether they really fulfil their requirements. Database searching is effectively only the first step of this process, which is probably why some participants took a more general approach for their search and spent more time browsing through records rather than reformulating their queries, which was particularly noticeable during the FACET 1 study (section 6.7.2). Given the option, museum professionals would thus do less database searching and move sooner to looking at either photographs or actual objects to verify they had found what they were looking for. One curator explained that in those cases where they actually had to conduct searches, it was clear that if the information could not be found in a few obvious places, it was likely that the request would take too much time and clients were not normally willing to pay for the cost of this, which presumably makes extensive searches of this nature unlikely.

The library professionals normally dealt with textual information, documents or document surrogates, rather than with actual objects. One commented that although in principle, it was the same approach, the object appeared to be more specific. Somebody else commented that searching based on an image was more vague than looking for a book, but admitted that it was similar compared to a vague enquiry where a client might not know authors or key references for a subject. One librarian indexed using the Library of Congress Subject Headings (LCSH). Although indexing with subject headings normally requires indexers to use a smaller number of identifiers than indexing with thesauri, she explained that they were not restricted to using fewer rather than more subject headings. Potentially, familiarity with this practise might have influenced this person to use fewer query terms.

Chapter 9 - Description of FACET 2 - data and findings

9.1 Introduction to the findings

The previous chapter has presented the data collection and analysis. From this point onwards, the findings are presented, starting with general observations on the search sessions in the form of statistical data, descriptions and graphical representations of log files (9.2). The overall search approaches are discussed in section 9.3. The origin of concepts used is analysed in section 9.4 together with the influence of the thesaurus on transforming concepts into thesaurus terms. The incidents which had been extracted and classified are then discussed under the three headings of training (9.5), thesaurus search (9.6) and query construction and reformulation (9.7). Query construction and reformulation are discussed together as similar issues occurred during both stages. However, it should be noted that from a conceptual point of view, query reformulation is more complex than the construction of the first query. Finally, this chapter addresses technical issues in the system and the thesaurus which potentially influence the success of information searching behaviour (section 9.8). The methodology is reviewed separately in connection with the conclusion on the methods used for all studies carried out for this thesis (section 12.2).

9.2 Description of sessions

9.2.1 Tasks completed by participants

The following table shows an assessment of how successfully each participant completed which tasks. Overall, participants used at least appropriate strategies to complete tasks and all sessions apart from the two labelled with "1" resulted in the identification of suitable results that the participants were satisfied with. The curators' own searches were difficult to judge as no comparison was possible, but their approaches were appropriate considering their aims, the records in the database and limitations of the thesaurus used, which are discussed in section 9.8.4 (Detail of the thesaurus structure).

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
19th century chair	1	1	-	-	-	-	-	-	-	-	-	-	-	2
Clock image	2	2	2	2	3	2	2	2	4	4	2	2	4	13
Upholstered chair	-	-	2	4	3	2	2	3	2	-	-	-	-	7
Alternative task	-	-	-	-	-	-	-	-	-	-	5	5	5	3

Table 9.1 Tasks completed by each participant, indicating levels of success

Legend for table 9.1:
1 – Task too complex, terminated by the author*
2 – Task completed successful
3 – Task conducted appropriately and the author decides to move on to the next task
4 – The participant feels they have completed the task and indicate they want to move on
5 – Curator's search in their own collection, difficult to judge, but appropriate approaches. Two curators actually completed two tasks.

* The author was again the main evaluator in this study.

9.2.2 Session statistics

Altogether, 13 sessions resulted in 13 hours of screen recording, most of which includes synchronised digital audio recording, while about two hours of audio were recorded on tape. The length of recordings ranges from 26 minutes to 1:10 hours. This includes in part the post-search discussion and filling in of the questionnaires. Based on the log files, the total length of sessions was calculated to be about 11 hours, ranging from 10 minutes for an exceptionally short session to just over one hour for several longer sessions. As anticipated, participants took up to half an hour for the training.

Some tasks included aspects that were more complex (e.g. expressing time – see section 9.6.1) and some participants took an interest in examining the interface, so that the time taken for a task does not necessarily indicate the amount of difficulties participants encountered.

9.2.3 Overview of the sessions

Participants used one of two strategies before running their first query.

- Entering terms describing several concepts, one of which expressed the main focus of the task (clock, chair or upholstery), into Term Finder.
- Entering one concept into Term Finder.

Afterwards, they then either browsed the thesaurus or dragged terms directly from the candidate list into the query. Most participants who had searched for several concepts also selected several thesaurus terms.

Participants who only entered one concept sometimes searched for further concepts individually and added these or semantically close terms to their query. Before running their query, some participants changed the term expansion. After running the query, participants normally inspected the results either by looking at the list of records or by opening individual records. If they reformulated their query, participants followed either one or a combination of the following options:

- Searching the thesaurus for terms they meant to add initially or that they felt they could add because the results were not yet specific enough. These were for example identified by looking back at the task.
- Searching the thesaurus for terms that they either saw in records, or where records or object descriptions gave them the idea.
- Replacing a query term with a term that occurs in a record (e.g. "wall clocks" for "clocks", "carving" for "carvings".) This required them to delete and add terms.
- Removing unsuitable terms.
- Changing the query expansion.

- Using the "More like this" option to create a new query based on indexing terms from a record.

Most participants reformulated their query more than once, and they often applied various different strategies.

9.3 Discussion of search approaches

During the previous FACET study, striking differences in the task approaches were identified (section 6.7.2). During the analysis of these data, the approaches were again examined (section 8.5.4). Participants often used similar interaction sequences based on basic interactions as discussed in section 9.2.3, but overall, differences were much more subtle. From the descriptive graphs of the sessions (section 8.5.2 and appendix 8.6) and free text terms extracted from the log files, three approaches to constructing initial queries were identified. Some participants searched the thesaurus for a single concept and ran their query containing only one concept. Others repeatedly searched the thesaurus and added several concepts to their query. Others searched Term Finder only once, but for several concepts simultaneously. In some cases, they then selected several thesaurus terms, in others only one.

Section 8.5.4 explains how the author processed the application log files and coded query modifications based on Fidel's and Bates' moves and tactics. The results of this encoding are shown in Appendix 8.8. These moves and tactics were again represented graphically using similar methods as used for representing the sessions as a whole (see 9.2.3) and figure 8.7 shows an example of the kind of graphs representing the sessions. These graphs refer to query reformulation independent of interactions with FACET. The interaction graphs might show that a term has been added and a term has been deleted, but the query reformulation graphs indicate cases in which a term has actually been **replaced** by a semantically close term or logically equivalent (e.g. one material for another in the context of the "upholstered chair" task). The following types of query reformulation were identified:

- Adding terms – values between 1 and 2
- Replacing terms – values between 3 and 4
- Deleting terms – values between -1 and -2
- Changing the query expansion – 1.4 for increasing and -1.4 for decreasing the expansion, as this equals the removal and addition of query terms
- Using the "More like this" option to create a new query based on a record, which us specific to FACET – value = 4.5
- values of 0 indicate neutral moves such as changing the focus term or the collection.

It becomes clear that adding concepts to the query is by far the most common modification (61 out of a total of 118). The next most common move is modifying the query expansion. All but one participant changed the term expansion rate at least once. One change only was the most common (10 sessions out of 25), but some participants modified the expansion more often. Hence, all options for query modification were used, although "More like this" was employed by few participants. When looking only at moves after the first query run, adding terms from the record (13 out of 49 – abbreviation s2/add3 and add3) is the most common. Only few terms were actually replaced (8 out of a total of 118 modifications). Moving from a BT to an NT (3 – t2) and from an NT to a BT (2 – t1) made up over half of the replacements. In the "upholstered chair" task, other replacements which had to be classified in a more general way were found. It happened that users searched for one material (such as "materials") first and then for another (such as "leather") later, which cannot be considered as a refinement of terms in the same way as moving from BTs to NTs and vice versa.

Although no distinct patterns emerged, the use of concepts in the first query potentially influenced reformulation. When participants started with few concepts, their behaviour alternated between adding or replacing on one hand and deleting terms on the other, and was often related to the inspection of records. If searchers have included several concepts in their first query, reformulation might possibly be limited because fewer concepts remained to be tried.

The curators all approached their own tasks in a similar way. They started with only few terms (in four cases only one term and in two, two terms), and thus made more modifications after the first query run than before. These sessions have certain characteristics in common:

- The participants were familiar with the items in the collection, but not so much with the indexing of the records.
- They all felt that the thesaurus was not detailed enough for their domain.
- They all commented on the fact that the current indexing did not permit very detailed searches.

Likewise, the two sessions searching for "19th century upholstered chairs" also have similar characteristics – they are short, with the same amount of modifications before and after the first query run. This is not very surprising – this task was difficult to complete and was terminated by the author in both cases. However, these incidents point to an interesting issue: the influence of task-related issues on the search approach. More specific issues on query formulation are discussed in section 9.7.

Participants actually displayed a number of different attitudes or dispositions when first starting the sessions: some were concerned about remembering the elements in the interface, others took a more light-hearted attitude. Some participants were more reluctant to continue searching and thus interacted less with the interface, but not being sure how to conduct a search did not necessarily mean they interacted less than others. Kuhlthau (1988) investigated in more detail the affective aspects of information searching. In the short term as seen in this study, frustration appears to be the most significant of these emotions. Potentially, it leads to abandoning the search, which has been addressed in the Risk Points in chapter 11.

9.4 Origins, representation and use of concepts

Studies where participants have access to thesauri often investigate the origin of terms to assess thesaurus use or neglect. In order to identify the origin of concepts, the author examined the audio transcripts and log files for the first occurrences of concepts and noted their origin, for example suggested by the participant, seen in a record etc., as described in 8.5.4. The following were identified as sources for concepts:

- Task description. Different from for example Spink and Saracevic (1997), the task description has not been written by the searcher or end-user, but by the evaluator.
- Participant
- Term Finder
- Browsing the thesaurus
- Records (both where the participants selected terms and where the author pointed them out). Records also functioned as a source of ideas; participants did not necessarily use the concepts they saw, but used them as a basis to think of others.
- Expansion display
- Existing queries.

Concepts used in the query	Clock	Upholstered chair	Other tasks	**Total**
Task	13	10	5	**28**
User (5 user-based tasks and 13 set tasks)	19	1	8	**28**
Term Finder	-	-	-	**0**
Browsing the thesaurus	-	3	-	**3**
Records (User)	5	3	7	**15**
Records (Evaluator)	3	-	-	**3**
From records as an idea	-	-	2	**2**
Expansion	-	-	-	**0**
Existing query	-	2	-	**2**

Table 9.2 Origins of concepts

Table 9.2 shows the concepts which were actually used in the query. Term Finder and thesaurus as such were not a good source of concepts (0 and 3 concepts respectively, for all participants in all sessions), but these sources probably help the users in selecting terms to represent their concepts rather than generating ideas for concepts. Candidate terms are returned which do not reflect the concept, for example terms which contain only part of a phrase searched for or which might be homonyms, but they are unlikely to result in finding a completely new concept.

Records on the other hand actually include additional concepts related to the searchers' topic and can thus more easily inspire the search (20 concepts in total). 15 concepts were taken directly from the records. Following the evolving scenarios method, the author pointed out an additional 3 during difficult sessions and in two cases, the participants saw terms in records, which made them think of additional concepts. One participant started a search with an existing query and used some of those terms. In this study, most concepts stemmed from the task description (28) and the participants (28). 9 sessions were conducted using a textual task where participants could lift concepts straight from the description. 13 out of the 28 task-based concepts come from the clock task (13 sessions), where participants were given an image and verbally instructed to look for a clock. The curators conceived 5 tasks themselves, where they are the sole initial source of concepts.

9.4.1 Exhaustive use of concepts

Inspection of the transcripts revealed some information on when in the search participants first considered different concepts (see 8.5.4 for the analysis and table 9.3 for the data). In 16 cases out of 27, all concepts mentioned up to the first query run are represented by thesaurus terms in the query, although four participants mentioned only one concept up to this point[1]. In the remaining 11 cases, participants were aware of more concepts than they used. However, in only 45.4% of these sessions, participants added additional terms after the first query run to represent other concepts, compared to 87.5% for those participants who used all of their concepts in their first query. In those 11 cases, the participants also made fewer moves to refine terms (18.2% to 56.3% of participants who had used all their concepts for the first query), which is striking.

In the next step, the amount of concepts used was examined, as this could potentially have explained why participants who had not used all concepts in the first query added fewer terms later. However, most participants used one or two concepts for their first query (average of 2 for those who used all concepts and 1.45 for those who did not), so that the group of participants who did not use all concepts does not stand out for using larger amounts of initial query terms. Thus, the sessions were examined for potential

[1] In two cases, additional concepts were mentioned after the first query run, but not used.

	Only 1 concept idea before the 1st query	Used all concepts for 1st query (when >1)	Several concept ideas, not all used for 1st query	No further use of previous concepts	More concepts from records	Refines existing concepts
Clocks						
Participant A		X			X	
Participant B		X			X	
Participant C	X		X			
Participant D		X				X
Participant E	X		X			
Participant F		X				X
Participant G	X				(x)	
Participant H		X			X	
Participant I		X			X	X
Participant J		X			X	X
Participant K			X		X	X
Participant L			X		X	
Participant M			X			X
Upholstery						
Participant C			X			
Participant E	(x)	X				X
Participant F		X				X
Participant G	X					X
Participant H		X			X	X
Participant I			X			
Participant M			X			
Other tasks						
Participant A – attempt 1	X					
Participant A – attempt 2	X				X	
Participant B			X		X	
Participant D			X			
Participant K			X	X	X	
Participant L – attempt 1		X			X	
Participant L – attempt 2		X			X	X
Total	6 (7)	12	11	1	13 (14)	11

Table 9.3 Origins of concepts used in queries
The (x) indicates debatable cases.

reasons which might have stopped participants from including more concepts. In some cases, the search had probably been influenced by personal styles identified already during FACET 1. In three cases, the problem was related to thesaurus access, i.e. participants were unable to identify suitable terms. Better means of accessing time-related terms would have helped in all these cases. In one case, the thesaurus was missing the necessary detail in the subject area. In two other cases, difficulties were related to the task. Participants did not continue their search as they felt the tools were inappropriate for what they were asked to do; this applied particularly to the "upholstered chairs" task. Considering their central role in the retrieval process, it is not surprising that the controlled vocabulary or insufficient indexing can have a negative influence on the search session. This highlights the need for good access mechanisms to thesaurus terms and high-quality indexing. It should be noted though that these problems also occurred in sessions where participants added more concepts. On the other hand, it appears that some participants who used all the concepts they could think of needed additional inspiration from the records in order to include more in their query or refine existing ones.

9.4.2 *Free text and thesaurus terms*

Task	Free text (average)	Range	Thesaurus terms (average)	Range
Clock	3.31	1 - 6	4.69	1 - 9
Upholstered chair	3.29	2 – 9*	5.86	1 - 12

Table 9.4 No. of free text and thesaurus terms used for "clock" and "upholstered chair" task
* One person searched for 9 terms, the others for 2 or 3.

In section 8.5.4, the analysis of free text (used to identify candidate terms from the thesaurus) and thesaurus terms was described. The only pattern between variables such as free text terms, thesaurus terms added to the query, thesaurus terms browsed and query terms deleted was that participants who searched for many free text terms tended to include more thesaurus terms in their query. A logical explanation for this is that participants tried to represent all concepts they thought of in their query. Looking at the sessions overall, participants either included many query terms when they were unsure of terms to use and reformulated repeatedly or when they were trying out the system. There does not seem to be a link between the participants' professions and terms used, although from the discussions, it appears that broad task approaches (i.e. avoiding more specific queries) could be linked to the participants' habitual tasks (8.7.4). However, when looking at the number of terms used, only few participants used the small number of query terms, which would have been the logical consequence of their habitual tasks. When looking at average terms submitted per query-run, it becomes clear that those participants who used a large number of distinct terms generally did not include more terms in a single query than other participants. Their larger amount of queries submitted accommodates a larger number of distinct terms.

Assuming that a Term Finder search such as "wall hanging clock" is one phrase, 21 different phrases were used in thesaurus searches for task 1 (clock). The phrases were grouped by concepts used (for 13 participants, considering e.g. "clock" and "clocks" as 2 distinct phrases):

- a type of clock (13 participants, 6 different phrases)
- wood (8 participants, 2 different phrases)
- pendulum (6 participants, 3 different phrases)
- Roman font (5 participants, 3 different phrases)

- Case (3 participants, 2 different phrases)
- 5 other concepts (carving, front, glass, rectangular, wall) by 4 different participants.

Concept	Employed by number of users	Expressed in number of phrases	Expressed in number of thesaurus terms
Type of clock	13	6	6
Wood	8	2	2
Pendulum	6	3	1
Roman font	5	3	2
Case	3	2	1
Other concepts	4	5 (all separate concepts)	9 (not necessarily the same concepts)

Table 9.5 Concepts, free text phrases and terms used in the "clock" task

The concepts were represented using 11 different thesaurus terms. The four main concepts are types of clocks (6 different thesaurus terms), wood (2), typefaces (2) and pendulum (1). Another nine terms were used but could not be grouped further. These include the last category mentioned above and other terms from records. Seven of the 13 participants used the first 3 concepts and another 4 participants "pendulums" and a type of clock. The remaining two participants searched either for "wood" and "clocks" or "clocks" on its own.

Concept	Employed by number of users	Expressed in number of phrases	Expressed in number of thesaurus terms
Type of chair	4	3	3
Material	5	6	9
Upholstery/related technique	7	6	9
Time periods	1	2	2

Table 9.6 Concepts, free text phrases and terms used in the "upholstered chair" task

For Task 2 (materials used for upholstering), participants used slightly fewer free text terms, but 12 out of 17 were used by only one person each, so that overall, the phrases were more diverse. They can be grouped into three main concepts (out of 7 participants):

- a type of chair (4 participants, 3 different terms)
- a material (5 participants, 6 different terms)
- upholstery or a related technique (7 participants, 6 different terms)
- 2 terms describing time periods used by one person who experimented with the interface

The same concepts are also reflected in the selection of thesaurus terms, but each concept is represented by more terms than in the "clock" task. There are 9 different material terms, 3 for chairs and 9 to describe upholstering. 4 participants combined all three concepts, 2 used a type of material and upholstering or chair respectively, and one person only looked for chairs.

In summary, participants used more concepts for the "clock" task, but the agreement on thesaurus terms is higher. For the "upholstered chair" task, they agreed to a higher extent on the concepts, but used more different terms within them. There seem to be two different effects of the thesaurus: for the "clock" task, people converge on terms, while for the "upholstered chair" task, they found more terms in the thesaurus than they originally started off with. The tasks were investigated for reasons which caused these two different effects. One of the most obvious differences is that one task is in the form of an image, while the other one is text-based. Although it can be said that from the image, participants had to come up with their own terms while they could use some words that appeared in the text description, this does not seem to be the primary reason. It might be more important that the clock has several features which could be described relatively easily even by somebody who is not a subject expert while the description in the "upholstered chair" task only provided limited information. The second difference is in the design of the tasks – the "upholstered chair" was primarily aiming at use of the term expansion and indexing terms in records. For the "clock" task, selecting a specific record can provide the answer to the task, but the "upholstered chair" task requires some additional processing. This task is thus more open and participants had more options, but relied on their own understanding to identify these.

9.5 Training and additional support

9.5.1 Formal training

During this study, the training was more structured than in the FACET 1 study through use of PowerPoint presentations (8.3.3). The training aimed specifically at re-enforcing the idea of search stages as the lack of knowledge of the search process had been identified as a problem in the previous study. Participants were also given a sheet with instructions to work through a hands-on exercise. Some of them referred to it later in the session and commented on its usefulness. In the library environment, this type of support material is common. Permanent accessibility of the information is probably an important issue. It would be possible to integrate the same kind of information into the interface, at least as a standard help file. However, issues of space and access arise from this.

In one session, two participants worked together and this appeared quite effective. Research on collaborative searching exists, but will not be considered here in any detail. Their different reactions to the same information however allow a glimpse of how different users are and reminds us of the fact that this needs to be taken into account when preparing training and support materials.

It appears that overall, the introduction and training prepared participants sufficiently to complete the tasks. Not everyone used the term expansion for every task, which might have been because they forgot about it and depending on the thesaurus terms used for the query, even a query with 0% expansion can retrieve sufficiently good results. This can also be considered to be a specific singular feature, and the overall search process seemed to have been remembered well.

9.5.2 Support through the evaluator during the search session

As the study focused in the observation of search behaviour and the resolution of problems rather than the evaluation of the interface, it was deemed appropriate for the author to provide assistance during the search sessions, as discussed in connection with the think aloud method in section 4.4.1. The interventions point to situations where potential support is required. Based on their analysis, suggestions are made for system interventions in the model section on Risk Points discussed in chapter 11.2.2. The audio transcripts were examined for every instance where the author gave participants pointers or help in any form. Each situation was assessed with regards to the content and type of intervention. In some cases, the author asked a question to make the participant think about an issue, sometimes, she provided hints as discussed in connection with the evolving scenarios (section 6.8.3, e.g. pointing towards an indexing term in a record which could be used in the query), and in other situations, she gave an explanation of system features. Through these categorisations, the interactions were generalised and resulted in the following main categories, which reflect the problems raised in the rest of this chapter:

- Use of the interface or specific features
- Expansion and related issues
- Identification of query concepts and query terms
- Assessment of query terms

One of the most common interventions was stopping people from browsing the thesaurus from the highest level, after associated frustration in locating terms had been observed in the FACET 1 study. Furthermore, the author pointed participants who had difficulties towards records which contained terms they could either use for their query or which should give them ideas regarding their reformulation, for example replacing one term with another one. The author also had to explain the expansion and related issues, for example the concern with match values (see section 9.8.1). Other aspects with which participants needed support were the use of categories when searching Term Finder and remembering to set focus term and theme. At times, the author had to repeat information on features of the interface, for example highlighting the difference between the Term Finder and Query panes in the Query Builder.

9.6 Searching the thesaurus and identifying query terms

9.6.1 Thesaurus access

Particularly when participants did not know what to search the collection for, it was difficult for them to find query terms. One curator, when searching his own collection, would have liked to see records in order to choose indexing terms. Exploratory search techniques might have been beneficial in this case, and could also help users familiarise themselves with a collection.

The present version of the interface does not provide support for use of Term Finder. In some way, participants probably approach it as they would a web search engine. Some participants enquired though about separators to be used between phrases. However, some syntax information might be beneficial in cases where retrieval causes problems, for example when searchers do not enter two terms of a phrase one after the other. The likelihood of this problem occurring is probably small but problems can still arise because the system will then not consider combinations of these two words when searching the thesaurus for matching terms and potentially miss out relevant terms. The same applies for the use of wild cards. If one participant had searched for "upholster*" instead of "upholstery", he would have retrieved "upholstering". Whether he would have used this term is another question, but at least the option would have been there. Use of wild cards might make more complex algorithms mentioned in section 9.8.4 – "Term expansion" to some extent unnecessary. However, the issue here is when users are made aware of these options. During initial sessions, when the training included information on selecting a category to search for thesaurus terms, participants tried to use categories and this caused more confusion than it benefited them (see also section 9.6.3).

Identifying terms to search for items from a specific time period is a particular issue. The problem here is that these terms exist in the AAT, but that it is not easy to access them. Drilling down, the categories are not very clear, and when searching in Term Finder, the results are not necessarily relevant. As one participant stated:

> I'm a bit confused at this point because there seem to be date ranges and time periods related to styles or regions of the world. I'm not sure where to go now to find my date period.

The problem is that the time periods are not in any way connected to date ranges. Some contain information on dates in the scope note, but this information is not standardised.

During the discussion, a participant pointed out that when talking about furniture, terms such as "Victorian" are familiar, but it requires an increased awareness to think of them. If a user cannot think of the name of a period, it is difficult to find a good starting point even for browsing the thesaurus in its local context. The thesaurus structure as displayed in FACET does not give any clues with regards to the chronology of eras. Within the AAT, the order of terms is chronological though and this could potentially be used in the FACET interface.

It should be noted that although the task on the 19th century chair was dropped, the issue of times still occurred in one of the curator's session. This might be an indication that use of time is indeed important and that this issue is potentially more important than it might appear from the observations in this study (see also Bates et al. 1993).

9.6.2 Breaking the task down into components

One of the issues during the previous study was that some participants struggled to break tasks down into their components. In this study, this seemed to be less of an issue. One of the contributing factors is probably the re-design of Term Finder. The system now searches for combinations of the entered free-text terms, rather than for a single phrase only. This is a better way of handling users' approaches which resemble the way in which somebody might search using an Internet search engine. Effectively, if a user does not realise that their phrase has to be represented through two separate concepts, the interface can still return the appropriate thesaurus terms even if they stem from different facets.

9.6.3 Searching the thesaurus by facet

Participants had some practical questions on searching the thesaurus in order to locate controlled terms to use in the query, for example regarding the formatting of search phrases, whether to enter terms individually or at the same time and the word order. Overall, information on this process seemed to be lacking. In addition to that, a re-occurring issue was restricting the thesaurus search to specific categories. The initial training presentation still included an introduction to selecting a category, but this was then removed as participants seemed to think they **had** to make a selection. However, during some later search sessions, some participants still tried to use the categories.

Selection of categories was the main situation in which the faceted structure of the thesaurus might have caused problems. At the same time, these could also have been a result of the participants' misunderstanding of the categories as database fields. Indications for this are comments along the lines of it being a more advanced search facility or filter. Another participant thought that maybe his results would be better if he searched for "carving" in objects rather than in processes[2].

One participant in particular seemed to struggle with the categories. First he searched Term Finder for "clocks". After running his first query containing this term, he decided: "So I wanted another term, what I was going to have was actually, a category, wasn't it? I was going to go for materials." He searched for "clocks" in materials, but no candidate terms were returned. The author pointed out that he had mentioned "wood" and should search for this phrase. Later on, he made a very similar attempt, searching for an object in the "Times" category. It could be suggested that these problems stem from being used to fielded searches, but in that case, the participant would still have searched for "wood" in materials. It seems more likely that the participant saw the categories as a way of filtering, as mentioned above.

Other participants seemed to understand better how the categories worked. One searched the thesaurus for "clock roman face drop case" and pointed out to the author that "it would be 'objects', surely." when she told him that there is no need to select a category. The participant was right, but such assumptions can potentially lead to problems. As another participant experienced, "upholstering" is a process, not an object, as some participants expected. This risk exists for anybody who is not really familiar with the AAT and the way terms are classified and used in the indexing. One participant for example assumed that "moquettes" was an upholstering material, but it is basically a type of carpet. The participant was going to locate the term by browsing, but would have struggled to find the term, as it is in the "objects" facet, not in "materials".

This takes us back to issues of providing thesaurus access points. The main motivation behind implementing searches by category was that it can help shorten long lists of candidate terms which are awkward to look through. However, these incidents demonstrate how this feature can potentially hinder users from finding any terms at all, and appropriate precautions need to be taken, for example expanding the search to all categories if no terms are retrieved.

The faceted structure of the thesaurus affected the participants' searches also in another way. They underestimated at times the distance between two thesaurus terms, for example "leather" and "imitation leather" or "clocks" and "pendulum" (see section 9.8.4

---

[2] He used the term "carving", which is understood as an activity. "Carvings", a different thesaurus term, refers to objects, and would thus appear under the "Objects" heading in the query.

– Detail of the thesaurus structure). Sometimes these terms even stem from different facets (e.g. materials and objects) and they are so far apart that even 100% expansion is not enough to include the second term in the query.

9.6.4 *Difficulties in term selection*

Despite scope notes, participants were uncertain about some of the terms. Participants decided for and against the same terms because they interpreted the features of the object or requirements of the task and the use of the terms differently. One participant for example explained that he could not be sure that the clock was indeed a wall clock, but was certain that "<weight-driven clocks>" was applicable. This term did not retrieve any results though, so that he settled for "clocks". In the discussion, the participant said that in real-life, he would confirm task information with the client. Thus, additional information which he felt was not available in the study, would facilitate selection of terms. A second conclusion from these incidents is that some people feel more comfortable using only terms which they are sure of. In cases where it is a matter of choosing a Narrower Term, the expansion allows users to select the term they are sure of and include the uncertain Narrower Terms through adjusting the expansion. Some participants seemed to be unaware of this approach, others chose it so they did not have to make a decision between a number of narrower terms they were unsure of. However, they did not always remember to adjust the expansion accordingly. Overall, participants did not seem to be interested in taking an experimental approach to searching. Only in a few cases, they tried out a larger number of terms to see how the results would change, which is an option in order to resolve questions regarding the effects a term will have on the results of a query.

9.7 Query construction and reformulation

General information on query reformulation observed has been presented in section 9.2 and 9.3. This section now addresses some specific issues, such as use of indexing terms from records and the expansion in order to reformulate a query, the use of semantically-close terms in order to distinguish between records of an advanced query and problems associated with 0 hit queries. Participants did not always consider the selection of terms aloud so that information on possible search strategies is not always available. However, points that they mentioned include concerns about making their query too specific or broadening it too much by including further terms. The later would not actually apply in this system where only records containing the focus term (or a Narrower Term of it) are retrieved. Some participants were wary of using more terms: "If I put another term in there, then chances are that I'll get 0. I'm probably not going to get any matches that have 100%.". If a query retrieved records with a reasonable match value, it is very unlikely that adding terms will result in zero hits though. The match value would indeed decrease and many participants were concerned about 100% match values (see section 9.8.1). As mentioned above, there was a certain reluctance of participants to test out modifications but no data on why this might have been the case is available. One participant added more and more materials for the "upholstered chair" task, but realised that this approach would not actually retrieve objects with a range of materials. At this point, he felt that his query had become too precise. He considered increasing the expansion but decided against this as he was concerned about including records with terms he is not really looking for.

Some of these issues might already arise when searchers select terms for their initial query. However, no particular distinction was made in this study between issues during query construction and reformulation so these issues are discussed here together.

9.7.1 Use of indexing terms from records

Overall, examination of indexing terms took place in every session either to assess the relevance of records or to select more suitable terms for the query, although the conclusions participants drew might not always have been correct.

One participant seemed to understand indexing terms which had not been matched by query terms as suggestions, and if query terms were not matched, as an indication that the term was inappropriate. As such, the participant was not wrong – if query terms are not matched by (any) records, then the term might not be very suitable in the context of the query or collection. If a record closely corresponds to what the searcher is looking for, then using more indexing terms for the query would result in retrieval of more similar items. However, this is an implicit rather than an explicit suggestion based on evaluative processes performed by the system.

Most participants indeed understood and used records as an implicit source of information on terms to add, replace and delete. Some also identified terms they would have liked to exclude, which is not an option in the FACET system. Others identified suitable terms but did not add them to their query, which later caused them some confusion. One participant decided to locate a record that referred to an item from the 19^{th} century in the list of results and thus identify an indexing term that would be suitable to describe this concept. This is a very good idea, especially in the light of problems with expressing time periods (see section 9.6.1). However, the records were not indexed in sufficient detail (see section 9.8.4) so an appropriate term could not be found. Other participants did not use exact terms from records, but used them as an inspiration. One record for example described the shape of the clock face as circular, so the participants decided to include "rectangular" for the shape of the clock case.

9.7.2 Use of query expansion

Participant	Task	Before 1st query run (only for participants who later changed the expansion)	After 1st query run, position in the sequence of query modifications	Total query modifications
E	Clock	-	1	2
G	Clock	1	**4**	6
K	Clock		**7**	10
M	Clock	1	2, **3**	4
G	Upholstered chair		1	6
H	Upholstered chair		3, **8, 18, 19, 21**	21
I	Upholstered chair	2	**3,4**	5
J	Upholstered chair	-	1	3
K	Upholstered chair	1	1	3
A	Other	-	1, 3, **6, 9**	11
L	Other	-	**5**	6

Table 9.7 Query expansion changes made after the first query run. Table shows affected sessions only. "Late" uses of the expansion are shown in bold (see Footnote 3).

Some participants made little use of the query expansion. They might have forgotten or it did not suit their approach. For the first four tasks, the default expansion was 100%, i.e. including the maximum amount of terms, so that these participants might have felt little need to change it. The default was then set to 0%, i.e. including no terms other than those explicit in the query, as an incentive for participants to modify the expansion. One person stated they did not really know how the expansion mechanism worked. Others also seemed to feel similarly – not knowing exactly how it affected the query made it more difficult to use the expansion in the most efficient way. Most participants modified the query expansion late[3] in their query (12 out of 20 modifications, bold in table 9.7), almost as a last resort, although in 7 out of 22 sessions, participants also set the expansion rate before the first query. Changing the value from the default of 0% to 100% certainly helped to increase the number of records and due to the ranking, it was not vital to reduce the set afterwards, although participants did take this approach. One participant stated (mistakenly) that he wanted to increase the match values of retrieved records and thus increase the expansion rate. Another situation for use of query expansion was inclusion of NTs that participants were unsure of or felt that several were applicable. Sometimes, they made their decision on search terms according to this, but later did not change the expansion. One participant noticed this, but then decided not to increase the expansion because enough records had been retrieved.

[3] A late reformulation was one either after half of the successful query runs or after 5 successful query runs (which is the average number of query runs) in cases of long sessions. These might be arbitrary values, but comparison using query runs seems justified as this is a measure for overall query reformulation. After half of the query runs, it can be considered that the searcher tried a number of options they were to try out overall in their session.

Some participants identified suitable terms for their queries from the list of expanded terms rather than the hierarchical thesaurus display. However, this seems to be more suitable for serendipitous discoveries, as in cases where participants wanted to include specific terms, they had trouble locating them in the list. Somebody then just set the expansion value to 100%, assuming that this would include the required term. This might not have been the case though - as is discussed in the example of "leather" and "imitation leather" in section 9.8.4 – "Detail of the thesaurus structure", participants sometimes did not judge the distance between terms correctly.

Participants also noticed terms they did not want to be included in their query amongst the expanded terms. Setting the expansion for each query term individually so that these terms could be left out, might be an option or even necessary in certain systems (see 9.8.4 – Term Expansion).

Another participant commented that using the expansion was "a good way of narrowing your search without having to find more and more specific descriptive terms." In fact, starting with a high expansion rate would result in more records being retrieved so that users have the possibility of identifying those terms which retrieve the most suitable records, put them into their query and then reduce the expansion so that records are limited to those where the terms occur.

9.7.3 Using semantically close terms to distinguish between records

One participant looked at some records of pocket watches and decided to include the term "Roman numerals" in the query. When looking at results for this new query, he noticed that some records contained the indexing term "Arabic numerals" instead of "Roman numerals". They were both considered a match due to the term expansion. The issue that becomes apparent here is that the participant selected "Roman numerals" as a term to distinguish the clocks that have Roman numerals from those which are not indexed with this term. The logic behind expanding this term is that users might also be interested in records containing "Arabic numerals", whereas this participant had reached a point of specifically fine-tuning the query. One of the current limitations in the implementation of FACET 2 is that the expansion value can only be changed globally for all terms. In an interface where terms can be expanded individually, terms used to distinguish between records can be left unexpanded, although this might not be the best solution. In this example, it would have been sufficient, as "Roman numerals" does not have any NTs. In certain situations, it might be beneficial to include NTs, but not siblings in order to make this distinction. The focus term has already been identified as distinct from other terms, and despite the fact that the current implementation in FACET 2 might not be ideal, this term has a role different from others in the query. Terms that users might employ later in their search to fine-tune their query might also need to be treated differently. Considering the level of complexity, this might only be feasible in an advanced interface and system.

9.7.4 Dealing with 0 hit queries

One of the most difficult problems to resolve for users is why a query does not retrieve any results. Some of the queries submitted (18 out of 147) did not retrieve any results. Some participants relied on assistance from the author in order to continue their search, while others were more proactive and attempted to modify the query themselves. Common modifications were changes to the terms, for example using a semantically close term or a completely different concept. Participants also increased the expansion in order to include more terms in their queries. Other modifications included changing the theme or the collection. Overall, participants managed to retrieve records by making one modification. However, in one session, the participant struggled to retrieve records and attempted a number of query modifications. In the FACET 1 study, a similar incident was observed where a participant could not retrieve any records other than the three initially found (section 6.7.9).

In the case encountered here, the author knew though that a query for the particular term should have retrieved results, and initially could not understand herself what the problem was. This very clearly demonstrates how difficult it can be to overcome problems with inefficient queries. The system provides very little information on options users have to reformulate in general, but also does not provide any information specific to the query. Participants thus attempted to rectify the problem by trial and error. Most of them were lucky that their modifications were successful. In a few cases, they were able to fall back on different terms and could sometimes even return to the initial concept when they saw appropriate terms in their result records. In real life, searchers cannot simply walk away from their tasks and they might (initially) only have one concept to work with. Support for reformulation is thus a major issue, and is discussed in detail in connection with the model, part of which specifically deals with avoiding and resolving problems with query construction and reformulation (see Risk Points in chapter 11.2).

9.8 Other observations

9.8.1 Knowledge transfer from other interfaces

One striking observation in this study was that participants frequently referred to Boolean logic. One participant even assumed that the system used Boolean logic without this being visible at interface level. Maybe it is not problematic that users do not understand exactly how a system searches for information, but it is important to assess the implications of these false assumptions, some of which have already been discussed. There is for example the reluctance to add further terms and the wish to exclude terms either explicitly in the query or from the list of expanded terms, which corresponds to using Boolean NOT.

One of the ways in which the Boolean mind-set finds expression is the assessment of the relevance of results. The interface provided participants with a match value that described the overall closeness of indexing to query terms and in the record display with a list of indexing terms as well as an indication of the distance between indexing and query terms. Participants needed some understanding of how the match values were calculated. They did realise that a 100% match meant that all query terms were matched, and were aware that records further down in the list were less close matches to the query. Some participants might not have had a very good understanding why that might be though. In itself, it is not essential that people know exactly how the match value is calculated, but in the study sessions, people applied a Boolean perspective. One participant for example did not want to add a third term as he feared he would not to retrieve any records. During the post-search discussion, he then commented:

> The thing that seems to be missing - and I'm sure it's running in the background - is the Boolean operators. And I think that's what's happening here, you're AND-ing things together. There is no way, there does not seem to be a way of OR-ing or NOT - using the "or" and "not" operators.

Thus, quite obviously, throughout about 50 minutes of searching with the interface, he had maintained the idea that the system was using Boolean operators. This did not influence his search session negatively in a noticeable way, although one of the effects of a "Boolean mind set" might be the concern with 100% match values and the amount of records retrieved, which was one of the first things participants checked when inspecting results. This might be the most tangible way in measuring success, although it seems that if the value did not change, they thought that their modification had not had much of an effect and some participants did not want to consider records that were not a 100% match. However, due to the semantic closeness expansion, records can have a match value of less than 100% but still contain all concepts. Indexing terms are then close, but not exactly the same as query terms, while in a Boolean system, a match value of less then 100% will indicate a missing concept.

Reformulation does not normally affect the number of records retrieved, as this is primarily determined by the number of records matching the focus term. Some participants were not aware that after reformulating their query, records would effectively be re-sorted, but that the expansion value might not necessarily affect match values, especially of higher ranked records. Some were unaware how adding more terms would affect their query, although others did take notice of this. Users need to learn to interpret the match values better to make the best use of the information, rather than concentrating on information which does not allow them to draw conclusions for their reformulation.

A second issue, although not as prevalent, were references to fielded searches, which might partly be related to the faceted structure of the thesaurus. The museums' management software has fields that users search although they mostly seem to use the descriptor field. Comments like

> Evaluator: Do you see why it's 75%? Or why it's not 100%?
> Participant: I wonder whether it is because I have "carving" in processes and not in object.
> - (See Footnote 2 on p. 151 for an explanation.)

seem to reveal this perspective.

9.8.2 Support of the search process

Most incidents seemed to occur at the reformulation stage. One reason for this is probably that this is the stage of the search at which people spent most time as most reformulated several times. Creating a first query takes comparatively little time. The interface also provides some guidance for creating this first query: Users click the "New" button which is similar in most Microsoft products, then they are instructed to enter terms and nearby is a "Find" button. After the training, participants were familiar with the idea that terms were then dragged into the query-half of the Query Builder window. Setting up the query is still not obvious enough however. As one participant commented:

> Intuitively you would probably think "Oh, right, I'll go to the 'Terms to find' ". Perhaps if you are used to searching and you have the "Find" button, so you might click there. And then after that, the dragging across and making the query is probably ….That's the bit that's not so obvious to start with. Having to move from there to there …
> **The participant moves the mouse from the "Found terms" pane to the query side of the Query Builder.**
> You have done something here…
> **The participant moves the mouse over the query terms.**
> then you come here …
> **The participant moves the mouse over the Theme menu and to the Query Expansion tab.**
> and then you have to say "okay" and then you come to that problem there,
> **The participant moves the mouse over the list of result records.**

```
where it looks [...]it's actually run it, particularly because
you can view these last searches it's saved..
```
The participant moves the mouse over the list of queries.

However, when reformulating, the guidance available does not apply to the same extent and is only really helpful if users want to include additional terms in their query. In several cases, participants started browsing the thesaurus from the top level in order to locate terms in the thesaurus. In other cases, they attempted to modify the query by deleting indexing or query terms in records rather than in the Query Builder. One participant said he wanted to refine his query, but did not know how. He almost seemed to expect to see the words "Refine" or "Reformulate" on the screen, which is a feature of many web-based interfaces.

Thus, although participants required more support when reformulating, comments such as:

```
This is what I've got. Now... basically, I'm at the refinement
stage, am I? That looks like something that - do I double-
click or right-click?
```

and

```
Maybe I will add 'wood' in because it has a wooden frame.
So, refine query. ... Let's go for a wooden one (clock). Hang
on... That(record)'s got no wood in.
```

show an increased awareness of the search stages compared to the FACET 1 study.

9.8.3 Low level issues in the interface

After an extensive discussion of conceptual issues, this section now briefly outlines lower level issues. Most participants found the interface complex and felt that despite the training, they might not be able to remember how to use it, but that this issue might be resolved with experience. A more long-term study to verify this was beyond the scope of this thesis. Participants who had taken part in the FACET 1 study gave an indication though that indeed, the FACET interfaces are either easy to remember or version 2 is intuitive.

A number of issues seem to be related to the fact that participants were unfamiliar with the interface. These include for example remembering what the angular brackets (guide terms) or the little arrow in the terms' icon (related terms exist) meant or how to display Narrower Terms. Two participants for example dragged thesaurus terms into the query before double-clicking them in order to see them in their local context in the thesaurus. Participants also had to be reminded to set the focus term and select a theme. The interface did this on closing the query, but participants felt that this was a bit late. The query did not run automatically so that users have the opportunity to include terms from several different records, but some participants waited without executing the query both after completing the first one and after reformulating it. Changing the background colour of the results might be an effective way of indicating that the current set is not up-to-date. Other minor design issues included the not-updating of the status bar which showed the number of records retrieved. This was information participants were particularly interested in and thus caused some problems (section 9.4.4). One participant for example thought no records had been retrieved even though they were displayed on the screen.

Participants commented that the thesauri they normally used could not be browsed and some thought the same type of relationships (e.g. associative relationships) did not exist (refer to section 8.7.2). Some participants understood the expanded terms as Narrower Terms of the query term. One participant commented on the fact that the expansion display obstructed the relationships between thesaurus terms, and other participant's comments also indicate problems with the concept ("91% of elbow chairs are made of leather?"). The list of expanded terms faded depending on the distance of the term from the query term and terms are displayed with percentage values, but this distinction from the hierarchical display appears to be insufficient. The motivation for showing the expansion in this form was that it was considered necessary to show users which terms would be included in their query, and participants of the study used this display to refine their queries, for example one selected here the term "wall clocks". More experienced users in particular might benefit from having information on the match value of a term included in the query through expansion.

Setting the expansion was in itself difficult. Using word processing programs and the Internet, people had probably come across sliders, but these might have been easier to use. One particular issue was that, unapparent to the users, the slider had to be moved further for some terms than for others due to the thesaurus structure. When reaching certain levels, many terms are included in the list and as the list is sorted by distance rather than alphabetically, it was difficult to judge whether a particular term had actually been included.

9.8.4 Potential system issues

Issues potentially related to database size and indexing "density"

Potentially, the expansion can result in many more records than is desirable, depending on the collection, the indexing, and the thesaurus, for example in collections where a subject area has many associated records but only few thesaurus terms to distinguish between them. When expanding a query term, more relevant terms do not necessarily have higher match values than irrelevant ones, so that less relevant records could be ranked equally high as relevant ones, thus using the expansion does not necessarily improve a query and users would require more detailed support on how to distinguish and separate more and less relevant records. The characteristics of thesaurus, indexing and collections thus need to be taken into account for each system when assessing the necessary support mechanisms.

Other functionalities, such as the "More like this" option to create a new query based on a specific record, might only have their full effect in these large collections. After using this option, one participant pointed out that the results did not seem to be much different from the original ones. Indeed, a query for "clocks" retrieves at any expansion rate a total of only 134 items, and any query containing the term "clocks" as a focus term and any combination of other terms will only re-sort the list of records, but not retrieve any additional items.

Indexing issues

Different institutions apply different indexing standards. Being familiar with one way of indexing, participants built their query accordingly as they probably expected that this is the most efficient way of searching. In a collection where a different standard has been applied, the same rules for searching might not work as efficiently or could even create problems. One participant specifically questioned whether he would be able to retrieve certain records which were only indexed with the Narrower Term, when using the Broader Term in the search. As discussed in section 9.8.1, participants need to work with the specific indexing of the collection, but are not always aware of this.

Indexing is also domain-knowledge dependant. Certain aspects are understood implicitly, for example that a clock has a pendulum. Only where exceptional circumstances regarding the pendulum apply (e.g. it might be missing), does the description or indexing explicitly mention this part. When searching, some participants made comments on allowing for limits in the indexers' knowledge as well, one said for example they might not have known what the exact description was and used a more general one. Participants also commented on having to make allowances for missing indexing terms, for example one of them had been searching for items from a specific time period but had to realise that this was not possible as none of the records had been indexed accordingly although the relevant concepts occurred in the free-text description. Another person had a similar problem in that he wanted to locate suitable query terms by looking in the results list for a matching description. He then found that not all elements of the description were translated into indexing terms and thus abandoned his otherwise good idea. Yet another participant had the opposite problem: The fact that less important features were reflected in the indexing made it more difficult to distinguish the relevant items.

These comments and incidents highlight the importance of high quality indexing not only for retrieval but also from the perspective of user searching behaviour. In all of the above cases, searchers had to adapt their approaches to particulars in the indexing.

Detail of the thesaurus structure

Especially subject specialists commented on areas of the AAT that are not very detailed. On the other hand, some areas appeared to be very detailed to somebody not familiar with the domain, for example "trimming (materials)" with 14 Narrower Terms. These terms are probably relevant to the kind of items the thesaurus was designed to describe. Those participants with more domain knowledge seemed to find it more difficult to select a term if the one they had been searching for was not in the thesaurus. Users with less domain knowledge might be less aware of the distinctions between terms and also of the gaps and inconsistencies that existed in the less defined parts of the thesaurus. These incidents indicate the usefulness of the thesaurus as a source of domain information, but also raise problems associated with it – if the area is not very detailed, the knowledge acquired is probably superficial.

Some participants seemed to expect terms to be closer together in the thesaurus than they actually were. For example, in the case of "leather" and "imitation leather", the terms appear to be very closely related. However, because "leather" is classed under "<processed animal materials>" and "imitation leather" under "<imitative materials>", there are actually six traversals (BT-BT-BT-BT-NT-NT) between the two. To take another example, although there is a hierarchy called "<clock and watch components>", which is already twelve traversals away from "clocks", "pendulums" is not actually here but under "<tool and equipment components>" (eleven traversals from "clocks"). It is understandable why somebody who is not familiar with the thesaurus underestimates these distances.

One participant would have liked to locate a term that signified the opposite of an indexing term. However, the thesaurus does not cater for this. He might have had the idea from English language thesauri which sometimes include antonyms as well as synonyms. He had a vague idea what he might have wanted to use, and could thus have searched the thesaurus for this term.

Term expansion

"Missing" related term relationships caused some problems. One participant expected the expansion of "upholstering" to allow him to browse to upholstering materials. Another participant ran into trouble when searching the thesaurus for "upholstery". This term was retrieved, but he remained unaware of "upholstering", the term actually used for indexing in the database. More complex Term Finder algorithms could overcome some of these missing relationships, but it can be considered inefficient to fall back on these techniques when using a controlled vocabulary in a system. Having the links built into the thesaurus might have effects on the expansion, and expansion along these relationships might actually have resolved some of the problems encountered.

9.9 Conclusions

9.9.1 Reaction to the interface

Participants agreed that the interface was complex but that they would probably be able to learn to use it well within a relatively short period of time. Participants also had some problems understanding the difference between the hierarchical and expanded thesaurus display, possibly because although several of them had previously used online thesauri, they were not really used to these types of display. However, interacting with them was not problematic. The mechanism for changing the expansion was more difficult to use though, for example, the slider was too small. Other than this, participants again were able to use the functionality for the interface without problems.

9.9.2 Search approaches

Support of the search progress was the single most important issue during the FACET 1 study. During this study, it is still an issue, although related issues are much less striking than before. It appears that the training and changes in the interface have successfully increased the support of users through the search process. As discussed in section 9.8.2, more support is still required at the reformulation stage.

Overall, differences between the participants' search approaches were much more subtle than in the FACET 1 study. This was potentially due to the use of evolving scenarios for the data collection. However, there are other influences which potentially had an effect on the search behaviour. These include issues related to the system, such as the detail of the thesaurus and indexing, as well as the task difficulty. The two different types of tasks also seemed to result in different patterns linked to the concepts which participants were able to identify. The participants' **perception** of the system also has an influence on their search behaviour as could be seen in connection with their "Boolean mind-set" (9.8.1). The level to which participants were sure of concepts and the difficulty they experienced in selecting terms seemed to have resulted in certain behaviour patterns. Even though tendencies for certain behaviour exist, at the decision points in a search, participants displayed a number of different reactions, and this could be a subject for future research.

9.9.3 Concepts

The participants themselves, the task description and the retrieved records were the three main sources for concepts (as opposed to free text or thesaurus terms) used and considered for queries (9.4). To help users identify additional concepts, the records seem to be the best source. Unless they browse extensively in the thesaurus, i.e. traversing many relationships and moving upwards and sideward (rather than down), the thesaurus is more useful in helping them refine query terms. However, used in combination with other aspects of the system, the thesaurus could be a very powerful tool for discovering new concepts.

9.9.4 Other influences

The indexing density and quality as well as the database size can limit the usefulness of the retrieval tools. Distinctions between records could potentially be less apparent and options like "More like this" might not bring up very additional results to those already presented (section 9.8.4).

Indexing standards users are familiar with might influence them in constructing their queries according to what they are used to rather than the way the collection at hand is indexed until they have adapted to it (9.8.4 – Indexing issues).

The thesaurus detail was also an issue which the subject experts brought up. Interestingly, when terms were missing, they found it more difficult to select alternative terms than those participants who knew less about the subject and can more easily accept alternatives as they are less aware of the distinctions (9.8.4 – Detail of the thesaurus structure).

9.9.5 Implementing search support

This research did not specifically look at implementing user support, however, the problems encountered pointed into this direction as a means of resolving some of the issues. Some potential solutions will be discussed in connection with the Risk Points in chapter 11. One related issue is the lack of space which makes it difficult to display larger amounts of information, either for feedback or to re-enforce learning, and the identification of the exact support required. In order to be useful to the searchers, support needs to apply to the specific situation, and possibly needs to be practically applicable. One observation during the sessions was that participants knew for example that they wanted to reformulate the queries, but were unsure about the exact actions to take. Making them aware of the options can thus already be a first step. Identifying the users' particular requirements and the best solution can still be difficult though, which is also related to the cyclic nature of the search process which lets users return repeatedly to the point of reformulating a query and potentially try out a number of different modifications one after the other. Subsequently, it appears to be important that the searchers are given the possibility to communicate their requirements in a form other than the query to the system. A wizard as discussed in 11.2.2 would be an option to guide searchers to the required information. This could possibly be presented in a stand-alone system as more advanced training options such as optional tutorials. These can provide instructions necessary for users to deal with a system without extensive training or experience.

9.9.6 Benefits of the thesaurus

The Risk Points discussed in chapter 11 detail more options for support both based on the thesaurus and other retrieval data. During the FACET 2 study, participants made use of the thesaurus in different ways. In a few instances, the thesaurus structure provided within a local context domain information to participants which helped them in selecting thesaurus terms. The scope note also provided information on suitability of terms. Some participants commented on the fact that using the thesaurus and the expansion allowed them to retrieve items even though they did not have detailed domain knowledge. Others said it had given them a different perspective on the tasks from the one they started off with. It also made it easier for them in that they did not have to think of terms without any support.

In section 9.4.2, a comparison of initial free text and thesaurus terms actually used showed that the thesaurus can have two different effects on the searchers' concepts and query terms. For the group of participants as a whole, it helped them for one task where they started with a large number of free text terms to converge on query terms. For the other task, where they only had few free text terms, it allowed them to identify more thesaurus terms.

However, the thesaurus did not seem to be very useful in helping people identify additional concepts. In many sessions, new concepts used in the search actually came out of the records, but not all participants took advantage of this source of information. As discussed in chapter 11, thesaurus-based tools could be developed in order to aid concept-discovery.

Concepts which cannot be expressed using thesaurus terms indicate problems, **but** it is unclear at the moment whether this could be due to issues other than the fact that the concepts cannot be expressed in the query. It appears that although general difficulties in selecting terms were encountered, participants who had better knowledge of the subject encounter more problems in selecting thesaurus term when they could not find the ones they expected to find. Alternative access to potential query terms is required, for example being able to see indexing terms from the collections. In special cases, as here with the time periods, alternative methods for locating terms in the thesaurus can be useful. These might require the implementation of additional tools, but this would probably pay off considering the general importance of this when searching object databases.

As discussed in section 9.7.2, one efficient way of using the system for a standard search seems to be starting with some broader terms and a wide expansion at first. Based on the records, searchers can then narrow these terms down and potentially decrease the query expansion. In future interfaces, this might be the approach to support in the interface and to teach users. Any problems related to this strategy, which was here not used by many participants, need to be uncovered in order to develop further mechanisms to fine-tune it and potentially to design the support within the interface and through other materials accordingly. Giving users instructions with this in mind might help them make the better use of the system than trying to explain how the expansion works, which is complex and depends to some extent on the individual situation, so that participants had difficulties in assessing its effect.

Overall, the system seems to have been improved from version 1 to version 2. The search stages were better supported through the interface and specific training, but certain unfamiliar features such as the adjustable semantic closeness expansion require either better instruction or more experience. Although search behaviour of individual participants appears to have been influenced by use of evolving scenarios, so that overall, differences between search approaches are more subtle, a number of factors that influence decision making during the search process were identified and considered in the model in chapters 10 and 11.

Chapter 10 – Model of the information searching process in controlled vocabulary enhanced systems

10.1 Introduction

This chapter proposes a model of information searching in controlled vocabulary-enhanced systems which was developed from the data collected primarily during the FACET 1 and 2 studies described in chapters 6, 8 and 9. This model does not aim at predicting searching behaviour, but at providing some explanation of how certain problems in information searching can occur, so that this can be taken into account in the design of systems and user support. At the same time, the model brings together and abstracts detailed findings from the studies, which could otherwise only be summarised with difficulty.

While existing models of information seeking and searching help our understanding of the information searching process, they do not provide practically-applicable guidance for systems design. This model attempts to provide this guidance by breaking the information searching process down into its stages and presenting influences from within the session, for example the number of records retrieved or data from individual records, on which searchers base decisions that motivate their search behaviour. For every stage, the model highlights information which searchers potentially require. One part of the model, the "Risk Points" (chapter 11), specifically deals with risks which can lead to breakdown at different points in the search – for example when no decision on how to proceed can be taken or where query modifications only have a detrimental effect on the retrieved results. These points form a basis which helps designers in making decisions to minimise problems during information searching and to implement support to resolve problems when these are encountered. The model also provides a framework for the evaluation of existing or prototype interfaces.

The documentation for the model consists of three parts. The first is a set of diagrams (figures 10.1, 10.4 and 10.5) which present entities and decisions together with the processes that connect them. These diagrams focus primarily on providing a readable presentation of the search process. The second part of the documentation is a tabular representation (figure 10.2) of the same entities, decision points and processes, which was developed while testing the model and which can serve as a basis for the evaluation of systems. The third part of the model documentation is a description of the Risk Points (chapter 11). These complement the description of the information search process with information on influences which potentially lead to breakdowns of different levels of severity, the most serious being abandonment of the search.

Section 10.2 describes the design of the model and section 10.3 provides a first overview over the model. Section 10.4 defines decision points and provides some explanation on the notation used. 10.5 deals with the testing. Sections 10.6 and 10.7 explain how the model can be used in design and evaluation of systems. Section 10.8 defines the scope and limitations of the model at present. In section 10.9, the information searching process described by the model is explained. Definitions of processes are given in Appendix 10.2 for reference. Some of the tools mentioned here are discussed in chapter 11 in connection with the Risk Points.

10.2 Design and development of the model

Section 3.2.1 in chapter 3 introduced Kuhlthau's and Marchionini's models of the information searching process. These models describe the information searching process in its basic form. The stages of problem definition, query formulation and execution and examination of results are relevant to the information searching behaviour investigated for this thesis. The incidents and comments collected during the FACET 1 study were first grouped into the proposed stages, and then ordered sequentially within each stage. For this purpose, the author identified the entities searchers interacted with and considered the mental and physical activities required to move between entities. This was for example done by comparing the participants' different search approaches as discussed in chapter 6. The individual phases were then fitted together which resulted in a basic structure for the model. Data collected during the FACET 2 study were used to develop the model further. Normally, several interactions can be performed on each entity. The author established these by inspecting the data for more evidence of how and why users moved between entities. It became apparent that certain conditions influencing the decisions also had to be represented, and this was done in developing the "Risk Points" (chapter 11). These outline under which primarily search- or system-related conditions specific problematic situations can occur and present possible solutions discussed during the studies. The model was generalised, for example by abstracting FACET-specific interactions, such as "dragging and dropping a term into the query" which in the model is represented as "including a term in the query" and through testing against existing systems. A tabular representation of the model was developed in order to facilitate this. This representation of the model is discussed in more detail in section 10.9. The model was tested against existing data and online interfaces which integrate a controlled vocabulary. The results lead to further corrections and raised some questions which require further investigations of search behaviour.

The model was thus developed under consideration of the literature using the data on information searching collected primarily during the in-depth studies reported in this thesis. Observations of the interactions with different interface components and the query enabled a description of the search process. The Risk Points draw together the analysis of difficulties encountered during the searches, their causes and potential means of avoiding these through the information searching interface or system. Both the description of interactions and the information contained in the Risk Points were generalised beyond FACET by testing the model against other controlled vocabulary enhanced systems.

10.3 Overview of the model

Figure 10.1 shows the main diagram of the model of information searching in controlled vocabulary-enhanced systems. Starting from the outside of the diagram, the outer white area represents the context of the search. The shaded oval represents information within this context that is more pertinent to the current search. These two layers place the current search session, i.e. white oval in the centre, in relation to models of information seeking behaviour which takes place outside of the search session and is for example discussed in models by Ingwersen (1996) and Wilson (1999) (see also section 3.2.1). The model presented here is only concerned with the search session itself, but needs to encompass the context of the search in order to take into account dynamic changes of the Information Need even during the search session. Block arrows (⇔) between the different areas indicate influences from one on the other. Entities are represented as rounded rectangles (▭), for example entity "(3) Selected Controlled Term". Decision points, at which searchers have several options to proceed, are represented as diamond shapes (◇) numbered with square brackets. Processes are represented by line arrows (→) and where the one arrow represents two processes, angular brackets (>) indicate the direction of flow. The numbering is explained in section 10.4. The black circles with white centre (◉) represent the end of the search session.

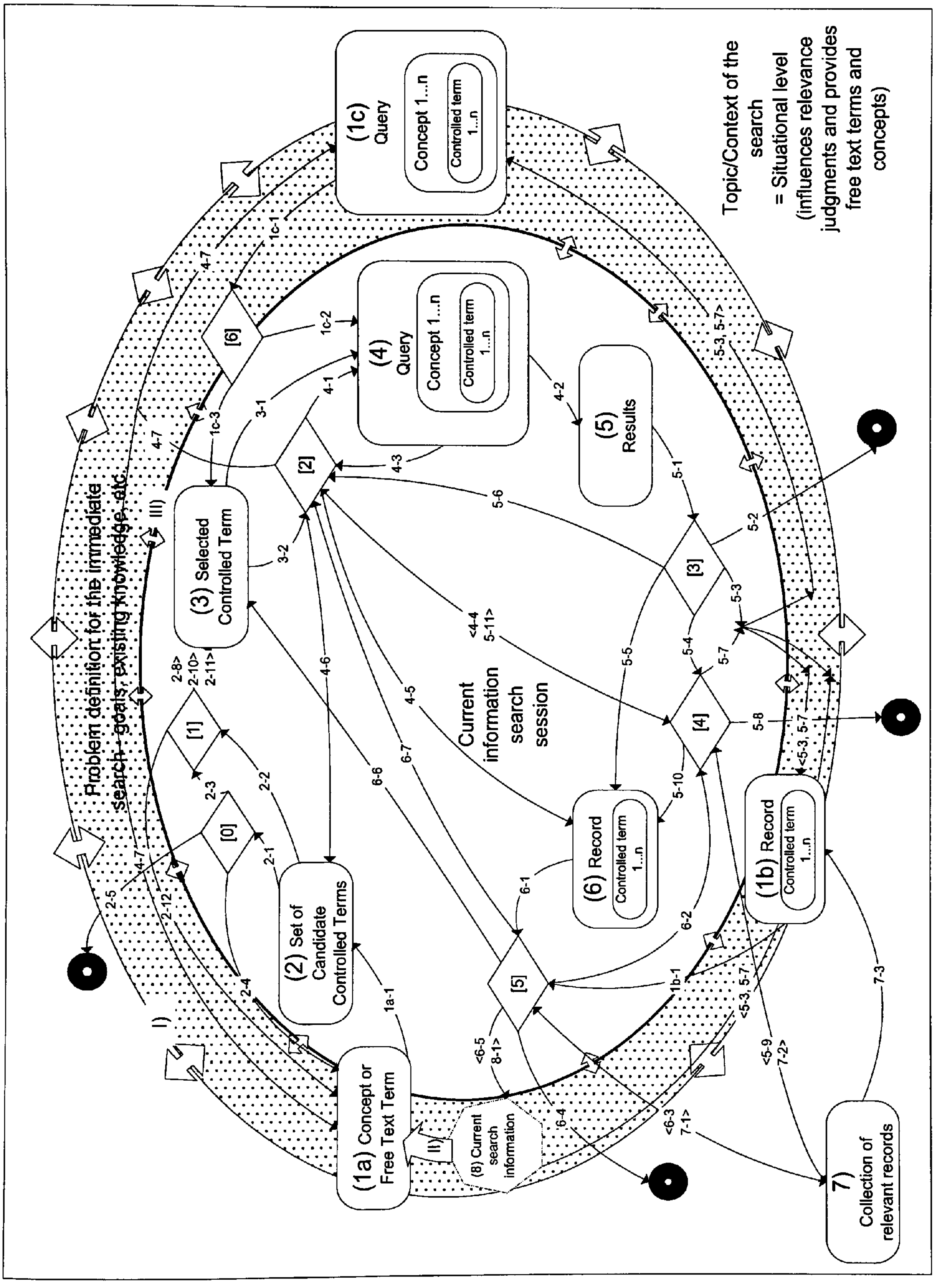


Figure 10.1 Model of the information searching process using a controlled vocabulary

10.3.1 Entities of the model

(1) Possible starting points

(1a) Concept or free text term

A task involves at least one concept or idea which can be expressed through free text terms or phrases. These are expressions which are not necessarily in the system's language/terminology and which thus need to be mapped to Controlled Terms.

(1b) Record

This entity represents a relevant record which is available at the beginning of the search. It could stem from an earlier search session, or be acquired through other means, for example from a colleague, or presented by the interface as an example which the searchers considers relevant to their topic. Records can be of any data type, as long as the query can be mapped to part of the record. They can thus also include non-textual data which have been indexed or can otherwise be translated into the format of the query.

This diagram implies that the record stems from the collection(s) accessed by the currently used system. Indexing terms are therefore in the right format and from the controlled vocabulary associated with the collection and system, so that they can be used as they are in the query. If the searchers have a reference to an article or a record which stems from a different collection, these are encompassed under entity "(8) Current search information", which serves as a source of free text terms or concepts, which need to be translated into Controlled Terms.

(1c) Query

Similarly to the record in 1b, searchers have access to this query at the beginning of the search session. It could have been saved during a previous session or be presented as an example. Again, it is assumed that the necessary steps to ensure relevance of the query have already been taken during the definition of the task and goals.

(2) Set of Candidate Terms

The Set of Candidate Terms is a stepping stone in the process of mapping concepts and free text terms to Controlled Terms. It is assumed that in many cases, the system finds several Controlled Terms that could be used to represent the concept. Some concepts might have to be expressed through several terms, for example "upholstered chair" (see section 6.7.5). In some implementations, the system selects the terms so that the Set of Candidate Terms is not accessible to the searchers.

(3) Selected Controlled Term

Controlled Terms are elements from the controlled vocabulary used to represent data in the collection. This model has been designed so that the controlled vocabulary can be of any type, for example a thesaurus, a semantic network or an alphabetical list.

Selected Controlled Terms differ from Candidate Terms in that the latter could potentially be used in the query while the suitability of Selected Controlled Terms has been assessed and the searchers have decided to use these terms in the query. The distinction is necessary as Selected Controlled Terms can stem from different sources, for example the records or other queries, and Candidate Terms can still be rejected as inappropriate.

(4) Query

This entity represents the query the searchers are currently working with. It consists of at least one concept, which is expressed through one or more terms from the controlled vocabulary.

Depending on the implementation, the searchers or the system can be required to set parameters, such as the collection to be searched, or options regarding query syntax or limitations.

(5) Results

The results are a set of records retrieved by the system according to the specified query. The model aims at being independent of the display and representation of results.

(6) Record

This entity represents a record from the set of results retrieved by the query. Records consist of a number of different aspects, indexing terms being the most important one in the context of this model. Others include a textual description, a photograph, information on the location of the item represented by the record, etc.

(7) Collection of relevant records

Searchers will want to use some of the records retrieved during the session, and thus need to for example print or save them. This entity represents a set of records selected from the collections accessed by the system. They can serve as a basis for a subsequent search at a later stage of the resolution of the searchers' Information Need.

(8) Current search information

Problem definition and intended use of the required information, different levels of goals, etc. form the context of the current search session whereas some information is more pertinent to the search session. Knowledge about the collection is for example acquired by (mentally) comparing two sets of results, or a record's indexing terms to the query terms, and feeds directly back into the reformulation. Although not strictly an entity in its own right, it was felt that this context should be made explicit in order to be able to represent some of the processes that take place, for example when generating free text terms, part of which need to be identified through future work.

10.3.2 The basic search process

The basic search process moves from entity (1) sequentially through to (7) and (8), i.e. identifying concepts and free text terms, mapping them to Controlled Terms, using these to construct a query, executing the query and evaluating the results. If more free text terms and concepts are identified during this sequence, the cycle restarts. Decision points also allow the return to earlier stages and thus iteration of processes, for example the assessment of several result records.

10.3.3 The contexts of the search

The contexts of the search are presented here to place the session into context with regards to information seeking models and to account for the possibility of evolution of search topic and Information Need. This larger context corresponds to the "situational level" in Saracevic's stratified model (Saracevic 1996b, 1997, section 3.2.1) or the long-term search goal in Xie's research (Xie 2002, section 3.3.2). Xie's leading and current search goals can be expressed through the shaded oval, i.e. the problem definition and the existing knowledge related to the search. This layer influences immediately the current search session in terms of relevance of terms and records, and the information that the users collect while searching is fed back into this existing knowledge about the topic. The block arrows between the white and the shaded oval represent this reciprocal influence between the search session and its immediate context. New relevant concepts and free text terms can be developed during the search based on the immediate search goals, and fed back into the current search. This process is here presented through the block arrow II) leading from entity "(8) Current search information" to entity "(1a) Concept or Free text term".

Here, the search context is presented in this manner, as it is difficult to include for example acquired knowledge otherwise in the search session. It will permanently be modified by the information discovered during the search – be it about the controlled vocabulary, the records or the concepts used to search – but it is also not an entity in itself like a "Controlled Term" or a "Query". The later represent logical objects that the user interacts with. Even "(8) Current search information", a manifestation of an otherwise invisible combination of information which influences the immediate interactions, and "(1a) Concept or free text term" cannot strictly speaking be put into the same category as the remaining entities. Entity (8) is thus represented using a different shape, and a broken line is used (not distinguishable in these figures).

10.3.4 Arrangement of entities

In the main diagram (figure 10.1), entities "(1b) Record" and "(1c) Query" are on the periphery of the current search session. During a first search on a topic, they will probably not even exist. Spink et al. (2002) have shown the commonness of successive searches, in which case a previous record or query potentially exists. The three options for entity (1) have been presented at the periphery of the current search session in order to express the fact that they form part of the knowledge of the topic although they also form part of the search session.

"(1b) Record" (an existing record) and "(1c) Query" (an existing query) have also been arranged so that they are positioned close to entities "(6) Record" (i.e. a record from the current search) and "(4) Query" (i.e. the current query). This has been done deliberately in order to counteract the perception that a search will always start from a concept or phrase. Searchers might just as well take up searching from other points in the process shown here. At present, systems limitations might prevent this in many systems.

10.3.5 Areas of application for the model

Although primarily based on usage data of a ranking search system integrating a controlled vocabulary for both searching and indexing, the model discussed here also informs and models to some extent systems which use Boolean matching functions and those which use a controlled vocabulary for searching only. The model applies to different types of controlled vocabularies. Tests were for example conducted with the OVID interface to Inspec which generates an index of author names based on the records in the collection. When users conduct an author search, part of this index, commencing with the closest match to the text entered, is presented as an alphabetical list from which they then select entries. This activity is also reflected in the model.

10.4 Definitions

In the model description, different fonts and numbering have been used to distinguish between decisions, processes and entities. Decisions are identified through numbers from 0 to 8 in square brackets. When decisions can be broken down into sub-decisions, these are distinguished using letters, for example [1a]. In the description, decisions are displayed centred, in a large bold font. Entities are also numbered, but have round brackets. When different types of objects exist for one entity, these are distinguished using letters, for example "(1a) Concept/free text term" and "(1b) Existing record". Process identifiers consist normally of two numbers. As a general rule, the first one identifies the entity where the process starts in order to facilitate the location of the process in the diagram. Process 1a-1 is thus in the proximity of entity "(1a) Concept/Free text term". After a decision, it is unclear from which entity the process initially originated. For example when looking at 4-1, the process of setting up the query, the searcher could in the previous step have selected a query term or inspected records. In these cases, mainly related to query reformulation, the first number is derived from the entity where the process ends. The second number of the process is used to distinguish several processes which start or end at the same entity. They have normally been numbered clockwise. Within the model description, the processes are largely self-explanatory, but their definitions can also be found in Appendix 10.2.

In the model description, processes are separated visually by displaying the different conditions under which they take place with an arrow:

→ NUMBER OF CANDIDATE TERMS IS ACCEPTABLE

Overall, it has been attempted to follow the information searching process in a logical way, but this was difficult due to the number of options which exist at each decision point and the linear nature of this document. When a process leads to a decision or entity other than the next one discussed, the page number follows in round brackets, so that the reader can more easily jump to that section.

10.4.1 Decision points

[0]

Number of Candidate terms retrieved by the matching mechanism

[1]

Relevance of Candidate Terms retrieved by the matching mechanism - is any of them suitable for the query?

[1a]

Is the expected term amongst the Candidate Terms? The "expected term" is here the term entered into the mapping mechanism or a slight variation.

[1b]

Does the expected term have the correct meaning?

[1c]

Are any of the other Candidate Terms suitable for the query?

[2]

Considerations regarding set-up and modification of the query: Are enough terms included in the query in order to proceed to deal with the set-up? Are the concepts sufficiently covered? Are the terms appropriate?

[2a]

Does the query need to be modified, and if so, which modifications need to be made?

[2b]

Have all steps necessary for the query construction and set-up been taken?

[3]

Number of results retrieved by the query

[4]

Relevance of results retrieved by the query – are any of them useful/relevant?

[5]

Relevance of aspects of records: Examination of the aspects of the individual record (indexing terms, description, etc.) leads to a decision on how to use the record, for example identifying indexing terms to include in the query, adding the record to the Collection of relevant records or reading it.

[6]

Assessment of the query: is the query to be modified, or individual terms to be selected?

[7]

Suitability of concepts (figure 10.5 only)

[8]

Suitability of individual terms (figure 10.5 only)

10.5 Testing the proposed model

The model was tested in order to clarify some of the processes, to identify errors within the model, particularly concerning generalisation, and to assess the extent to which the model applies to different types of systems. The tests were conducted using a number of existing interfaces and the data collected during the studies discussed earlier in this thesis (chapters 5, 6, 8 and 9). At this point, it has not been possible to resolve issues regarding the processes of query reformulation, but other processes have been confirmed or where applicable, modified.

10.5.1 Testing against existing interfaces

As mentioned in chapter 2, the author identified a number of interfaces on the Internet which are enhanced with controlled vocabularies. These were used in testing the coherence and completeness of this model. Ideally, stand-alone systems would have been included in this evaluation, but unfortunately, the author did not have access to any other than FACET. The main differences between stand-alone and web interfaces appear to be in the interaction mechanisms and certain functionalities which require high bandwidth and other resources which can be restricted on the Internet. The interfaces considered in this evaluation are mainly front-ends to complete systems which have been available online over a number of years, presumably designed to contain a maximum amount of feasible functionality. The OVID interface[1] (discussed in sections 2.4.2 and 5.3.1) is a commercial citations database. ERIC (section 5.4.1) and CHIN (section 2.4.3) are public interfaces which provide access to online resources, Flamenco (section 2.4.4) is an experimental system specifically developed to facilitate faceted data retrieval. The subject search facility of the OPAC system of the University of Glamorgan Learning Resource Centre (LRC) was also used in order to broaden the types of interfaces.

The tabular representation of the model

In order to assess the model in an objective way, the author developed a tabular presentation of the model in the form of a pivot table (figure 10.2). The entities and decisions of the model correspond to rows and columns. The processes that connect them are represented through the cells where the respective rows and columns intersect. Moving from a free text term to the set of Candidate Terms is for example presented in cell A2, where the row for entity "(1a) Concept/Free text term" intersects with column "(2) Set of Candidate Terms". The tabular layout thus allows the representation of connections between any pair of elements even if they do not exist in the model. References to identified processes are found in the respective cells, cell A2 for example contains the reference to process 1a-1, which corresponds to the process from entity (1a) to entity (2). Some cells contain several references because the row or column summarises two entities, for example column 7 contains entities 1c and 4, which are both queries; or because decisions are summarised, decision [1] in column 4 can for example on a more detailed level be broken down into three separate decisions.

[1] Refer to the end of this chapter for urls.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	MODEL	1a) Concept/free text term	2) Candidate term list	Number of candidate terms [0]	Suitability of candidate terms [1]	3) Selected controlled term	Enough query terms [2]	1c/4) Query	Suitability of query terms [6]	5) Results	Number of results [3]	Suitability of results [4]	1b/6) Record	Aspects of record [5]	7) Collection of re
A	1a) Concept/free text term		1a-1												
B	2) Candidate term list			2-1	2-2										
C	Number of candidate terms [0]	2-4			2-3, 2 4										
D	Suitability of candidate terms [1]	2-12			2-6,2-7.2-9	2-8,2-10.2-11									
E	3) Selected controlled term		3-2				3-2	3-1							
F	Enough query terms [2]	4-7			4-6			4-1/4-4/1c-4				4-4	4-5, 6-9		
G	1c/4) Query						4-3		1c-1	4-2					
H	Suitability of query terms [6]					1c-3		1c-2							
I	5) Results										5-1				
J	Number of results [3]	5-3					5-6	5-3		5-5		5-4	5-3, 5-5		
K	Suitability of results [4]	5-7					5-11	5-7					5-7, 5-10		5-9
L	1b/6) Record						4-5						6-x	1b-1, 6-1	
M	Aspects of record [5]					6-6	6-7					6-2		5-9	6-3
N	7) Collection of records											7-2	7-3	7-1	

Figure 10.2 Tabular presentation of the model

The author proceeded to conduct searches in each interface, marking the connections between entities as she progressed. The tabular representation of the model was then compared to the entries in the tables for each interface. Some processes existed in the model but not in the interfaces and vice versa. Causes for each difference were examined. Some were due to the author's mistakes, others were indeed identified correctly.

The confirmed differences were entered into spreadsheets and missing and new connections were accumulated. Figure 10.3 shows the original table, where the separate rows and columns for decisions 1a, b and c, have not yet been condensed into one. This table has also not been updated according to the new version of the model described in this document.

The "1" in cell A2 for example shows that in one of the interfaces examined, no possibility exists to move from free text terms to a set of candidate terms. The F in cell G12 indicated that the Flamenco interface allows users to move from a selected controlled term to the number of records, because postings give information on how many records the term will retrieve. Each difference between this spreadsheet and the model was considered for the revision of the model, but not necessarily implemented.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		(1a) Concept/free text term	(2) Set of Candidate Terms	[0] No of candidate terms	[1a] Suitability of candidate terms	[1b] Suitability of candidate terms	[1c] Suitability of candidate terms	(3) Selected controlled term	[2] Enough query terms	(1c)/(4) Query	[6] Suitability of query terms	(5) Results	[3] No of results	[4] Suitability of results	(6) Record	[5] Aspects of record	(7) Collection of records
A	(1a) Concept/ Free text term		1														
B	(2) Set of Candidate Terms			1	CF												
C	No of candidate terms [0]	1			1												
D	[1a] Suitability of candidate terms					0	0	0									
E	[1b] Suitability of candidate terms						0	0									
F	[1c] Suitability of candidate terms	2			5			0									
G	(3) Selected controlled term		4						4	CFO			F				
H	[2] Enough query terms	2	5					F		0				2	2		
I	(1c)/(4) Query								0		3	0					
J	[6] Suitability of query terms							4	E								
K	(5) Results									CFL			0				
L	[3] No of results	0							0	3		0		0	0		
M	[4] Suitability of results	0							0						1		E
N	(6) Record								1			LFC			CF	0	
O	[5] Aspects of record							2	3	GFO				0	CFLO	G	5
P	(7) Collection of records											L			5		

Figure 10.3 Missing and additional processes identified in a preliminary version of the model

Legend:

C – CHIN F – Flamenco L - LRC OPAC
E – ERIC G - FACET 2 O - OVID (INSPEC)

Numbers indicate the number of interfaces in which the respective processes are missing.

10.5.2 Testing against existing data

During the SOCS study (chapter 5), the author developed a flowchart of the OVID interface (Appendix 5.8) from which a record sheet for gathering information on the participants' progress through their searches was derived. The author attempted to match processes from the model onto the flowchart. This proved to be problematic due to the different aims of model and flowchart. Although the flowchart shows all relevant entities, it does not cater for the decision making process and for example reformulation is thus highly condensed.

The author then tested the model against the data from the SOCS study. Occurrences of decisions and use of entities were identified and connections between them were established. At this point, the incompleteness of data became striking. Although the transcripts and record sheets give some idea of what happened during the sessions, many of the interactions are implicit. From the participants' comments, it is possible to reconstruct some of their interactions, but knowledge of the interface is essential, for example knowing that to view an abstract, the user has to open a record. One benefit of using the transcripts compared to the analysis of existing interfaces is that they explicitly report the decision making process with the consideration of different options. Analysing interfaces heuristically, the author could effectively only establish the existence of a decision point without being able to confirm considerations, for example as they are outlined in the process of selecting Candidate Terms. In the following, the author thus progressed similarly with data from some FACET 1 sessions (study reported in chapter 6) which were either particularly interactive or problematic.

10.5.3 Results of the tests

Overall, most of the processes identified in the model were confirmed through the tests. The author had to realise that certain decision processes are system-independent (e.g. in figure 10.3 cells D5, D6, D7, E6, E7 and F7, which represent the selection of Candidate Terms) and the respective columns were thus condensed. As expected, the FACET interface corresponds most closely to the model. Potential reasons for this are the fact that that the data collected with this system was instrumental in the development of the model. On the other hand, this interface is also the only one which did not have to deal with restrictions and functionality linked to Internet use. It would thus be useful to validate the model against other stand-alone interfaces in future work. One of the main (and sometimes only) differences in the tabular presentations is linked to the fact that most web-based interfaces limit the windows and panes they use and thus do not keep the set of candidate terms available to the user. Some of the interfaces also make little use of a controlled vocabulary. The OPAC for example use keywords, which are not even displayed in the records. In the ERIC interface, the thesaurus seems to be viewed as a help to get started. Once the initial query has been executed, the controlled vocabulary cannot be accessed, unless the searchers use the back button or start a new search.

Some changes had to be made to the presentation of the use of existing queries. In FACET, terms can easily be dragged from one query to another. Although many systems allow searchers to save queries for later use, these interfaces are mainly designed to work with the existing query, not to re-use terms for new queries. The model was adapted in order to accommodate this approach better.

Many of the differences identified initially proved not to be correct. Some missing connections are linked to optional components which do not exist in the systems concerned. One example is entity "(7) Collection of relevant records". Out of the interfaces inspected, only OVID provided users with the functionality to create such a collection. Other missing connections are linked to restrictions on interactions with controlled vocabulary terms. Although sets of candidate terms exist in the interfaces investigated, they are not accessible in some interfaces once the users start browsing the controlled vocabulary or select terms, which affects cells G2, G4 and H2. In many of the interfaces, the users did not have the choice of whether or not to add more Controlled Terms (process 4-6). In some interfaces, selecting the Controlled Term executes the query, so that the users can only consider more terms after the query has executed (processes 5-11). In interfaces where queries cannot be saved, it is not possible to assess the use of an existing query (1c-1). In interfaces where selecting a term executes a query, no zero hit results occur as only terms which have been used in indexing are shown. This means that the process 5-6, the reaction to reformulate a query that retrieves zero hits, does not exist in this interface. Users can still find the results unsuitable and reformulate without viewing any. This process is represented by the sequence of processes 5-4 and 5-11. In some of the interfaces, users cannot move from a record to reformulating their query (process 6-7) as indexing terms are not shown in the records (e.g. OPAC) or they are only text, so that the users can only take the concept (6-5) and map it to controlled vocabulary terms to use in the query.

In conclusion, although none of the interfaces investigated corresponded exactly to the model, the comparison resulted in the clarification of some processes presented and in the correction of others. The remaining differences are linked to limitations of the systems.

10.6 Using the model for design

In the design of information searching interfaces, the model can serve as a basis for the specification of interface and system functionality. The entities are represented through one or several components in the interface, and the interface can be designed to follow the information searching process as presented in the model. Designers can for example select those processes which they want to include in their system and using the Risk Points, consider the implementation of features to avoid certain risks. One of the strengths of the model in this respect is that it makes decision points as well as components explicit, thus combining functional and conceptual perspectives on the information searching process.

10.7 Using the model for evaluation

The model of information searching in a controlled vocabulary enhance system can also be used to identify shortcomings in interfaces.

In the first step of the evaluation, the components in the interface which correspond to the entities in the model need to be identified. In the next step, the relationships between them are established. The tabular view of the model can be used, and the process is similar to that described in connection with the testing of the model against existing interfaces (section 10.4.1). While inspecting the interface, processes are identified and entered into the table by marking the cells concerned. The table is then compared to the tabular model representation (Figure 10.2, Appendix 10.1). The evaluator then examines the differences and identifies explanations for their existence, such as the lack of a component like the Collection of relevant records (see example at the end of this section). Some processes are overlooked when establishing what is happening in the interface and should be revisited using the interface to assess whether some of the differences are based on these mistakes or whether the system evaluated differs from the model. One of the main confusions arises after results have been retrieved. It seems logical to say that the searcher moves from the results to the records, but in fact, the decision, even if the searcher is not conscious of it, of whether there are enough records (i.e. at least one) will always have to be made.

The model does not claim to be exhaustive, so differences are possible and can constitute an improvement of the model. However, they might indicate potentially missing features. Some interfaces, including FACET, for example do not allow searchers to store relevant records separately from the set of results. This limits their options in using the data they have found. In the tabular view, the respective connection (Cell M14, process 6-3) to entity "(7) Collection of relevant records" is missing as a result.

The Risk Points can be used additionally to draw attention to missing information or the potential risks at each decision point in the interface.

10.8 Current limitations

In section 10.2, the scope of this model has been addressed and it was pointed out that the model does not yet cater for free text searching. In order to assess the limits of the model, the author tested the model against Google (a free-text search engine), WebBrain (a dynamic interface to a subject directory), and the OVID interface to Inspec with a combination of free text and controlled terms. These data provided a glimpse of the effects that the controlled vocabulary has on interactions during the information searching process and showed that the model could be broadened. Processes regarding the viewing of results and query reformulation correspond to those described in the model. Differences were identified in the term selection processes, as the interface did not require searchers to map free text phrases to controlled Terms to construct their query. Further research using empirical data could thus permit the adaptation of the model to include free text or hybrid searching.

Other areas of the model require further investigation to confirm the model as presented in this document. One area which has been completely left aside is alternative data access methods, such as browsing of records, starting for example from a sample record or query by navigation, which would be an important method especially on the Internet or in many Internet-based system. Another aspect not yet investigated is the importance or frequency with which certain process or problems occur.

Some systems perform the selection of controlled terms without the users' intervention. It is possible that differences exist when the decision process has been automated as designers can decide on a different sequence of processes and decisions. However, according to the model, either the system or the users can execute the processes regarding term selection. Other systems use controlled vocabularies only after the query has been executed, which could influence especially the reformulation process.

Reformulation is a complex issue, involving aspects of the broader context beyond the immediate focus of the research conducted and consequently, this model. Thus decision [2] as described in section 10.8 and in the sub-diagram of query reformulation (figure 10.5) does not fully capture the complexity of query reformulation and should be understood as preliminary. In particular the consideration of concepts, options and query terms has not been analysed in depth. The current diagram assumes that the searchers consider concepts, terms and options one by one, but the data collected during the studies indicates that searchers also take different approaches. When they have executed the query a second time, they appear to make comparisons between the result sets in respect to the two queries. This process is influenced by several aspects and it is not reflected in this version of the query reformulation diagram, as large parts of the decision making process were apparently not expressed explicitly by participants during the search and not captured by other specific data collection methods either. Bates (1979b) and Fidel (1985) have both separately identified moves and tactics which represent possible choices during the search process. Bates discusses tactics related to monitoring, dealing with file structures, search strategies and term strategies independent of the medium. Fidel's moves complement these, but apply strictly to online searching. Therefore, they are more detailed especially regarding limiting queries to for example publication date and language. Using these moves and tactics and under consideration of existing literature on reformulation, for example Spink's work on different types of relevance feedback (Spink 1997) and Barry's on users' relevance criteria (Barry 1994), a study which focuses more specifically on the processes that take place during query formulation could be prepared. Collecting data for example by asking specific questions during and after the search process could lead to a better description of the potential problems encountered during query reformulation, which proved to be one of the most complex stages of the search process (see section 9.8.2).

10.9 Detailed description of the model

In this section, the model is described in detail. Although some of the processes are illustrated with implementation-dependant examples, this document does not intend to cover all possibilities. This applies particularly to cases where the searchers terminate their search sessions. It is understood that, if they have at this point identified useful records, they will perform actions necessary to make use of these, for example saving or printing them, as well as saving the query or logging out of the system. As these actions do not directly impact the information search processes, they are not presented in this model.

10.9.1 Identifying Candidate Terms

1a-1 – From Free Text Terms/concepts

In order to map a free text term or concept to a Controlled Term which can be used in the query, the searchers normally have to enter their phrase into a mechanism provided by the interface. The system will then retrieve entity (2) Set of Candidate Terms, which is the next element the searchers interact with (section 10.9.2, p. 184).

1b-1 – From an existing Record

If searchers have an existing record, this can serve as a source of query terms. The process of selecting terms is effectively identical to selecting terms from a record from the results set, which is discussed in section 10.9.10 - "Inspecting individual records". The search continues with entity (3) Selected Controlled Term (section 10.9.3, p. 187).

1c-1 – From an existing Query

Alternatively, the searchers assess an existing query to see whether they want to work with it or select individual terms from it. The assessment of an existing query leads to decision [6].

[6]

→ THE SEARCHERS USE INDIVIDUAL QUERY TERMS FOR THE CURRENT QUERY

1c-3

The searchers select query terms from the existing query to include in the current query. The search continues with entity (3) Selected Controlled Term, (10.9.3, p. 186).

→ THE SEARCHERS WORK WITH AN EXISTING QUERY

1c-2

An existing query can be reformulated and executed like a new or current query. Once the searchers decide to make modifications, the entity can be treated as entity (4) Query i.e. a current query (10.9.5, p. 190).

10.9.2 From (2) Candidate Terms to (3) Selected Controlled Terms

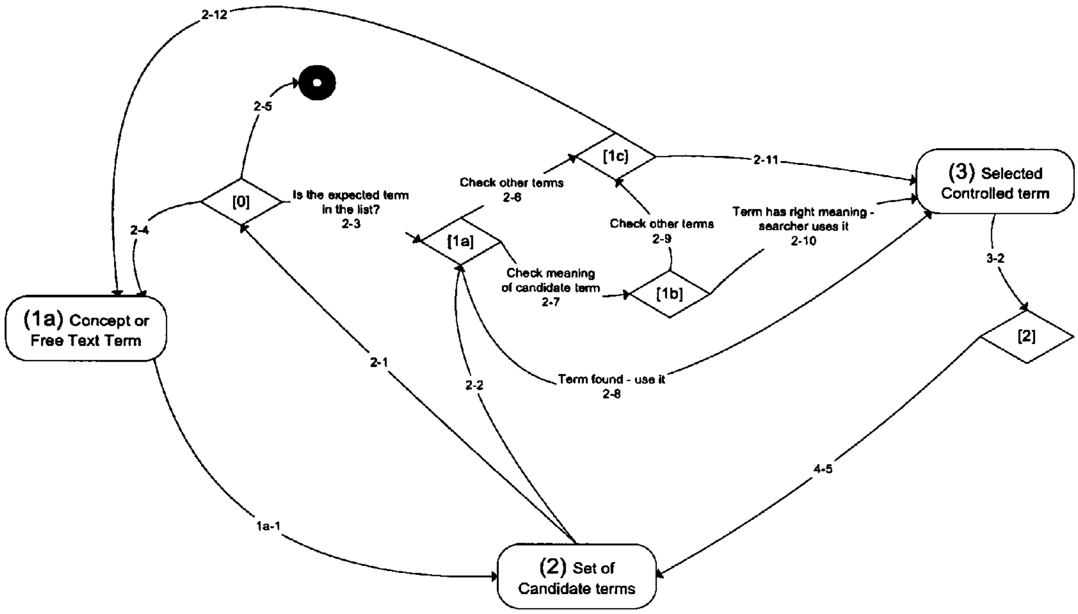


Figure 10.4 Term selection – detail

The "selection" of Candidate Terms referred to in this section refers to the process of considering a term and assessing its suitability. The selection can for example by made by the system. Searchers will also have different approaches to making a selection. Depending on the system implementation, they could browse the controlled vocabulary starting from the Candidate Terms or use other means, for example scope notes, in order to clarify the meaning of terms. The different options have to be considered when designing a system.

2-1

The Set of Candidate Terms is returned. This leads to decision [0], the assessment of the number of Candidate Terms. Some systems perform the selection themselves according to specific rules, implemented in the system. The decisions and their sequence can in this case be modified.

2-2

Some systems allow searchers to browse and select Controlled Terms without having to map concepts to the controlled vocabulary. This functionality leads from the Candidate Terms straight to decision [1a], the assessment of whether any of the available Candidate Terms conform to the searchers' expectations (p. 186).

[0]

→ NUMBER OF CANDIDATE TERMS IS 0, TOO LARGE OR INSUFFICIENT

Through algorithms, the system can detect a problematic number of records and execute alternative retrieval procedures that use for example more advanced stemming and natural language processing to avoid these problematic situations.

2-4

The searchers return to the starting point, (1a) Concept or Free Text term (10.9.1, p. 184).

2-5

The searchers terminate the search.

→ NUMBER OF CANDIDATE TERMS IS ACCEPTABLE

2-3

The searchers look for the expected term, i.e. the phrase entered into the mechanism, or one that is close. This leads to decision [1a], the assessment of whether the expected terms have been mapped or not.

[1a]

→ THE FREE TEXT TERM HAS BEEN MAPPED TO A SIMILAR CONTROLLED TERM

2-7

The searchers inspect the Controlled Term, for example by viewing the term in its local context in the controlled vocabulary or reading the scope note. This leads to decision [1b], the evaluation of the term's meaning.

2-8

The searchers select the Controlled Term. The search continues with entity (3) Selected Controlled Term (10.9.3, p. 187).

→ THE RETRIEVED CONTROLLED TERMS DO NOT RESEMBLE THE FREE TEXT TERM THE SEARCHERS ENTERED INTO THE MAPPING MECHANISM.

2-6

The searchers inspect other Candidate terms. This leads to decision [1c], the assessment of other Controlled Terms' suitability (p. 187).

[1b]

→ THE TERM HAS THE EXPECTED MEANING.

2-10

The searchers select the Controlled Term. The search continues with entity (3) Selected Controlled Term (10.9.3, p. 187).

→ THE TERM DOES NOT HAVE AN APPROPRIATE MEANING.

2-9

The searchers inspect other Candidate Terms. This leads to decision [1c], the assessment of the suitability of Candidate Terms.

[1c]

→ CANDIDATE TERMS ARE NOT SUITABLE.

2-12

The searchers return to the starting point, (1a) Concept or Free Text term and modify the entered phrase or think of a new one and recommence the mapping process (10.9.1, p. 184).

→ CANDIDATE TERMS ARE SUITABLE.

2-11

The searchers select appropriate Candidate Terms for the query.

10.9.3 From (3) Selected Controlled Terms to (4) the Query

3-1

Some systems, especially those based on the Internet, allow searchers to execute queries by clicking a term, for example implemented as a hyperlink. Searchers are not involved in the query set-up and modifications are made after the query execution. In these systems, the search continues with executing the query (10.9.6, p. 190).

3-2

The searchers add a Selected Controlled Term to the query, for example by checking a box or dragging the term across.

Selection of a Controlled Term leads to decision [2], the considerations regarding set-up and modification of the query. These processes are modelled in a (preliminary) sub-diagram shown in figure 10.5.

[2]

In the more detailed diagram of the query reformulation process (figure 10.5), decision [2] has been split into [2a] (p. 188) and [2b] (p. 190). [2a] represents the decision on which changes are to be made while [2b] represents the decision whether or not the query needs to be modified further.

The following sections describe the query reformulation process together with the query set-up because these are very similar. However, some considerations on setting options only apply after the query has been executed at least once.

[2a]

→ THE SEARCHERS CHOOSE TO RECONSIDER CONCEPTS.

When reconsidering concepts, the searchers refer back to the session information, for example concepts from retrieved records; reflect on the task requirements, for example extract more concepts from the task, or consider the task requirements in the light of the retrieved results, for example to see whether concepts need to be refined in order to retrieve better matches. When making these considerations, the previous query and its results as well as co-occurrences of terms in records play a role. The choice to reconsider concepts leads to decision [7] (p. 189).

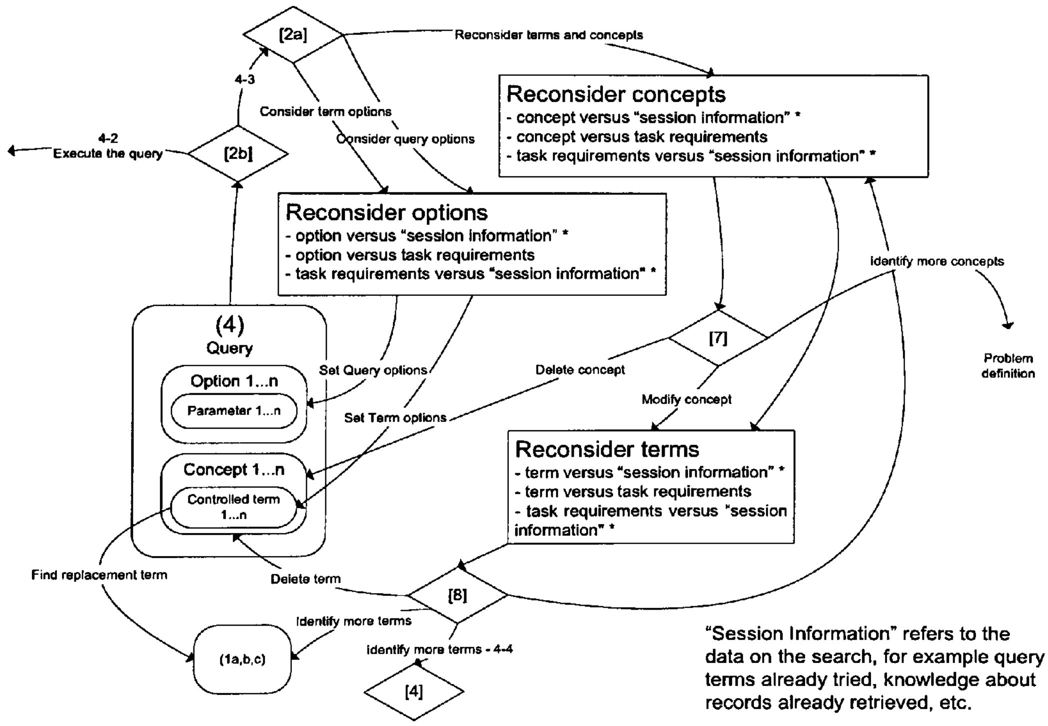


Figure 10.5 Query reformulation

→ THE SEARCHERS CHOOSE TO RECONSIDER THE QUERY OPTIONS.

When reconsidering the query options, searchers refer back to session information, for example the experiences they have made with the current setting; or to the task requirements, for example whether they need the document to be of a specific format.

They can compare the task requirements to the session information, for example to judge whether they need to broaden or narrow their search.

Depending on the outcome of this process, the searchers set the query options or term options, which takes them to decision [2b], the decision on whether the query is ready to be executed (p. 190).

→ THE SEARCHERS CHOOSE TO RECONSIDER QUERY TERMS.

When reconsidering query terms, the searchers refer to the session information, for example whether the term retrieved any records; the task requirements, for example how well the term represents concepts of the task, and compare session information with task requirements, for example to see whether other terms in the records retrieved represent the concept better. This leads to decision [8], the consideration of individual terms.

[8]

→ THE SEARCHERS DECIDE TO IDENTIFY MORE TERMS.

The searchers return to entity (1) Starting point (10.9.1, p. 184) or select other records from the results set ([4] Assessment of the results retrieved, 10.9.9, p. 192) to identify more terms.

→ THE SEARCHERS DECIDE TO DELETE A TERM.

The searchers delete a term from the query. This leads to decision [2b], the decision on whether the query is ready to be executed (p. 190).

→ THE SEARCHERS CHOOSE TO RECONSIDER CONCEPTS.

This process has been described above under decision [2a] (p. 188). The reconsideration of concepts leads to decision [7] Assessment of the suitability of concepts.

[7]

→ THE SEARCHERS DECIDE TO DELETE A CONCEPT.

The searchers delete the relevant terms from the query. This leads to decision [2b], the decision on whether the query is ready to be executed (p. 190).

→ THE SEARCHERS DECIDE TO IDENTIFY MORE CONCEPTS

The searchers go back to the problem definition in order to identify more concepts. This leads to entity (1a) Concept/Free text term (10.9.1, p. 184).

→ THE SEARCHERS DECIDE TO MODIFY A CONCEPT

This leads to the reconsideration of query terms. The searchers select a term to replace and to look for a replacement, they return to entity (1), their initial Starting point (10.9.1, p. 184), or the problem definition, which leads them to (1a) Concept/Free text term, or to suitable records in the results set ([4] the consideration of the results as a whole, 10.9.9, p. 192).

[2b]

→ CONCEPTS ARE MISSING FROM THE QUERY, OR NOT EXPRESSED SUFFICIENTLY BY QUERY TERMS.

10.9.4 Options for identifying and including more terms in the query

4-4

The searchers return to the result records (decision [4], 10.9.9, p. 192).

4-5

The searchers return to the record they were looking at (10.9.10, p. 192).

4-6

The searchers return to entity (2), the Set of Candidate Terms (10.9.2, p. 185).

4-7

The searchers return to entity (1) Starting point (10.9.1, p. 184) or reconsiders the task definition.

→ SET-UP PARAMETERS NEED TO BE CHOSEN OR MODIFIED

10.9.5 Options for query set-up

4-1

The searchers set up the query. All steps for query set-up are interface and system-dependent, i.e. not all are required, the system can take some of the decisions and impose a sequence in which they have to be selected. Possible options are:

- Arranging terms – applying a syntax
- Selecting a collection
- Setting query options such as language, document type….
- Setting term options, for example weights, expansion, etc.

4-3

This process allows for the fact that several options might have to be set by returning to decision [2] (10.9.5, p. 190).

10.9.6 Executing the query

4-2

The searchers submit the query after set-up. In a dynamic system, the execution is triggered by the modifications.

5-1

The system retrieves records which match the query from the database and presents them as entity (5) Results to the searchers. This leads to decision [3], the assessment of the number of records retrieved.

[3]

→ NO, TOO MANY OR TOO FEW RESULTS ARE RETRIEVED.

10.9.7 Modifying the query

If a problematic number of result records is retrieved, the system can use algorithms which increase the expansion or search a different collection.

5-2

The searchers terminate the search.

5-3

The searchers return to entity (1) Starting point (10.9.1, p. 184) or reconsiders the problem definition.

5-6

The searchers decide to make modifications to the query.

→ A REASONALBLE AMOUNT OF RESULTS HAS BEEN RETRIEVED.

10.9.8 Inspecting the results

The options mentioned above in connection with a problematic number of results (5-2, 5-3, 5-6) still apply. Additional options are:

5-4

The searchers consider the results retrieved by the query as a whole. This leads to decision [4], assessment of the results retrieved (10.9.9, p. 192).

5-5

The searchers open a retrieved record in a random manner. The search continues with [5] (in 10.9.10, p. 192).

5-9

The searchers add useful or suitable records to entity (7) Collection of relevant records (10.9.11, p. 193).

[4]

→ THE RESULTS DO NOT FULFIL THE SEARCHERS' EXPECTATIONS.

10.9.9 Options for modifying the query

5-7

The searchers return to entity (1) Starting point (10.9.1, p. 184) or reconsider the problem definition.

5-8

The searchers terminate the search.

5-11

The assessment of the query results leads the searchers to make modification to the query set-up (decision [2], 10.9.5, p. 187).

→ THE RESULTS FULFIL THE SEARCHERS' EXPECTATIONS.

10.9.10 Inspecting individual records

The options mentioned in connection with unsuitable query results (5-7, 5-8, 5-11) still apply. Additional options are:

5-10

The searchers choose to inspect a specific record. This process can involve selecting or opening the record or a shift in focus from the set overall to the individual record.

6-1

Once a record has been selected, the searchers inspect this record in detail and extract information from it. This can require them to move from one aspect of the record to another aspect, for example from description to indexing terms. In order to for example use indexing terms from the record in the query, their suitability needs to be judged.

6-8

In some interfaces, searchers have the option to move from one record to the next without returning to the record set. This move does not require the searchers to make a decision on which record to view, as the sequence is generally predetermined by the system.

Inspecting the record in detail leads to decision [5], the assessment of whether and how to use the information from the individual record contained in the different aspects.

[5]

→ NO ASPECTS OF THE RECORD ARE OF PARTICULAR INTEREST.

6-2

The searchers return to the record set to select a different record or to take any of the other options possible that that point. This leads to decision [4], assessment of the results retrieved (10.9.9, p. 192).

6-4

The searchers terminate the search.

→ ASPECTS OF THE RECORD ARE OF INTEREST.

6-3

The searchers add useful or suitable records to entity (7) Collection of relevant records.

6-5

The searchers return to entity (1) Starting point (10.9.1, p. 184) or reconsider the problem definition. The inspection of individual records is probably the point in the search session where the searchers learn the most about the topic of the search and its representation in the system. Therefore, this process has been explicitly included in the diagram. However, the acquisition of knowledge is not a separate process but occurs alongside the others.

6-6

After assessing its suitability, the searchers select a Controlled Term from the record. The search continues with entity (3) Selected Controlled Term (10.9.3, p. 187).

6-7

The searchers decide to make modification to the query set-up. This leads to decision [2] (10.9.5, p. 190).

10.9.11 From the Collection of relevant records

7-1

From entity (7) Collection of relevant records, the searchers can return to the record they were inspecting in order to interact with other aspects, for example select Controlled Terms, or gather more information. The search continues with decision [5] (10.9.10, p. 192).

7-2

After adding records to the collection, the searchers can return to decision [4], assessment of the results retrieved (10.9.9, p. 192), for example to choose another record to inspect.

7-3

The searchers can select a record from entity (7) Collection of relevant records to use as a starting point, which leads back to entity (1) Starting point (10.9.1, p. 184).

8-1

After considering current search information extracted from a record, the searchers focus again on the aspects of the record in order to interact with other aspects, for example selecting Controlled Terms or gathering more information. The search continues with decision [5] (10.9.10, p. 192).

II)

From (8) Current Search Information, the searchers can also generate new free text terms or concepts and effectively recommence at the beginning of the search cycle by mapping these new terms to the controlled vocabulary (entity (1a) Concept/Free text terms 10.9.1, p. 184).

10.10 Urls for chapter 10

CHIN: http://www.chin.gc.ca/Artefacts/e_artefacts_canada.html

ERIC: http://searcheric.org/scripts/ewiz/amain5.asp

Flamenco: http://bailando.sims.berkeley.edu/flamenco.html

LRC OPAC, University of Glamorgan: LRC OPAC:

http://waclient.lrc.glam.ac.uk/www-bin/www_talis1

OVID Inspec: http://edina.ac.uk/inspec/index.shtml

WebBrain: http://www.webbrain.com/open_IE.htm

Chapter 11 - Risk Points

11.1 Introduction to Risk Points

The Risk Points (RPs) complement the model diagram and description discussed in chapter 10. This part of the documentation details potential errors and problems which occur at various stages, especially at decision points, of the search process and suggests potential solutions. Section 11.2 provides details on some of the recurring points regarding risks, solutions and limitations. Conclusions on dealing with risks in information searching are discussed in section 11.3. Section 11.4 explains the Risk Point documentation at the end of this chapter, to which table 11.1 (p. 205) gives a brief overview.

When the documentation refers to factors or influences, these are normally from within the search session, or closely related, such as the situation where searchers are already aware of retrieved records. Of course, other broader factors will influence searching behaviour, for example the intended use of the information or the domain (section 3.3).

For all suggested solutions, the question remains whether the implementation is feasible with the programming language and mode of delivery used to implement the system and interface in question, for example not to increase the response time unacceptably.

Interface display space will be an issue with most Risk Points discussed. Being able to show more information will normally enable searchers to gain a better understanding of the processes, but the display has to be readable and ideally not obstruct the search interface so that users can easily refer back to their query. Some types of records contain large amounts of data which need to be displayed so that designers are hard-pressed to include any additional information. The RPs do not prescribe how support should be implemented. Some options are discussed in section 3.5. The main focus here is to suggest how data in the system, especially the controlled vocabulary, could be exploited more thoroughly for the benefit of searchers.

11.2 Explanations of risks, solutions and limitations

Certain issues apply throughout the search process. In order to keep the description of Risk Points brief, these have been summarised at the beginning of this section and are referred to later.

11.2.1 Risks

General

Confusion and frustration are two of the main risks whilst using any type of interface. In the context of information searching, they can result in searchers abandoning their query without retrieving the required information. Searchers can take (apparently) reasonable actions, and not retrieve the expected results, which can make it difficult for them to resolve the problem. Searchers can also follow interesting leads which do not immediately relate to the tasks they set out to resolve, and have subsequent difficulties in returning to the point where they left off. Another issue which applies particularly to Risk Points [1] and [4] (Assessment of query terms and records) is the time it takes users to find what they are looking for. However, this issue extends to the effort searchers have to make after the search, for example if they have taken away a large set of results which contain many records that are not useful, or if they realise upon closer inspection that they have to return to the system to conduct further searches (e.g. [4].2).

Selection difficulties

Selection difficulties can occur during the selection of Candidate Terms, query options and result records. Factors like large sets make the selection of Candidate Terms and results difficult. The searchers can encounter difficulties in identifying the items they require, or if they resemble each other, it can be difficult to distinguish between them. Searchers can also have problems in identifying the correct controlled vocabulary terms if concepts need to be expressed using several terms. If any terms or records are retrieved, possible support mechanisms to identify further items are possibly not initialised, which leaves the users to their own devices. In terms of query options, the effects that different options have can be unclear to users so that the selection is difficult.

Selection of sub-optimal options

A number of factors can result in the selection of items or options which are detrimental to the query results or even result in the retrieval of zero hits. Selection difficulties and simply misunderstandings of the meaning and scope of terms contribute here as well as the fact that searchers do not browse the controlled vocabulary or that not sufficient information is available. Auxiliary information such as postings can mislead users.

11.2.2 Solutions

Training

Training is necessary to use most complex information searching systems. Some functionality can more easily be understood while it is used rather than through instruction or teaching. In order for this to happen, searchers need to be able to assess the effects of query modifications, especially as they vary from system to system. The system can highlight changes, for example in the order of results, from one query run to the next, or, in a more complex implementation, generate alternative queries. Hands-on exercises are an alternative to highlighting. Without knowing what exactly happens in the background, searchers can learn how modifications influence results and subsequently apply this knowledge actively. Developing this type of understanding is particularly important in systems where it cannot easily be predicted how exactly the change will affect the results. Highlighting changes gives searchers a more realistic impression of how effective the query modification was. Users who are new to the system will also benefit from information on how to interpret query-related information.

Paper-based support materials can help avoid space issues on the screen. While searching, short tutorials, possibly like Microsoft's "Show me" can help searchers to quickly learn about functionality. Searchers are not always able to state what kind of assistance they require, and a wizard can be used to guide them to the appropriate tutorial.

Wizards

Searchers can at times be unsure about how to progress with their searches. One possible way of assisting in this situation is a wizard. In various interfaces, wizards are used to assist searchers in making decisions on settings and selections. Wizards often hide plain option screens from the searchers by presenting them with several screens, each containing either more information or expressing the options in terms of aims and objectives. Examples for these wizards can be found in Microsoft products, in particular in Access, or Nero CD burning software.

The type of wizard suggested in this context does not necessarily guide searchers through the whole search process, but provides very specific support determined by data collected during the session and provided by the searchers themselves. Effectively an electronic intermediary, a wizard could for example be based on the decision-making tree based on professional searchers' behaviour regarding the selection of free text and controlled query terms developed by Fidel (1991b). "Who, what, where, when" questions are an alternative to help clarify the search task. The main uses of the wizard appear to be in suggestion of query options or terms which can be identified using techniques addressed in the following sections (particularly "Extracting information from data in the system"), or to guide searchers to the appropriate section in the help file. These suggestions will depend particularly on the stage in the current search session and the stage in the general information searching process, i.e. how far searchers have resolved the task for which the information is required. Depending on the domain, the wizard can include support for specific strategies, for example an author search in a library catalogue. Below is an example based on a simple tree structure that accommodates the searchers' reactions to retrieved results.

Screen 1

Please select the option which best describes your thoughts about these results:

1. I am not interested in records which contain a particular term.
2. Suggest other concepts which I could use in my query.

The searchers select option 1 and screen 2 appears, requesting users to enter the term(s) they are not interested in. Once this is done, the system adds them to the query using Boolean NOT or changes the necessary parameters to exclude the records concerned.

If the users selected option 2, the system first of all performs an analysis of the indexing terms of records retrieved. A simple algorithm could show all terms not matched by the query, a more advanced one might group these according to the structure of the controlled vocabulary. Co-occurrence analysis might hereby provide some indication of more useful concepts. Screen 2 thus shows the following information:

> The indexing terms below occur in the records retrieved and might be of interest for your query. Please select those which you are interested in by checking the boxes.
>
> [] Term1
>
> [] Term 2
>
> [] ...
>
> [] No terms are appropriate.

The searchers then select the appropriate terms and the system adds them to the query statement. By selecting "No terms are appropriate", the users can indicate to the system that other assistance is required. The system displays either options for actions which can be performed to generate more terms, or instructs the users on how they themselves might be able to identify more terms.

Access to items

Both Candidate Terms and results can be retrieved in large numbers. Search mechanisms or highlighting of specific indexing terms in records helps to quickly identify items of interest. It appears that when searching the controlled vocabulary for Candidate Terms, the searchers first try to find the terms lexically close to the phrase they entered - these could be highlighted initially. One efficient way of reducing the number of Candidate Terms retrieved is to display only those actually used in indexing. However, other terms could be more beneficial to use in connection with expansion mechanisms. Ranked information retrieval displays records closest to the query towards to the top of the list. Resorting result or grouping records by specific criteria are other methods to facilitate selection. The controlled vocabulary can be used to perform some of these operations. It should be noted that cut-off points might need to be defined in ranked systems in order to avoid distraction through records with low match values when resorting. A tool which allows users to search retrieved records for specific indexing terms or words in other fields would speed up the inspection of records as might the highlighting of query terms.

If searchers are likely to execute repeated searches in the interface, it can be helpful to maintain information such as records previously accessed from session to session, or within one session from search to search.

Extracting information from data in the system

Examples of records and queries can provide searchers with information on how to construct a query and which terms to include. Provision of sample queries covering a range of collection items potentially allows searchers to run an existing query which requires only minor modifications. Records can provide information on how to reformulate a query. Searchers can examine them manually, but are likely to be much less efficient than if this is done automatically. The analysis can be conducted on all records retrieved or only on those which have a certain match value. Some examples are:

- Terms that have actually retrieved records
- Other terms and concepts which have been used to index retrieved records, i.e. other terms/concepts that might work
- Terms which do not retrieve any records
- Terms which could be refined, for example if the searcher has searched for “clocks”, but most records are in fact indexed with the term “wall clocks”.

The controlled vocabulary can be used for example to “cluster” indexing terms by checking in particular the hierarchical relationships between them. Where appropriate, the users could be shown broader terms representing these clusters. This would result in a better representation of common concepts than if every term is considered individually.

Relevance feedback is another technique which can be employed in a number of different ways (see section 3.5).

Automated modifications

Algorithms can be used at different stages of the search process to improve the retrieval of Candidate Terms, the selection of query terms and settings of options. They can for example detect problems with the free text entry, for example spelling mistakes, spacing and syntax errors. They can otherwise limit or extend a search in the controlled vocabulary or the database. Fuzzy algorithms and string processing techniques such as stemming and wildcards or natural language techniques serve to increase the number of results retrieved. Expanding query terms along hierarchies of a thesaurus is another technique that can be used to this end. The expansion, i.e. the amount of terms included in the query, can potentially be adjusted automatically or indirectly, for example based on relevance judgements of records. In some searches, users benefit from using query terms not too close in meaning to the original ones in order to escape from a deadlock (e.g. Ruthven 2003a). In addition to increasing the expansion, the query could be modified, for example terms could be ignored if necessary. Alternatively, in hybrid systems, controlled vocabulary terms could be used as free text search terms, so that searchers retrieve at least some results, and can gain an insight into the collections and the indexing practice, which they can in turn apply to query reformulation. Changes in the expansion rate can also be used to reduce the number of records, should this be deemed necessary.

11.2.3 Limitations

One major limitation to many of the advanced suggestions is the confusion that they can cause, which potentially creates more problems than the solution resolved. If the searchers are for example not aware of automated processing, they might have problems to understand how query results relate to the query terms and set-up. If complex processes are explained or made explicit in the interface, this might also create confusion if searchers misunderstand the explanation.

Some solutions depend on the size of the database and the controlled vocabulary. It is for example possible that many records are retrieved with the same match value, so that it is difficult to distinguish them further.

Computational limitations

Computational limitations concern mainly three issues: the time required to execute the calculations, the complexity of the problem and the correct detection of the problem. Certain solutions suggested require long calculations, for example the "clustering" of controlled vocabulary terms to identify concepts, which can result in long delays while retrieving results or before the searchers can attempt query reformulation. False assumptions on the searchers' intentions and goals as well as incorrect identification or avoidance of a problem, can lead to problems if automated processes are triggered without consulting the users, or frustration if users are repeatedly interrupted unnecessarily. As far as the complexity of the problem is concerned, some require sophisticated monitoring of the search behaviour during the session. The necessary variables need to be identified, and their influence on each other has to be assessed carefully in order to avoid the above-mentioned false identification of problems.

Not all problems which are identified correctly can be resolved easily. For example when attempting suggestions to reformulate a query, it can be difficult to predict which modification or combination thereof will lead to the best results. Searchers often have several options, for example to replace a term or to change a setting, and the more options are available, the more complex it will be for the system (and the searchers) to assess the possibilities.

Some issues already need to be resolved at the design and implementation stage of the system. Showing postings can for example be difficult in a system that used expansion, as the number of associated records varies depending on the expansion rate. Some techniques are limited by parts of the system itself, for example when using a non-hierarchical controlled vocabulary, no expansion can be used at all.

11.3 Conclusions on dealing with risks

Generally speaking, three areas in which solutions to the information searching problems addressed in the Risk Points can be identified: the training or support during the search in particular with regards to the stages of the information searching process, the presentation of data and the development of controlled vocabulary based tools. Training and support have already been discussed in section 11.2.2. Much of the actual implementation will probably depend on the specific system, although the FACET studies suggest that more general solutions exist, for example the support of the search process within the interface. Research on computer-based learning and help manuals can inform the development of solutions in this area.

Application of existing findings in information visualisation research can assist in the sensible presentation of data, for example to highlight query terms in records or to show changes in the results after re-executing a query. This area is related to the controlled vocabulary tools, for example when suggesting additional query terms, consideration needs to be given to their presentation. Research in the use of controlled vocabulary and thesauri in particular has focused on assistance in query formulation, for example through suggestion of terms from the thesaurus which match entered free text phrases (OKAPI project, section 2.4.2) or query expansion through use of the hierarchies as in FACET. More recent interfaces like Flamenco have used the hierarchies for a novel way of browsing records. However, further possibilities exist and should be explored as they might limit some of the issues associated with controlled vocabularies, such as the additional cognitive load required in order to make efficient use of them.

11.4 Presentation of Risk Points

For easier access to individual Risk Points, table 11.1 (page 204) summarises information from the documentation on the Risk Points. The table for example shows more clearly the conditions which can apply at each decision point. The column "Broad contexts" permits at a glance information on which Risk Points are relevant to the search stage of "Identifying controlled vocabulary terms". The columns correspond to parts of the documentation, which are now described in detail before being presented.

The description of each RP follows the same format and is presented in the form of a table. The first field contains the name of the Risk Point. In order to distinguish Risk Points from decisions, entities and processes, each identifier starts with "RP", followed by the decision or process it refers to. At many decision points, searchers have under different conditions various options to progress (for example no, too many or an acceptable number of records retrieved). These options are distinguished through a full stop followed by a number.

The **Type** explains whether the Risk Point represents an unexpected outcome (e.g. zero hits) from which the searchers needs to recover (Error recovery), whether they need support with a decision (Support/Facilitation) or whether they need background information on the search process (Support/Instruction).

The **Broad Context** signifies the general activity within which the situation occurs, for example "Identifying suitable controlled vocabulary terms", and the **Model reference point** refers to the exact position in the model documentation and diagram. The **Condition** distinguishes between several possible outcomes (for example for the number of records retrieved) because the risks depend on these conditions. The condition that either follows automatically after the previous step or which corresponds to the least problematic/best condition is marked as "Ideal/Routine outcome", see for example RP[1]. For other conditions, the section **Potential Causes** gives some information on why searchers might encounter the problematic situation. The points mentioned here are not exhaustive as causes can also be system-dependant. However, when dealing with a risk, it is useful to consider why a problem occurred as it could potentially be prevented in the first place.

Risks show some of the problems that can arise if the situation is not dealt with appropriately. A common risk is that searchers give up on their search, but they can also progress in a way that causes them further difficulties.

Solutions/support options outline ways to deal with the particular situation. These points refer to some extent to the potential causes and offer solutions, but also include suggestions to support the progress of the search. Solutions cannot necessarily be implemented exactly as described here because whilst resolving one issue they produce another difficulty. Hence the specific design of the system, the data types in the collection and the domain need to be considered. The section **Limitations to implementation** draws attention to some of the problems associated with solutions.

Minimum information required refers to essential information which searchers need in order to continue with the search. **Other potentially useful information** lists other data which can help users in making informed decisions. This section refers to data that is probably available to the system during the search process anyway, but which is not made visible in a meaningful way. Information referred to in the Solutions normally requires more involved implementations of additional processing methods.

Other Behaviour refers to other choices of actions that searchers have at this point of the search because the existence of a risk only indicates a possibility that searchers may choose a detrimental action. One major factor on this choice is that searchers who have encountered difficulties can be affected on an emotional level by the situation they are confronting; one of the most obvious manifestations is frustration.

Other factors which influence behaviour at this point lists reasons why searchers behave in one way rather than another. These factors cannot necessarily be identified in the situation itself, but only when considering the search session or its context as a whole. If somebody for example runs a query weekly to monitor a collection, the person might be content with zero hits. If a searcher reformulates a query repeatedly without retrieving results, they are likely to become frustrated because they are unable to find relevant records. Factors mentioned in this section go beyond influences from within the search session. They were all observed during the studies or are commonly accepted, and it was thus considered acceptable to include them, especially as they have direct implications for the search behaviour discussed. In RP[7]/[8] for example, the query can potentially not be reformulated if the restrictions related to the task can make it impossible for searchers to select other query terms. This section does not attempt to present all possible influences on a search though, as this would require further analysis and it is potentially prohibitive due to the vast amount of factors (see 3.3 for a brief overview of some areas).

On the following page, the Table 11.1 now gives an overview of the Risk Points, and RP[4].1 is then presented as an example. The complete Risk Points can be found in Appendix 11.1. The reader is reminded that some of the points mentioned in this documentation have been described in section 11.2 and are here only mentioned using a keyword.

Broad context	Model reference point	Condition	Name	Type
General approaches to searching	III) – Permanent influence on searching behaviour	N/a	RPIII	Support/Instruction
General approaches to searching	(1a), II) - Preparing to enter free text terms/phrases in the controlled vocabulary search mechanism	N/a	RP(1a)/II)	Support/Instruction
Identifying controlled vocabulary terms	[0] Number of candidate terms retrieved	The number of candidate terms is 0 or insufficient.	RP[0].1	Error recovery
		The number of candidate terms is too large.	RP[0].2	Support
	[1] Locating terms for the query amongst candidate terms.	The number of Candidate Terms is acceptable and the searchers attempt to identify suitable terms.	RP[1]	Support/Facilitation
Setting up the query	[2] Modifications required before executing the query	- Searchers have selected at least one controlled term for the query.	RP[2]	Support/Instruction
Query reformulation	[7] Reflection on concepts/ [8] Reflection on individual query terms - reflection on query options.	The searchers have either executed a query or they are in the process of setting up the first query.	RP[7]/[8]	Support/Facilitation
Assessing results retrieved	[3] Number of result records retrieved	No or too few results are retrieved.	RP[3].1	Error recovery
		Too many results are retrieved.	RP[3].2	Support
Assessing the records retrieved	[4] Relevance of retrieved results	An appropriate amount of records has been retrieved. The searchers judge the records as generally not suitable.	RP[4].1	Error recovery
		An appropriate amount of records has been retrieved. The searchers judge the records as not quite suitable or satisfactory.	RP[4].2	Support
		An appropriate amount of records has been retrieved. The searchers judge the records as generally suitable.	RP[4].3	Support/Facilitation
Inspecting individual records	[5] Aspects of a record to interact with	The searchers have selected a record from the results set which is of interest to them.	RP[5]	Support/Facilitation

Table 11.1: Summary information on Risk Points addressed

11.5 Exemplary discussion of Risk Point [4].1

RP [4].1 represents a situation where a seemingly appropriate amount of results has been retrieved, which can raise the impression that the records are suitable, so that automatic query reformulation or other support processes are not triggered. However, in this situation, the records retrieved do not fulfil the searchers' expectations. They might believe that no matching records can be found in the databases searched, which leads them to stop their search and to potentially miss important information. If searchers decide to continue their search, they can encounter serious difficulties in trying to retrieve more or different records. An example of a searcher who got stuck with three records even after repeated, sensible query reformulation is discussed in section 6.7.9 (from the FACET 1 study) and in 9.7.4 (from the FACET 2 study). The particularity of this situation is that as opposed to other cases (RP[4].2 and RP[4].3), the searchers potentially have no information to work with.

As far as reasons for this search outcome are concerned, sometimes, inappropriate query terms have been selected. This can happen because they are used differently in the context of the collection than what the searchers expect. In order for them to reformulate their queries successfully, they need to become aware of this issue. Through highlighting, resorting or searching on the results set, searchers can identify records where specific query terms occur and assess whether the indexed records correspond to the type of items they are looking for. Those functionalities can also speed up the process of assessing records and selecting those to inspect in detail.

RP[4].1	**Type: Error recovery**
Broad context	
Assessing the result records retrieved	
Model reference point	[4] Relevance of retrieved result records
Condition	
An appropriate amount of records has been retrieved, but the searchers judge the records as generally not suitable.	
Potential causes	
- Selection of an inappropriate term - Inappropriate indexing - The combination of query terms represents concepts other than those intended by the searcher.	
Risks	
The searcher makes (reasonable) modifications to the search which do not lead to better or different results.	
Solutions/support options	

- Searchers need to be made aware of the reasons for which unsuitable records have been retrieved. This could be done through displaying a list of suggestions, or be specifically based on the results retrieved (e.g. through an analysis of co-existing indexing terms). - Access to items - Automated modifications
Limitations to implementation
The information available in the controlled vocabulary or records can be too limited to be useful.
Minimum information required
Displaying information on the retrieved documents, document surrogates or records.
Other potentially useful information
- The match values that the system has calculated for the record, especially where this is part of the retrieval process (for example in a ranking system) - Information on why the record is considered to be a match for the query, for example highlighting relevant terms in the record or representing the record in a visual way, for example using bars etc. - Information on the scope of the query terms used if they stem from a controlled vocabulary to make searchers aware of inappropriate terms they might have used.
Other Behaviour
The searchers inspect results records and find reasons why their query has not retrieved more appropriate records. They reformulate the query by changing query settings or terms according to those reasons.
Other factors which influence behaviour at this point
- Searchers' ability to understand why some records are considered more relevant than others and how the match with the query comes about will make it easier for them to reformulate their queries. - Quality of indexing, i.e. the detail and the correctness. - Use of other systems might have lead searchers to establish what they consider to be a "good" match, for example in cases where a percentage is given as an indication of closeness to the query. - Search experience or experience with the particular system.

Chapter 12 – Conclusions

12.1 Introduction

The research conducted for this thesis set out to investigate how end-users interact with thesauri, to identify their potential problems in using this tool and to find solutions in order to solve or reduce these. These research questions required the development of an effective methodology for data collection and analysis, which formed part of the objectives for this research project. During the studies, the aspect of integrating thesauri further into the information searching process was raised as an additional point of discussion, and the development of a model of information searching was chosen as an appropriate means to synthesise and generalise findings so that they could inform systems design, evaluations and user studies.

In two preparatory and two in-depth studies, the author combined a number of qualitative and quantitative data collection and analysis methods such as think-aloud protocols and content analysis, application logging, observation and description of sessions, which provide detailed accounts of information searching behaviour in systems with integrated controlled vocabularies. Detailed analysis of these data allowed both further development of an experimental prototype, FACET, as well as extraction of enough information to develop a model of information searching behaviour in controlled vocabulary-based systems which can be used in the development and evaluation of such systems. This model has the particular objective to draw the readers' attention to potential problems which can occur at different stages of the information searching process, which could be avoided or resolved through appropriate design decisions. The information searching processes for example needs to be specifically supported in the interface in order to avoid conceptual problems during the search.

This chapter reviews first the methodology applied, then describes findings and the original contribution to knowledge and finally gives an overview of future research based on the work carried out up to date.

12.2 Review of methodology

During the four studies conducted, the author employed a number of data collection techniques depending on the circumstances under which the individual study was conducted. Variations of questionnaires were used to collect information on the participants' searching backgrounds, but also for session-specific information. In the first study, the author was present during the sessions, but only supervised the general progress without observing individual participants. All later studies were conducted on a one-to-one basis and during these sessions, audio recordings of the discourse between author and participant were made, then transcribed and analysed in detail for indications of problems and misunderstandings. Open coding proved useful here as it permitted the identification of issues not previously anticipated.

The author attempted for all studies to collect information on the participants' actions. In the first study, this was left up to the participants, then the author used a record sheet which required her to make manual entries and later in the FACET studies, application logging and screen capture software were available. The latter files were transcribed for easier use, and the log files processed for a number of purposes, such as the analysis of free text terms entered into the mapping mechanism and the query terms used. The data collected was collated to reconstruct a better picture of each search session. The following sections now discuss the author's experience with the methods used.

12.2.1 Evolution of methodology

Especially during the initial studies carried out for this thesis, the assessment of suitable methodology was an important aspect of the investigations, but even during the in-depth studies conducted with the FACET system, the author still aimed at optimising the methods employed. During the later studies, the author's research interests crystallised. While the initial studies were geared towards investigation of thesaurus navigation and use, the thesaurus was later considered as a tool within the search process that can be beneficial in different ways. This change of focus of the research as well as the availability of more tools, for example logging and screen recording software, allowed a move from primarily quantitative, rather limited methods to a combination of quantitative and qualitative methods. Using triangulation of methods was one of the key elements in data collection. It allowed the analysis to focus on qualitative aspects extracted primarily from participant-evaluator discourse, which were backed up by other methods, including quantitative data such as the log files which revealed very detailed information on the participants' information searching behaviour.

12.2.2 Training and instruction

During the studies conducted, it was always appropriate to instruct participants in the use of the interface they were expected to search with. In a few cases, the search session took place in connection with a consultation of the participants' opinion on the interface or system. The author found that too much information which was not directly relevant to the search session confused participants. If they were for example shown tools under development, users enquired how they could be employed in the context of the tasks. Thus, before a search session, only the features necessary to conduct the search should be introduced. It also helped participants if either menus or buttons were used in the training, even if both performed the same operations. On the other hand, participants need to be made fully aware of the functionalities of the tools they are expected to use, especially as novel functionality is particularly difficult to get used to and it is in the interest of an evaluation that users actually attempt to use these features. During the FACET 1 study (chapter 6) for example, some users tried to locate very specific terms. They struggled with this as they lacked the necessary subject knowledge, and the query expansion moreover made it unnecessary to find the most specific suitable term to begin with. These problems might have been avoided if the training had been explicit enough on the advantages of the query expansion.

The author found that hands-on training exercises were beneficial in re-enforcing the participants' ability to use an interface quickly. An accompanying hand-out was particularly useful as participants could refer back to it at a later point. This type of instruction was appropriate here as assessing searching behaviour was considered more important than evaluating the users' capabilities of using an unfamiliar interface.

The setting of the individual sessions varied greatly. Under these conditions, it is helpful to structure the training as far as possible, for example using an automated presentation rather than one held by the researcher, in order to increase consistency. PowerPoint proved to be a useful tool for this as the easy-to-create animation allowed some realistic presentation of interactions and the package also enabled the playback of screen capture video of more complex interactions.

12.2.3 Real life tasks versus scenarios

Whereas limitations associated with scenarios or set tasks are well known, the author also experienced some problems with supposed real-life situations. The main restriction which applied to all the studies conducted was that users could not be observed over a long time period. In the FACET studies, the system was only available to the participants during the their study session. In the SOCS study (section 5.4) where participants used a fully-functional system, no data collection methods for long term studies such as application logging were available. Participants were asked about their habitual searching behaviour and the author had to realise that although these users had real Information Needs, which they normally tried to fulfil using the investigated interface, they did not have an

Information Need at the time of the search sessions. Monitoring the literature is one of their regular information searching tasks, but this type of behaviour was not of particular interest to the investigation. In a setting where participants are recruited for a study at the time when they have an Information Need, for example when they approach a library OPAC system, much more relevant data could be collected. In the FACET studies, the author had to give participants set tasks as they did not use the interface and many of them did not use the collection data in this context for their regular work tasks. However, most participants in these studies search for information mainly because of customer requests or for a colleague. Therefore, a set task reflects to a large extent a real-life situation.

When designing a set task, the researcher should consider the features of the interface or the type of interactions which are to be observed. For example, if the evaluator wants to investigate browsing of the thesaurus, this activity should form a logical component of the process required to resolve the task. As discussed in Chapter 6.8.3, the desired behaviour does not always occur during a session. In order to be able to observe certain behaviours and to collect relevant data, the author developed "evolving scenarios", which were to be adapted according to the participants' progress and search behaviour in the FACET 2 study (chapter 8 and 9). However, she had to realise that it was difficult to steer the participants' behaviour to the extent imagined, and it also seemed inappropriate to influence the natural progress of the session too much. In fact, the main resulting action was to prompt users to continue the search by pointing out a possible interaction they could take, or by indicating that the participant could maybe retrieve better results. Although evolving scenarios are thus useful for the investigation of problems encountered, for example with specific interface features, the analysis of the search processes becomes less relevant. It appears that the evolving scenarios make the sessions more uniform to some degree, in that some are extended beyond the point where a user might stop searching. However, some sessions still remained fairly short (10 minutes) despite the author's efforts.

When designing tasks for sessions, researchers should accommodate potential time restrictions. A task might for example take up more time than expected and the participant cannot extend the time sacrificed for the session, or it might be clear from the start that a participant cannot spend the time normally required to complete all tasks. The evaluator should design tasks bearing these situations in mind and be able to suppress a task rather than requiring the user to complete tasks faster. It is primarily the depth of data which contributes the most valuable information to the analysis of issues related to searching behaviour, not the fact that all tasks have been completed.

12.3 Findings

Altogether, four studies were conducted for the work of this thesis. The first two were conducted with the participation of students and staff of the University of Glamorgan using proprietary interfaces on the Internet. These studies resulted primarily in the development of the methodology and data analysis techniques, as discussed in the section above. Due to their preliminary nature, data analysis was limited. However, some results, for example the realisation that participants did not necessarily make a distinction between the thesaurus component and the remaining system, led to the refinement of the research questions, which moved from specifically investigating interactions with the thesaurus to analysing its impact on and contribution to the search process overall. The latter two studies were conducted with the FACET system designed at the University, in order to assess methods of thesaurus retrieval. The first FACET study resulted in significant changes to the interface design, for example re-structuring the thesaurus access and query building facilities to reflect the search process better, redesign of the list of queries and display of records in order to make comparison of records and queries easier, and changes in thesaurus access functionality (chapter 7). Findings from the second study, particularly the possibility to change query expansion for individual terms and the use of option buttons rather than sliders to do this fed into a web-based version[1] used to demonstrate possibilities for thesaurus integration (see also Binding and Tudhope 2004). The participants in these studies were library and museum professionals, who corresponded to the anticipated user group for a system of this degree of complexity. The author set out in the initial studies to investigate thesaurus interaction in particular. As a result of the initial experience, the later focus became the role of thesauri in information searching in order to assess how users approach these tools, how they inform their searching and how they could be used to inform searching further. Related aspects investigated were the kind of problems participants encountered and how these could potentially be resolved. These findings are discussed in detail in the chapters describing the studies and in the model developed in chapters 10 and 11. Some of these include for example the fact that the users' activities potentially need to be channelled, for example browsing from the top-level terms is usually not very efficient, but browsing can be very beneficial in a local context. This section provides a summary of the more significant

[1] http://www.comp.glam.ac.uk/~FACET/webdemo/

conclusions. At all stages of information searching, problematic situations, connected to lack of support of the search process, could occur and this is discussed in section 12.3.3.

12.3.1 Thesaurus use

Overall, thesaurus interaction did not pose problems to the participants, although in the FACET studies, where the thesaurus plays a more prominent role in the interface and where data collection was more detailed, conceptual difficulties such as misunderstandings of what happened while browsing the hierarchies (e.g. believing that this process filtered records, section 6.7.4) and the purpose of guide terms, were identified in particular with those participants who were less familiar with thesauri. However, the participants appreciated the possibility to browse the thesaurus.

In chapter 2, advantages of thesauri have been discussed. Amongst these are for example the fact that false hits can be avoided to a large extent, the possibility of discovering the most suitable query terms by navigating the thesaurus and terminology control through use of thesaurus terms rather than free text terms. During the FACET 2 study in particular, the latter two effects of thesaurus use could be observed.

However, access to thesaurus terms needs to be improved for users to benefit more fully from them. Difficulties in initially selecting thesaurus terms occurred repeatedly, for example in making a choice amongst several options. All interfaces investigated provided a more or less sophisticated mapping function to candidate terms. However, some users misunderstood the meaning of retrieved candidate terms or had trouble to enter the right combination of words in order to retrieve appropriate candidate terms. Incidents observed in the FACET 1 study show that browsing the thesaurus on a global scale is not an alternative solution to access terms as identifying the relevant categories can be very difficult for users (see e.g. 6.7.7), and it is thus reasonable to conclude that any (inherent) limitations of mapping approaches can have a detrimental effect on term selection and thus query construction. However, it was also found that as long as searchers were able to identify some query terms, further ones could be extracted from records retrieved, and thus provide additional options for query reformulation.

As discussed for example in sections 6.3.2 and 6.7.8, term expansion is another method to compensate for a sub-optimal selection of query terms. Users need not choose a specific Narrower Term or include large numbers of terms in their queries. Expansion was investigated using the FACET interface, and was successfully used as intended. The mechanism was implemented in the FACET 2 interface so that users could modify the expansion rate themselves. This caused some problems, partly due to the complexity of the calculations which make the effects difficult to explain and predict. Thus searchers felt unsure about using it, but by the end of their sessions, some expressed that their understanding had grown through the experience. The implementation resulted in some usability issues, for example setting the expansion to a precise value requires much dexterity, which need to be addressed, but overall, this mechanism can be used by those users who are not professional searchers. Extended hands-on training and highlighting of changes in query results that might trigger more conscious deliberation on the effects of query reformulation appear to be ways in which users can be sensitised to the effects of the expansion and learn to apply it in a directed manner.

User behaviour could also be channelled in this aspect. During the FACET 2 studies, it was found that a useful strategy was to set the default for expansion to "maximum" for the first query executed, which should contain not too specific terms. This tended to retrieve a broad set of results, which allowed users to gain an overview of the type of records in the collection. They could then refine their query terms using indexing terms from the records, and if required, reduce the query expansion to narrow the results set. This approaches takes full advantage of the expansion and allows more experienced users to modify to expansion parameters as required, but also reduces the cognitive burden for less experienced users.

In terms of the implementation of the expansion mechanism, further investigations might be required. Query terms can play different roles in the query, for example the focus term identified during the study of the FACET 1 interface. Other less distinct types of terms, which would allow fine-tuning of queries on a specialist level by distinguishing between similar records, seem to exist (9.7.3).

12.3.2 Tools based on controlled vocabularies

Apart from the expansion mechanism discussed in the previous section, other tools could be developed based on controlled vocabularies. A mechanism to identify additional concepts based on record indexing which are clustered using the thesaurus has been mentioned in section 11.3.2 - Extracting information from data in the system. The additional options suggested here would have been useful in situations encountered in the studies, but the tools need to be integrated into the system while considering the domain.

Distance between controlled vocabulary terms

When expanding terms, it can be useful to calculate the distance between two terms in the controlled vocabulary in order to see if specific terms of interest to the searcher are included under the current settings. The system could, for example, identify lexically similar terms and warn searchers if the term expansion is not large enough to include both in the query. This approach could by-pass missing associative relationships in the controlled vocabulary. In the AAT, for example, "imitation leather" and "leather" are far apart but searchers assumed that the maximum expansion would include both terms in the query, as the two materials did not seem so different to them as they are considered in the AAT (section 9.8.4). Lexically close terms are not necessarily semantically close though or in any way relevant, so that establishing under which circumstances such warnings are issued is vital in order to avoid too many confirmation requests or warnings. Another issue is that expanding terms too far is likely to result in the inclusion of many irrelevant terms. In the case of "leather" and "imitation leather" for example, it could be that effectively all material terms have to be included in the query, and the effect of this on the results remains to be established. Rather than modifying the expansion (and potentially even the term included in the query – here, "<animal materials>" would have to be used instead of "leather" in order to include both terms in the query), a "Hedge" tool might be more appropriate. Hedges, i.e. the combination of terms through Boolean "OR", would significantly increase the complexity of the retrieval mechanism. However, it might be possible in the future to integrate semantic expansion with the Boolean operators.

Identifying additional concepts from specific categories

At the start of a search on a new topic, it is unknown to the searcher which terms will result in hits. When searchers have selected a main concept they are interested in and are aware that their query should be more specific, they do not always know which other concepts to use, especially when they have already executed a "0 hits" query. A tool can be imagined that allows searchers to investigate combinations of concepts which co-exist in records by entering a controlled vocabulary term describing their main concept, for example "chairs". They then select a category or facet, for example "materials". The system searches the database for records containing the main term and returns all indexing terms from this category, or alternatively, all indexing terms, which could again be clustered into concepts using the thesaurus. The Flamenco system uses a similar approach where indexing terms are displayed alongside the results for searchers to refine their queries.

12.3.3 Information searching

The more detailed data collected in the studies with the FACET system allowed an investigation of problems encountered at different search stages. Potential reasons and solutions have been discussed in the model documentation on Risk Points. Generally speaking, the difficulties observed, which primarily occurred on the conceptual level, show that the search process in itself needs to be supported by the interface. The initial version (1.0) of the FACET interface allowed users to access the windows which represented different stages of the search process in any order. Redesign in the following two versions (1.1 and 2) resulted in a more structured interface, which provided the necessary access to functionalities at the search stages when it was required, which largely reduced confusion on the next steps to take, although this issue has not yet been completely resolved. Specifically at the query reformulation stage, further support is required. Superficially, this search stage resembles query construction, but searchers have to take into account the results they have retrieved and the extent to which any of them correspond to their expectations has an impact on their options. This stage is thus more complex than the initial query construction. If results generally correspond, query terms can be extracted from the records or refined based on the indexing terms found. However, even these options are not necessarily obvious to the searchers. Thus, interface should support reformulation more specifically in that the available options are presented to searchers, potentially in combination with more general instructions, such as inspecting records for indications of query terms which resulted in retrieval of unsuitable records.

Two specific observations regarding search behaviour were made during the FACET studies. The first one is that two distinct approaches to information searching were applied. The second was the fact that participants appeared to transfer knowledge on Boolean searching to the FACET system which used ranked retrieval.

For the first search approach, searchers put more effort into query formulation than in the other approach where they preferred to browse records in order to identify suitable results. Factors which influence search behaviour towards one or the other approach are for example the time available to complete the task or habitual search approaches, which are in turn related to different motivations behind the searches.

Attempts to transfer knowledge about Boolean systems were identified in several sessions. It was unclear how exactly this affected search sessions, but it appears that for example the number of query terms selected and the wish to exclude terms from results (i.e. using Boolean NOT) are an expression of this. A concern with 100% matches was another potentially related observation.

12.4 Original contribution to knowledge

The main contributions to knowledge of this thesis concern the information searching process in controlled vocabulary-enhanced system. Firstly, the necessity of supporting the interface searching process on a conceptual level within the interface, and particularly during query formulation, has been shown, with various issues discussed. Related findings have been applied in the FACET 2 interface. Secondly, aspects of making better use of the thesaurus and integrating it into the interface have been identified. Accessing thesaurus terms is important for users to efficiently construct their queries, and query expansion is one way of counteracting some of the negative influences that the users' choices can have. It has been shown how semantic expansion, a complex processes in itself, can be used to the benefit of users without imposing too much cognitive load on those who are not professional or well-trained searchers. It has also been found that thesauri could form the basis of more tools or functionality which can support searchers, for example by extracting new concepts from retrieved records or by making the analysis of results easier.

The findings have been brought together in the model of information searching in controlled vocabulary enhanced systems discussed in chapter 10 and 11, which forms a further contribution to knowledge. This model consists of three parts. The descriptive part (graphical and textual representation) presents the stages of information searching with their corresponding interactions:

- Identifying keywords and mapping them to thesaurus terms
- Constructing or setting up the query and executing it
- Considering and selecting results
- Reformulating the query (this step has significant parallels to the first two)

This descriptive part is complemented by a set of so-called Risk Points. These identify for most stages and decision points of the model potential difficulties and makes suggestions towards their solution. A third part of the model is its tabular representation, which can be used in the evaluation of existing interfaces or those under development.

This model does not intend to predict user behaviour, but to assist interface and system designers in making decisions regarding new systems by providing them with a description of the information searching process which draws attention to session-related factors which potentially have an influence at different search stages and can lead to problematic situations. A number of support mechanisms ranging from simple displays of information to more complex, algorithm-based functionalities are suggested. These are meant to be exemplary rather than prescriptive. Risk Point [4].1 for example deals with a situation where the searchers consider the retrieved results unsuitable. An example of a searcher who was stuck with the same three records even after several reasonable query modifications has been described in section 6.7.9. This kind of situation can be difficult to resolve because first of all, searchers are unaware of why inappropriate results were retrieved. Under certain circumstances, query reformulation can be difficult, for example because the searchers need to replace the terms they have used but are not aware of alternative query terms. Automatic query modification, for example further expansion of the query terms or removal of the terms which are least efficient in retrieval might lead to a change in records retrieved. As this approach is not always successful, the interface could provide some guidance on how to identify causes for the unsatisfactory retrieval. Examples include inspecting the records to see how a term is used in indexing, or to verify its meaning and scope, so that terms which need to be replaced by others which are not semantically close can be identified. In this case, it is important that searchers have the possibility to easily inspect records and thesaurus terms.

Most importantly, the model presents the search process onto which a controlled vocabulary information searching system should be based in order to allow a natural flow of interactions through the stages of the search under particular consideration of the difficulties that can arise at the query reformulation stage.

12.5 Future work

The work carried out for this thesis raised questions concerning a number of research areas. These can be grouped into five categories: Methodological development, issues related specifically to information searching in the domain of museums, the further development of the model, the application of the model, and the development of tools to support information searching.

12.5.1 Methodological development

The improvement of the methodology has been an important aspect of this research. Aspects of the methodology which would benefit particularly from further development are the data analysis techniques. In the studies reported here, collation of data proved to be useful in order to construct a rich picture of the search sessions. However, sometimes it was awkward to work with these data, for example when examining solely the evaluator's interventions, which relied to start with on the audio transcript. Thus, while generally speaking, it was useful to have data fully collated, a tool should be developed which makes it also possibly to display only one data type. Although software exists for example to annotate video files, this tool could be specific to information searching and permit easy access to important values, for example query terms submitted. It could allow annotation of interactions, for example to attribute moves and tactics, and might then even do some of the descriptive statistical analysis used particularly in chapters 8 and 9 to illustrate sessions, permit definition of sequences of interactions as "patterns" and include a feature to summarise the sessions at different levels of abstraction in the form of graphs, or in the form of a textual summary using moves, tactics or "patterns".

When using prototypes for information searching behaviour studies, the realism of Information Need, motivation to complete tasks, but also the approaches taken by participants can cause problems for collecting relevant data. During this research, the author developed a type of set tasks, called "evolving scenarios". This method was intended to help observe situations relevant to the research interests, for example interaction with the thesaurus and reformulation. So far this method could not yet be refined to a satisfactory level. Although the idea to let an information searching task evolve according to the approaches taken by participants was useful, the practical execution was difficult to an extent that the technique was used little in the study where it was intended to be used (FACET 2, see section 12.2.3). One of the main problems was to be prepared for all possible approaches. Based on the experience of these studies, and possibly also using information from the model in chapters 10 and 11, a framework for constructing evolving scenarios could be constructed, which would help anticipate common situations and standardise input from the evaluator. Use of evolving scenarios should then specifically be tested in user studies.

12.5.2 Information searching in a museum's context

Although it was not appropriate to pursue this analysis in the context of these studies, some issues specific to information searching in the museum's context were encountered. It was found that many of the curators' habitual searching tasks do not require them to construct very specific queries (8.7.4). Bates, Wilde and Siegfried (see Getty project reports, in particular Bates et al. 1993) have already pointed out particularities of humanities scholars' information searching behaviour compared to that of science researchers. Although museum professionals might at first sight be considered as part of the first category, there appear to be additional differences between what appears to be their information searching activities and needs and those described in the Getty project. These differences probably relate to the different tasks, as museum professionals perform for example many searches for others, but might also be related to the fact that here, the collection in question contained objects as opposed to documents. Further studies comparing museum professionals information searching behaviour with that of other humanities-related professionals might thus provide further information on how access to museums or more generally, cultural heritage data should be provided. Considering the museums' efforts to make their collections available on the Internet, this also raises questions concerning users who are not cultural heritage professionals. These might potentially approach collections with different aims and objectives, and consequently, their approach would differ from that of the professionals. Particularly for these users, alternatives to standard queries seem appropriate which could be based on browsing interactions and the "More like this" principle.

12.5.3 Development of tools to support information searching

In the Risk Points documentation, a number of ways in which information searching can be supported have been addressed. Tools of specific relevance to this research are access mechanisms to thesaurus terms and tools based on controlled vocabularies, using for example expansion. As mentioned above, identifying specific query terms was problematic in the studies despite the mapping functionality implemented in all the interfaces. These mechanisms could for example benefit from natural language processing which allows a more complex assessment of the expression entered by the users, and in turn retrieve controlled terms which better represent the required concepts. An example is the disambiguation of homographs. When for example the term "table" is entered, an analysis of the remaining terms could be conducted. If "graphs" is amongst the other terms, the system could assume that "tables (document components)" is probably the required thesaurus term, while the presence of "cast iron" would indicated that "tables (support furniture)" is more relevant.

In combination with records from the collection, controlled vocabularies and thesauri in particular could be used to assist searchers in discovering new concepts, which previously has been one role of search intermediaries, which cannot be fulfilled by the thesaurus on its own, unless the mapping mechanisms are improved significantly.

A further type of tool which could support searching are tree structures, from which the interface guides searchers to appropriate instructions and suggestions for query reformulation or search progression. In order to develop these decision trees, it is first necessary to identify potentially problematic situations and possible strategies to resolve these. In the next step, the problematic situations need to be categorised in order to allow selections which will progressively reduce the number of situations and solutions which apply. Possible solutions will probably require the identification of appropriate behaviour, such as browsing the thesaurus or inspecting records to identify potential query terms. As seen here, behaviour should be channelled, for example to dissuade navigation of the thesaurus from the root term or top-level categories, in order to avoid further problems.

In terms of expansion, the suggested methods for using it should be further assessed through experiments (i.e. conducting searches using different parameters and comparing the options to progress in terms of types of items retrieved, indexing terms available in records etc.) and user studies. The implementation of this mechanism could be simplified on the interface level in order to make it easier to use, or it could possibly even run automatically in the background.

Although the collections used in these studies were indexed with the controlled vocabularies in question, thesauri can also be used for searching only, in which case Alternate Terms might also need to be included in query expansion. Recently use of the underlying FACET components for a web interface has shown that even the more computationally-expensive functions can be used over the Internet, and not only locally in a stand-alone system. Research can thus also be expanded to this domain.

12.5.4 Further development of the model

The model of information searching in controlled vocabulary enhanced systems proposed in chapter 10 and 11 is not complete. The section on query reformulation for example is only preliminary and requires specific research on this behaviour.

Further development of the model by including more factors is also possible. It would be necessary to decide whether this extension is to be theoretical or practical. From a theoretical perspective, other researchers' work could be integrated. Spink's work on successive searching (e.g. Spink et al. 2002) would here be of particular interest. Influences such as previously viewed documents or records and identified query terms or concepts become potentially more significant and instead of being tacitly included in the contexts of the query, they might be made explicit in the graphical presentation of the model, the tabular view and dealt with specifically in the Risk Points. Kuhlthau's work on affective factors (e.g. Kuhlthau 1988, 1991) for example mentions in addition to negative feelings, such as frustration, positive ones such as clarity or confidence.

From a practical perspective, the model could potentially be adapted for use with user profiles, for example to collect data on the number of query runs executed. This could possibly become the basis for personalised support mechanisms which efficiently avoid or overcome problematic situations in information searching.

12.5.5 Application of the model

So far, the model has been tested against user data and existing interfaces. It is meant to assist evaluators and developers in their work, thus the next logical step is to design or redesign an information searching interface using the model. Considering that the suggestions for solutions are exemplary, this does not require implementing all of them, but selecting the most suitable ones for the specific context. This will serve two purposes, first of all to assess the difficulties that might arise in using the model for the design of an interface, and secondly to assess the correctness of the model itself, especially with respect to the finding that implementing the interface according to the stages of the search process will facilitate information searching.

In order to render the model more usable for interface design, it could be presented in an interactive form which would allow a better logical connection between search stages and factors, risks or potential solutions than the linear presentations of the Risk Points and the search process.

12.6 Final thoughts

This research set out to investigate possibilities of using thesauri to make it easier for end-user searchers to find relevant information. It was found that using standard methods of interaction (e.g. selection using mouse clicks and drag-and-drop), end-users who have a certain knowledge of thesauri can deal with one, even in a complex context with expansion and a facetted thesaurus structure. A more significant problem than the use of thesauri was the support of the whole search process. The stages of the search – identification of concepts and selection of query terms, construction of the query, analysis of results and reformulation in particular – should thus be supported within the interface. Semantic expansion opens up possibilities for use of the domain knowledge, inherent particularly in the hierarchical structure of thesauri, in query construction. Searchers are not required to have the knowledge to select highly specific terms and can initially run general queries with a large expansion rate, which permits them to gain an overview of the possible results. However, in order to use expansion functionality in a meaningful, controlled way, users will have to gain practical experience of how it affects results.

Additionally, thesauri could be employed in the background in order to assist with identification of concepts and results analysis.

It is hoped that this research will contribute to the understanding of the controlled vocabulary search process and the development of future tools based on thesauri that will enable end-users, particularly those who use a system on a regular basis, to search more efficiently for information.

A qualitative study of thesaurus integration for end-user searching

- Appendices -

Dorothee Blocks

A submission presented in partial fulfilment of the requirements of the University of Glamorgan/Prifysgol Morgannwg for the degree of Doctor of Philosophy

July 2004

Table of content for appendices

Glossary of terms and acronyms

AAT	Art and Architecture Thesaurus, developed by the Getty Institute.
ASK	Anormal state of knowledge, Belkin 1980 (section 4.2).
BT	Broader Term, used in context of hierarchical controlled vocabularies.
Camtasia	Screen capture software.
CHIN	Canadian Heritage Network, developed the Artefacts Canada site discussed in 2.4.3.
Candidate term	In the FACET interface and the model (chapter 10 and 11), a controlled term which is potentially appropriate for the searchers' query.
Compound term	Term constructed from single concept terms, for example "upholstered seat".
Controlled term	A term from a controlled vocabulary.
Drilling down a facet	Browsing a thesaurus along NT relationships.
ERIC	Educational Resources Information Center, maintains an online database discussed in connection with a preliminary study in section 5.3.1.
Expansion (of term/query)	See semantic closeness expansion.
Explode	An option in the Ovid interface to include NTs of the selected query term (see 5.4.1.3.).
Facet	mutually-exclusive fundamental categories into which terms from a controlled vocabulary are arranged.
FACET	Faceted Access to Cultural hEritage Terminology, an experimental system developed at the University of Glamorgan. "FACET 1" and "FACET 2" refer to two versions and the studies carried out with these.
False hit	Record which is retrieved because query terms occur in a different context than that intended by the searcher.
Free text	Phrases that do not stem from a controlled vocabulary. Used also short for "free text retrieval", an alternative retrieval technique. See also "NLP" and section 2.1.
Flamenco	IR system for faceted image retrieval developed at the University of Berkley (see section 2.4.4).
Focus	A query term option in the Ovid interface which permits users to define the primary topic of retrieved items (see 5.4.1.3.).
Focus term	In FACET 2, searchers can select a term which is expanded differently from the others to focus the query on this concept, see 7.5.
Hierarchy	Although not mutually-exclusive by definition, this term is similar to "facet" and refers to the thesaurus or a section there-of. In the context of this thesis, "objects hierarchy" for example refers to the terms found below "Objects".

IN	Information Need
Information searching	Process of looking for information in an online environment.
Information seeking	Process of looking for information using potentially a number of different sources and different media types.
IR	Information retrieval
LCSH	Library of Congress Subject headings
LIS	Library Information Science
Mediated searching	End-users acquiring information through a professional searcher (also called intermediary).
MeSH	Medical Subject Headings, used for example by the Ovid interface to Medline (refer to study in chapter 5.4.1.)
Move	Change in the query formulation (Fidel 1985).
MultiMimsy	Collections administration software used at NMSI museums
NMSI	National Museum of Science and Industry
NKOS	Networked Knowledge Organisation Systems. A registry for NKOS is referred to in chapter 2.4.
NLP	Natural Language Processing, an alternative information retrieval technique. See also "free text" and section 2.1.
NRM	National Railway Museum, York, UK, forms part of NMSI.
NT	Narrower Term, used in context of hierarchical controlled vocabularies.
OKAPI	Experimental OPAC system developed at City University, London. Discussed in chapter 2.4.2.
OPAC	Online Public Access Catalogue.
Ovid	Proprietary interface to a number of online databases such as Medline and INSPEC. Discussed in chapter 5.3.
Posting	Number of records associated with a specific controlled vocabulary or query term.
Post-coordination	Controlled terms are combined to form concepts at the time of indexing. Refer to chapter 2.3.3 and "pre-coordination".
Precision	Ratio of relevant items retrieved to total items retrieved.
Pre-coordination	Controlled terms are constructed from several concepts so that fewer terms are needed for indexing. Refer to chapter 2.3.3 and "post-coordination".
Query Builder	A part of the FACET 2 interface used to construct queries from thesaurus terms, see section 7.2.
Recall	Ratio of retrieved relevant items to relevant items in the database.
Risk Points	Part of the model documentation described in chapter 11 which specifies risks, causes and potential solutions for different stages of the search process.
Root	The highest or top level of a hierarchical controlled vocabulary.
RT	Related term, used in context of controlled vocabularies.

Search approach	General progress or process of a search..
Search stage	Distinct phase of the information searching process, for example.
Search tactic	An action in information searching to further a search (Bates 1979).
Search strategy	A plan for (part of) a search session (Bates 1979).
Semantic (closeness) expansion	In a hierarchical vocabulary, the hierarchical relationships can be used to "expand" terms semantically, i.e. to include those nearby in a query. See section 7.3.2.
Scope note	In a controlled vocabulary, an explanation of how the term is to be used.
SOCS	School of Care Science. See section 5.4 for the SOCS study.
Screen capture	Recording of activities visible on the screen. See e.g. section 4.4.3.
Term Finder	Part of the FACET system which retrieves candidate controlled terms from the thesaurus, see section 7.3.1.
Terminology control	Use of a controlled vocabulary for indexing.
Thesaurus Browser	In the FACET system, the tool which gives searchers access to the thesaurus.
Traversal	Moving from one controlled term via a relationship (e.g. BT, NT, RT) to another term.
TREC	Text REtrieval Conference, http://trec.nist.gov.
UF	Use for, i.e. refers the user to a synonym. Used in context of controlled vocabularies.
Zero hits/0 hits	Another way of saying that a query has not retrieved any records.

URLs

AAT: http://www.getty.edu/research/conducting_research/vocabularies/aat

ADAM: http://www.adam.ac.uk/sindex.html - select "multi option browser" from the list

BHInet (obsolete): http://www.bhinet.co.uk/content/free/welcome.asp

BUBL LINK: http://www.bubl.ac.uk/link/

Camtasia website: http://www.camtasia.com/products/camtasia/camtasia.asp

CHIN: http://www.chin.gc.ca/English/Artefacts_Canada/index.html

ERIC: http://searcheric.org/

ERIC wizard: http://searcheric.org/scripts/ewiz/amain5.asp

Extensible Markup Language (XML): http://www.w3.org/XML/

FACET project : http://www.glam.ac.uk/soc/research/hypermedia/facet_proj/index.php

FACET web demo : http://www.comp.glam.ac.uk/~FACET/webdemo/

Flamenco: http://bailando.sims.berkeley.edu/flamenco.html

Gateways listing, University of Glamorgan:

http://www.glam.ac.uk/findit/elib/internet/gateways.php

Iconclass: http://www.iconclass.nl/

ITIS (obsolete): http://habanero.nhm.ukans.edu/aves/default.htm

National Museum of Science and Industry picture gallery:

http://www.nmsi.ac.uk/piclib/index.asp

Netherlands Institute for Scientific Information Services (NIWI) (registered users only):

http://www.niwi.knaw.nl/nl/SWL/intro.htm

NKOS thesaurus registry site: http://nkos.slis.kent.edu/Thesaurus_Registry.html

OKAPI Pack: http://www.soi.city.ac.uk/~andym/OKAPI-PACK/index.html

OVID for Inspec (registered users only): http://inspec.edina.ac.uk/

OVID for Medline (registered users only): http://gateway2.uk.ovid.com/

Plumbdesign's visual thesaurus: http://www.visualthesaurus.com/online/index.html

Resource Description Framework (RDF): http://www.w3.org/RDF/

Resource Discovery Network (RDN): http://www.rdn.ac.uk

Revealing things: http://www.si.edu/revealingthings

TREC: http://trec.nist.gov

Waterways thesaurus information: http://www.mda.org.uk/waterw/index.htm

Appendix 2.1 Example for interface scheme

Thesaurus and environmental attributes - CHIN:

Some of this information is based on the Networked Knowledge Organisation Systems (NKOS) Registry help document[1]. Those fields will not be explained in more detail. Additional fields are dealing mainly with user-thesaurus interaction. Fields from the NKOS registry will be marked with an [N], with NKOS registry required fields marked with an *.

i. Product information (thesaurus)

This section provides basic information on the thesaurus, on who created and maintains it and where on the Web it is available.

Product name/title [N*]: CHIN Artifacts (Humanities) [integrating the AAT]
Variant Product Name/Title [N]:
Type of product [N*]:
Product description [N*]:
Auxiliary lists [N]:
Current edition/version [N*]:
Date of current version [N*]:
Product update frequency [N*]:
Author/Editor [N]:
Technical development:
Content provider: CHIN (Canadian Heritage information network)
Online availability [N]: yes
URL: http://www.chin.gc.ca/Artefacts/e_artefacts_canada.html
URL for examples [N]:
Examples of records or other interface information in case the thesaurus application is not freely available.
Related publications:
Publications regarding the product, its application or use.
Notes [N]:

ii. Application

This section refers to the application in which the thesaurus is integrated or used with.

Name of the application: Canadian Heritage Information Network
Version: ?
Last update: 1998
Major Subject coverage [N*]: information from museum and galleries collections on e.g. archaeology, decorative arts, fine arts, ethnology, and history
Minor subject coverage [N]: N/A
Type: database
Database integration: AAT integrated into search interface:
Databases/resources that the thesaurus applies to: various
Implementation: HTML/CGI
Password/authorization required: no
Time out: no

[1] http://nkos.slis.kent.edu/Thesaurus_Registry.html

iii. Availability for use

Online browser- location: http://daryl.chin.gc.ca/Artefacts/e_MasterLayout.cgi
Downloadable: no
Availability on last access: yes
Printed/paper versions: ...
Available format(s) and size ^{N*}: ...

iv. Terms and conditions

Purchase price by format (or cost-free statement) ^{N*}:
Subscription price by format ^{N}:
Licensing Availability $^{N:}$
Restrictions (or non-restrictions statement) ^{N}:

v. Vendor

Vendor Name ^{N*}:
Vendor Street/Post Office Box ^{N*}:
Vendor City ^{N*}:
Vendor State/Province ^{N*}:
Vendor Country ^{N*}:
Vendor Postal Code/ZIP Code ^{N*}:
Vendor Voice Phone ^{N*}:
Vendor TDD/TTY Phone ^{N}:
Vendor Fax ^{N}:
Vendor Email ^{N}:
Vendor Logo URL ^{N}:
Vendor Web Site URL ^{N}:
Vendor Hours of Service and Timezone ^{N*}:
Vendor Service Description ^{N*}:

vi. Contact

Contact Name ^{N}:
Contact Voice Phone ^{N}:
Contact Fax ^{N}:
Contact Email ^{N}:
Contact Web Site URL ^{N}:
More Contact Information ^{N}:

vii. User characteristics – who is this system designed for?

This section deals with attributes related to targeted users of the application. Most of this information will only be based on the evaluators' estimation and these attributes mainly distinguish between the level of training or experience that might be required to use the applications.
Used by (user community and applications) ^{N}: anybody
Level of expertise: probably less experience with searching, there is quite a lot of support available in the interface
Frequency of use: various
Subject knowledge: not required
Purpose of information use: finding pointers to more information, study and personal interest, possibly research

viii. Tools and types of assistance

Accessing support/assistance: help link in the header; instructions in the interface
Aims of help: make users aware of different options for searching and explain how to use functions.
Indexing/Searching of documents: searching; documents have been indexed with the thesaurus
User modeling/profiles: N/A
Different approaches: search by concept, command or using the AAT. Searching different fields – who, when, where, how, why, what. The users can also search on their own or browse through sample queries.
Statistical analysis (e.g. postings): yes, by field (who, where, what...)

Relevance feedback: no
Search strategy suggestions: not strategies as such, but suggestions of different approaches and examples to look at.
IR-specific error recovery: suggestions to deselect "Search for files with images on" when not many results are found, but no suggestions for very large results sets.
Help files: instructions in the interface, help file accessible from header of page which includes more instructions on how to use different interfaces.
Is the help file required already to complete basic tasks or only for more exceptional tasks?: no.
Available modes: different types of queries have different "displays"
Wizards: no, but possibility to browse existing queries
Bookmarks/histories: N/A

Other potentially useful functionalities

Cut and paste: N/A
Saving records/data/terms or queries: no, results can be treated as web sites though
Add value: N/A
Post-processing of data/information: results can be treated as web sites

ix. General interface attributes

Description: There are 4 frames when the user comes to the searching interface. The header remains constant throughout the search. The help file can be accessed from here. On the left, there are some links to more information. On the right, there is the frame from where the users search alone and the one where they can look at pre-programmed queries.
Changes of display: The thesaurus, if this search type is selected, replaces the links to information in the frame on the left. Thesaurus information replaces the frames on the right. When the results are displayed, only one large frame remains underneath the header.
Language(s) [N*]: French, English
Metaphors: N/A
Dialog techniques: form fill-in, drop down menus, selecting hyperlinks
Alphanumeric presentations: Terms as hyperlinks
Graphical representations: Arrows for the different fields in different colours, but they also have their meaning in text.
Distractions: There is quite a lot of information fitted into the one window. Within one frame, the users are probably okay, especially as displays of other frames "disappear" at different points of the search.
Knowledge integrated into the interface and integration/access of this knowledge: N/A

x. Networked Knowledge Organisation Systems characteristics

Language(s) [N*]: English, French
Types of terms [N*]: concepts
Description of overall structure [N*]: Hierarchical thesaurus terms are displayed as hyperlinks and when clicked on, a search with this term is executed.
Source of new terminology [N*]: AAT
Number of preferred terms or nodes [N*]: ...
Number of non-preferred terms [N]:...
Types of relationships [N*]: BT/NT
Depth of hierarchy (maximum number of levels) [N]: N/A

xi. Presentation of the thesaurus

This section refers to the structure and display technology of the thesaurus.

Structure: Hierarchical tree display
Technology: Hypertext
Display: 2D
Standards applied in thesaurus/structure development: N/A
Standards applied for display technology: N/A
Presentation of terms: Text/hyperlinks

xii. Presentation of documents by the interface

This section refers to the conceptual display (Marchionini 1995) of documents or information in the system. The display of documents can help users clarify their search question in their mind by identifying relevant and irrelevant aspects.

Document structure: Object Name; Manufacturer; Latest Production Date; Material; Accession Number; Institution; (example, would vary for different types of records.)
Alternative levels of detail: short record or more detailed record. Records can also images.
Indexing terms: Category and subcategory are displayed, but are not navigable.

xiii. Query formulation in the interface

Query styles: form fill-in, query by selection, command language

Search features: Boolean operators (masked: All/any terms), search for records with images
Free text terms: yes, users need to refine query after running it by clicking onto a term in the AAT display.

xiv. Thesaurus activation

Thesaurus activation: user selects this search mode. The thesaurus can actually only be used separate from different queries, unless the users modify a query including a thesaurus term.
Point of activation: from main interface
Thesaurus entry points: only by selecting this search mode. It is not even possible to add more thesaurus terms!

xv. Thesaurus interaction/navigation

Interaction supported: searching of documents/records
Object management: N/A
Order of action: users decide on whether they want to execute a query by themselves or not, and then decide on the query type. After results are displayed, they can either modify the query or execute a new one.
Mode of manipulation: mouse, keyboard entry
Browsing supported: flat browsing of the thesaurus after entering into categories by clicking on hyperlink titles. Apparently it should be possible to search the thesaurus, but hits does not seem to be possible or does not work.
Jumps in list/Scrolls/Go-tos: jump in list for index
Interaction styles: selection of hyperlinks

xvi. Other (overall, quite subjective judgements)

Ease of use:
Easy to learn:
General impression:

Appendix 5.1 ERIC study – Handout/Questionnaire

Thesaurus workshop

Objectives

Explore the use of a thesaurus for searching a database in comparison to free text search. The database searched is "ERIC", which contains educational references related to information technology. There are two interfaces that we will use, on with and one without thesaurus access.

Scenario

In work, you are involved in a project group looking at developing multimedia materials for distance learning. These are targeted at adults who wish to acquire new knowledge and skills for a wide range of reasons. You and your colleagues are now interested in making some learning material available on the web. Without considering issues like regulating access and charging users, imaging that it is your task to write a report to discuss with your group and project management. In this tutorial, you are to create a bibliography of about 5 references, which seem to provide a good coverage of the issues involved.

If you wish to, you can execute your own search. Please explain what you are looking for on the back of this sheet.

Exercise A

Build a subject search for this scenario or your own information need by thinking of the concepts implied in the scenario and formulate a search statement. The query interface is at: http://www.ericae.net/aesearch.htm

Note:
OR: (termA,termB) (Don't leave spaces.)
AND: Simply type terms in separated by spaces
Put terms which contain spaces between "help line"

Write the statement down here:

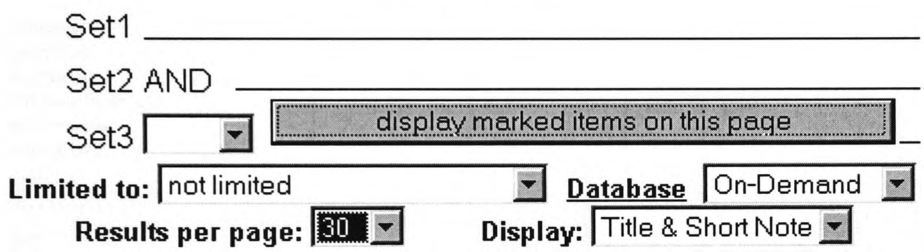
Change the "result per page" to 20 and execute your query. Have a quick look through the results displayed. You can click on the title to read the abstracts.

Exercise B

Now, execute the same search using a thesaurus. When you first open the site specified below, you will be requested to enter your first term. So look back at components you identified earlier. If your term is not a thesaurus term, the wizard will find one, which is spelled similar. Try and find descriptors (thesaurus terms) for all your concepts.
The address is: http://www.ericae.net/scripts/ewiz/amain5.asp

Note:
Terms in the same set are ORed.
All sets are ANDed, unless you choose to exclude the third set.
If there are not enough sets, execute your query and add more concepts in the textbox on the next screen as you would have done in the previous search.

Please enter your query here as you did on screen.



Now refer back to the scenario and select 5 references by checking the boxes on the left. At the bottom of this page, you will find this button:

Click and save the new page in the following location:
If you want your results to be printed for you: Students on F:\ERIC\print.
If you don't want a printout of your results: Students on F:\ERIC\results.
Please call your file "«Experiment_number».txt". If you want a printout, keep the bottom part of this page for information.

Please return this handout now. (This is a condition for having your results printed.☺)

Your experiment file was called «Experiment_number».txt. You will be able to collect printouts from tomorrow (6.4.00) at the lated from outside JJ 7 (If there's still a photocopier in the corridor, they will probably be on it.).

Appendix 5.2 ERIC study – Query terms and records retrieved

Bold= thesaurus terms *if several searches were executed for any type, there were drawn together. Shaded terms, e.g. "development" are terms from the scenario.	Beta	Beta (thesaurus)	records retrieved	Gamma	Gamma(thesaurus)	Delta	Delta(thesaurus)	records retrieved	Epsilon	Epsilon(thesaurus)	records retrieved	Eta	Eta(thesaurus)	records retrieved	Iota	Iota (thesaurus)	records retrieved	Kappa	Kappa(thesaurus)	records retrieved	Theta	Theta(thesaurus)	records retrieved
adult education																						1	2
computer assisted instruction																			1	1			
computer based learning																		1					
computer mediated communication		1	5																				
continuing education										1	1												
Delivery																					1		
Development	1			1		1	1																
distance education											1					1	4		1	6			
distance learning	1			1	1	1	1		1	1		1	1		1			1					
education environment																						1	5
educational media							1	1															
home study									1	1													
Hypermedia													1	5									
independent study										1	3												
learning																					1	1	
learning materials									1	1													
lifelong learning										1	2												
Multimedia	1	1		1	1	1	1					1						1					
multimedia instruction							1	6					1			1	5						
multimedia materials				1					1	1					1				1	2			
non traditional education										1													
online systems																						1	2
open universities		1								1													
part time students										1													
study material									1														
Web	1											1	1								1		
world wide web																						1	10
Total query terms*	4	3		4	2	3	5		5	10		3	4		2	2		3	3		3	5	
Non-thesaurus terms	3	1		2	2	2	2		3	2		3	2		1	0		3	0		2	1	
Thesaurus terms	1	2		2	0	1	3		2	8		0	2		1	2		0	3		1	4	
Exact hits in descriptors			5					7			7			5			9			9			19
Total documents retrieved			5					9			5			10			5			6			12
Docs. retrieved by 2 terms			0					0			3			6			4			1			6
Docs. retrieved by 3 terms			0					0			0			0			0			1			1

Appendix 5.3 – ERIC study – Thesaurus versus free text terms

Non thesaurus search	Times used
computer based learning	1
delivery	1
development	3
distance learning	7
home study	1
learning	2
learning material	1
multimedia	6
multimedia materials	4
study material	1
web	2
Amount of terms used	27
Thesaurus terms used in the free text search	**8**
Amount of distinct terms	**11**
Amount of distinct thesaurus terms used in the free text search	**4**

- Terms in bold are thesaurus terms
- Arrows indicate where terms from the free text search have been used in the thesaurus search

Thesaurus search	Times used	
adult education	1	(Thesaurus term)
computer assisted instruction	1	(Thesaurus term)
computer mediated communication	1	(Thesaurus term)
continuing education	1	(Thesaurus term)
development	1	(Thesaurus term)
distance education	2	(Thesaurus term)
distance learning	4	
education environment	1	
educational media	1	(Thesaurus term)
home study	1	(Thesaurus term)
Hypermedia	1	(Thesaurus term)
Independent study	1	(Thesaurus term)
Learning	1	(Thesaurus term)
learning materials	1	
lifelong learning	1	(Thesaurus term)
Multimedia	3	
multimedia instruction	3	(Thesaurus term)
multimedia materials	2	(Thesaurus term)
non traditional education	1	(Thesaurus term)
online systems	1	(Thesaurus term)
open universities	2	(Thesaurus term)
part time students	1	(Thesaurus term)
Web	1	
world wide web	1	(Thesaurus term)
Amount of terms used	**32**	
Free text terms	**5**	
Amount of distinct terms	**24**	
Amount of distinct thesaurus terms	**19**	

Appendix 5.4 ERIC study - Thesaurus term clusters

For more information on traversals see section 5.3.3.4

1. Distance learning (free text term) ->

Relation	Term	Times used	Traversals	Users
Thesaurus term	Distance education	2	0	Iota, Kappa
Related terms				
	COMPUTER MEDIATED COMMUNICATION	1	1	Beta
	CONTINUING EDUCATION	1	1	Epsilon
	HOME STUDY	1	1	Epsilon
	INDEPENDENT STUDY	1	1	Epsilon
	LIFELONG LEARNING	1	1	Epsilon
	NONTRADITIONAL EDUCATION	1	1	Epsilon
	OPEN UNIVERSITIES	2	1	Epsilon, Beta
	PART TIME STUDENTS	1	1	Epsilon

2. Multimedia (free text term) ->

Relation	Term	Times used	Traversals	Users
Thesaurus term:	Multimedia instruction	3	0	Delta, Eta, Iota
Related terms				
	HYPERMEDIA	1	1	Eta
	EDUCATIONAL MEDIA	1	1	Delta
Relation	**Term**	**Times used**	**Traversals**	**Users**
Thesaurus term:	Multimedia materials	2	0	Epsilon, Kappa
Related terms				
	{HYPERMEDIA – alternative possibility}	1	1	Eta

3. Computer base learning (free text term)->

Relation	Term	Times used	Traversals	Users
Thesaurus term:	Computer based laboratories	0	0	
Alternative	COMPUTER ASSISTED INSTRUCTION AND LABORATORIES	0	1	(traversal by Kappa)
Alternative	COMPUTER ASSISTED INSTRUCTION	1	1	Kappa

4. Internet

Relation	Term	Times used	Traversals	Users
Thesaurus term:	Internet	1	0	Theta
Related term	ONLINE SYSTEMS	1	1	Theta
Narrower term	WORLD WIDE WEB	1	1	Theta

5. Learning (free text term)

Relation	Term	Times used	Traversals	Users
Thesaurus term:	Learning	0	-	
Narrower terms	ADULT LEARNING	1	1	Theta
Related term[2]	EDUCATIONAL ENVIRONMENT	1	I	Theta

[2] Of Learning!

Appendix 5.5 SOCS study – Mini questionnaire on searching background

How long have you been working on your project?

How regularly do you normally search a database?

Which databases do you normally use? Are they available on the web or only in the library?

Which other information sources have you used? (For example references in papers, but anything really. Please indicate main sources with X)

Have you had a seminar/workshop (e.g. by a librarian) on how to search for information in electronic databases or on the internet? If so, please indicate when or how long ago this was.

Appendix 5.6 SOCS study – Search request form

Project stage
Orientation phase
Already past an initial orientation, but still clarifying some areas
The project has just been defined
The project has been defined some time ago, but new areas that need clarifying have appeared.
Monitoring the progress of the field
About to come to a conclusion

Subject of the search
Please describe the topic area for which you would like to locate references. Be as detailed as possible, listing any terms that may have special meaning in your request.

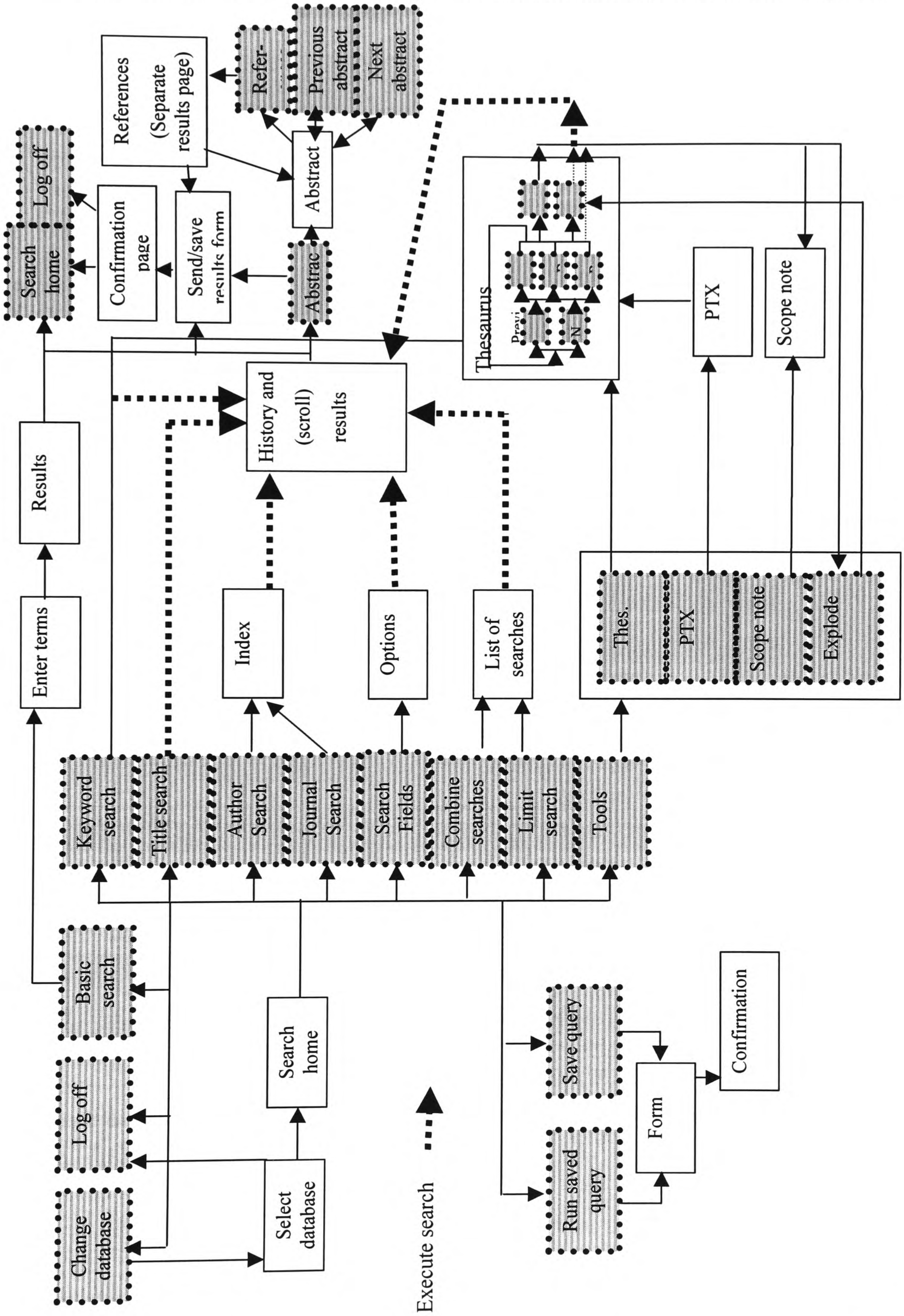
Use of results
Speech
Research paper
Thesis/dissertation
Article/book publication
Other (please specify)

Purpose of the search
Defining the project (etc) scope/topic
Looking for information on the project topic
Clarifying an area of the project
Clarifying an area related to the project
Monitoring the progress of information in the field (e.g. re-executing a search)

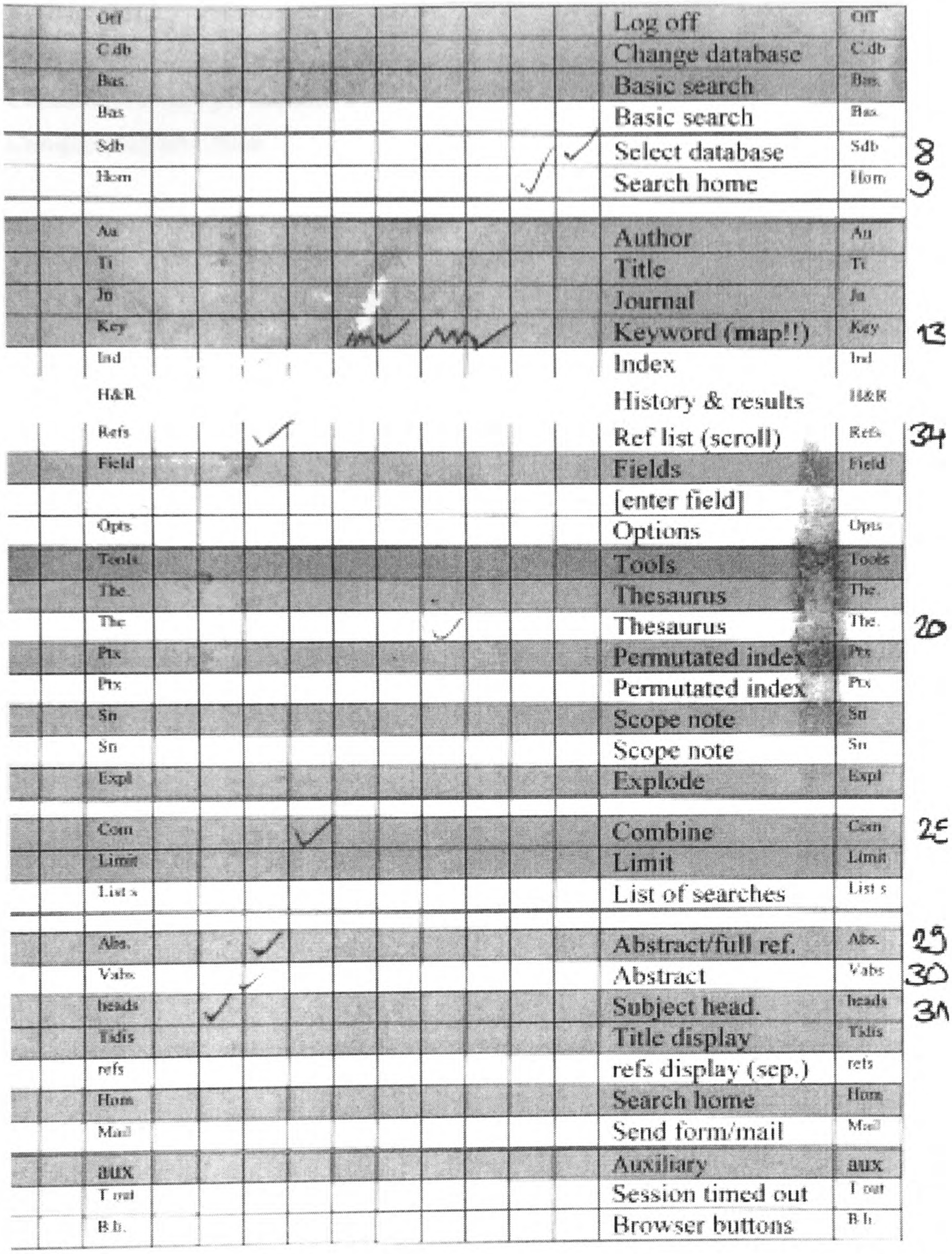
Search requirements
The search should be
- A narrow search, locating some relevant articles
- A comprehensive search, possibly including some peripheral material

Search limitations
Please specify any limitations which are applicable to your request, including such things as languages of interest and publication types (e.g. journals, books, dissertations) to be included.

Appendix 5.7 SOCS study – Flowchart of the Ovid interface



Appendix 5.8 SOCS study– Part of a filled-in record sheet



Off	Log off	Off	
C.db	Change database	C.db	
Bas.	Basic search	Bas.	
Bas	Basic search	Bas.	
Sdb	Select database	Sdb	8
Hom	Search home	Hom	9
Au	Author	Au	
Ti	Title	Ti	
Jn	Journal	Jn	
Key	Keyword (map!!)	Key	13
Ind	Index	Ind	
H&R	History & results	H&R	
Refs	Ref list (scroll)	Refs	34
Field	Fields	Field	
	[enter field]		
Opts	Options	Opts	
Tools	Tools	Tools	
The.	Thesaurus	The.	
The	Thesaurus	The.	20
Ptx	Permutated index	Ptx	
Ptx	Permutated index	Ptx	
Sn	Scope note	Sn	
Sn	Scope note	Sn	
Expl	Explode	Expl	
Com	Combine	Com	2£
Limit	Limit	Limit	
List s	List of searches	List s	
Abs.	Abstract/full ref.	Abs.	25
Vabs	Abstract	Vabs	30
heads	Subject head.	heads	31
Tidis	Title display	Tidis	
refs	refs display (sep.)	refs	
Hom	Search home	Hom	
Mail	Send form/mail	Mail	
aux	Auxiliary	aux	
T out	Session timed out	T out	
B.b.	Browser buttons	B.b.	

This is the record sheet for User 3's session. Numbers on the right correspond to assigned values (See section 5.4.4.1).

Appendix 6.1 – Camtasia settings

System, display colour : 16 bit

System, display resolution: 1024x768

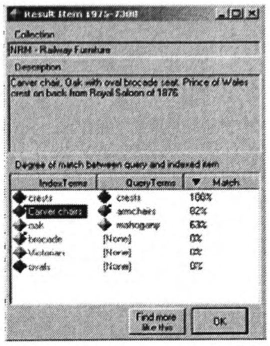
Camtasia, Frames per second: 5

Compression rate: Best

Appendix 6.2 – Scenarios

In order to assess this interface, we have compiled a number of short tasks which look at using the interface in different ways. Please read and complete the tasks below. If you have any questions about what to do or regarding the interface, please ask. We are interested in how usable the interface is, and if there is anything unclear, it is particularly important to us to know about it. If possible, please tell me what you are thinking while going through the tasks. Remember, we are evaluating the interface, not you.

Scenario 1 (Finding records in the collection)



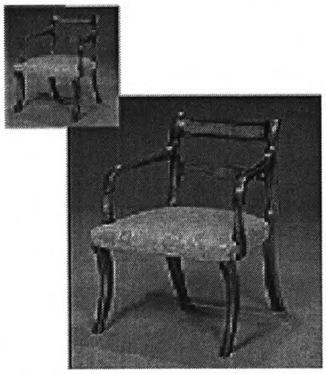
Create a new query (e.g. select "Query" menu and "New"). Please search the collection for a record of a table that would go best with this chair. (Presented here on the right.)

Remember: You can drag terms from the record, the thesaurus browser or the term finder into the query form.

Scenario 2 (Finding records in the collection)

Please search the collection for items that are decorated with text of some form. You may look at records and modify your query if you wish.

Scenario 3 (Finding records in the collection)



Please search the collection for similar to the item on the photograph (Presented here on the right.). Please try to be specific in the first query.

Scenario 4 (All interface components, finding records in the collection)

This is not a set task, but gives you an opportunity to look at or ask about certain aspects of the interface which you might have noticed while completing the tasks. Feel free to explore!

Temporary tasks

X
The following tasks are browsing tasks and should be answered using the thesaurus browser only. The scope notes in the top right corner of the thesaurus browser window will provide definitions of terms.

a) What is the name for a tall building without lifts?
b) How do you call an instrument used to calculate natural phenomena such as the rising of the tides?
Y
Through this task, we would like to find out about how many hierarchical levels of the thesaurus to present. Please use the term finder to find term A, then double-click it to open the thesaurus browser. Look for term B in this window.

A. Colours	B. Dark red
A. Attributes and properties	B. Gloss
A. Disciplines	B. Logic

Appendix 8.1 – FACET 2 – Training handout

The instructions from the presentation are shown here in an expanded form so that you might need to swap less between FACET and PowerPoint.

Finding query terms – left hand side of the Query Builder

- Click the "new" button on the toolbar to open the Query Builder with a new query.
- In the Query Builder, enter "wooden chair" into the text box and click "Find".
- From the list of candidate terms, drag "wood" into the query
- Double-click "chairs" to bring up the thesaurus
- Browse the thesaurus a bit to get used to doing this. You could use the "Back" button.
- Drag "Chairs" into the query on the right hand side.

Setting up the query – right hand side of the Query Builder

- Right-click on "chairs" and select "Set focus term" from the pop-up menu.
- From the options box above the query terms, select "Furnishings" as the theme of the query.
- Click on the term expansion tab. Change the term expansion by dragging the slider and moving it down until it is large enough to also include terms other than "chairs" and "wood".
- Click the "OK" button on the bottom of the Query builder window. The window will close and the query appears in the list.
- Press the "Run" button on the main toolbar to execute the query.

Judging the results – Main window

- Have a look through the list of result records.
- Double-click a record that does not have a 100% match to see the record form.
- Compare the indexing and query terms.
- Which query terms are not matched exactly?
- Which indexing terms caused the record to be retrieved?

Improving the query

- Open a record that has indexing terms which might be appropriate for the query.
- Right-click a term you want to add to the query.
- Select "Add to query" from the pop-up menu.
- Look at the query in the list – the term will have appeared with the other query terms.

Appendix 8.2 – Outline of the training presentation

Session introduction

Slide 1 – Overview

Slide 2 – Information on thesauri

Slide 3 – Use of thesauri in searching and indexing

Slide 4 – Basic searching in thesaurus-based systems

Slide 5 – Overview of the search process

Slide 6 – Steps: Identifying query terms

Slide 7 – Steps: Searching the thesaurus

Slide 8 – Steps: Setting up the query

Slide 9 – Steps: Judging records

Slide 10 – Steps: Query formulation

Slide 11 – Recapitulation of the search process

Use of FACET

Slide 1 – Finding keywords

Slide 2 – Searching the thesaurus – information

Slide 3 – Searching the thesaurus – demonstration

Slide 4 – Setting up the query – information

Slide 5 – Setting up the query – demonstration

Slide 6 – Judging results – information

Slide 7 – Judging individual records – information

Slide 8 – Judging results - demonstration

Slide 9 – Improving the query –terms – information

Slide 10 – Improving the query –term expansion – information

Slide 11 – Query reformulation – demonstration 1

Slide 12 – Query reformulation – demonstration 2

Slide 13 – Other options for starting a query

Slide 14 – Recapitulation

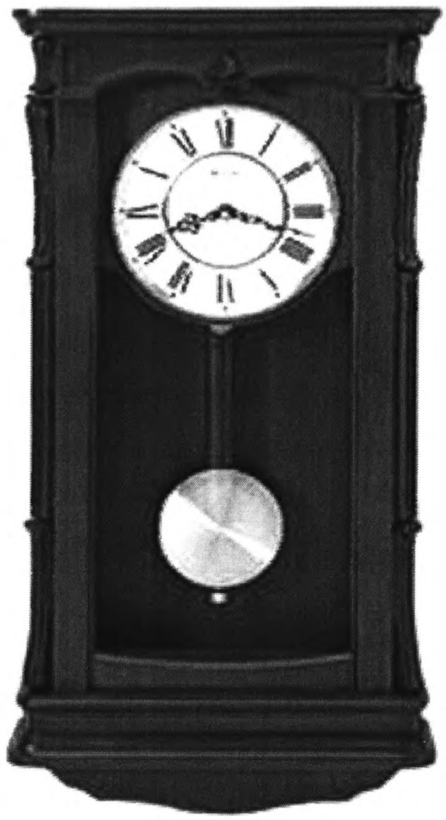
Appendix 8.3 – FACET 2 - Scenarios

Rejected task 1

Please search the database for a chair with carvings, ideally from the late 19^{th} century.

Task 1 – Standard Scenario

Please find a record in the database that could describe the item in the image below[3]. Imagine this is a picture I took of a clock that I have at home. Find the closest matching object in the museum collection so that I can go and compare it to mine.



Task 2 – Standard scenario

This task is more open to give you the opportunity to try out some features of the interface such as the term expansion mechanism. Modify and run either one of the existing queries or create a new one. Please discuss your ideas with the evaluator before you begin.

Task 2 – Scenario for the curators

This task is more open to give you the opportunity to try out some features of the interface such as the term expansion mechanism. Modify and run either one of the existing queries or create a new one to find out about the different materials used to upholster or cushion chairs. Which ones seem to be the most common ones used? (You do not need to work this out exactly!) Please discuss your ideas with the evaluator before you begin.

[3] Image here in reduced size on the right.

Appendix 8.4 – FACET 2 – Self-administered questionnaire

I would like you to answer the following questions to provide some background information for the data analysis. Please let me know if this sheet does not seem to let you answer a question appropriately and explain the situation. This information will be treated confidentially and anonymously.

What is your present occupation?

Roughly, how long have you worked in this and other positions where you did comparable work?

Information on computer use and information searching

1. Which of the following packages do you use? Please indicate how regularly you use them.

	Never	Occasionally	Monthly	Weekly	Daily	Unsure
Email						
Word processing						
Web surfing						
Games						
Multimedia (e.g. music, films,…)						
Programming						
Computerised library catalogues						
Local databases, inc. those where you might maintain your own data						
Web databases which might or might not require you to log in						
CD-ROM based systems, e.g. Encarta						
Other research on the web						
Other packages, please specify below						

2. Did you ever take courses in information searching/seeking?

Yes | No

If yes, when?

Secondary school | Professional training

University or equivalent | Personal development

3. Do you spend any time searching for information?

Yes No

3.1 If you spend some time searching for information, what are the main purposes for this?

3.2 How much time do you spend searching for information? (e.g. "1 hour per week" If this varies largely, please specify the circumstances.)

3.3 Are there any databases or information collections that you search regularly?

Yes No

If yes, which are the main ones?

4. Have you ever used a thesaurus when searching for information?

Yes No

4.1 If yes, in which context was this?

4.2 If yes, how long was this ago?

4.3 How regularly did you use a thesaurus?

Occasionally	Monthly	Weekly	Daily	Unsure

4.4 If you remember it, please write down the name of the thesaurus or the application it was used with. If you do not remember specific information, you might want to tell the evaluator about information such as the subject area, type of database/system, etc.

5. Do you currently use a thesaurus? How regularly do you use one?

Never	Occasionally	Monthly	Weekly	Daily	Unsure

5.1 Which thesaurus is it/thesauri are they?

6. Do you enjoy online searching?

Appendix 8.5 – FACET 2 - Post-search questions

General

Was it easy or difficult to learn the system?

Was it easy or difficult to use the system?

Did you feel rather familiar or unfamiliar with the topic?

Comparing to the information searching you would normally do

Where the tasks similar to tasks you would normally complete when searching for information? In how far did they differ?

In how far is the system different or similar to systems that you normally use?

If you have used a thesaurus before, how was using this one/ this one in this system different or similar to that?

Term expansion function

What do you think does the term expansion do?

Was it easy or confusing to understand what it did?

Did you feel your queries were improved or not improved due to the term expansion (not so much the thesaurus)?

Query reformulation

Did you feel that modifying your query improve it or not?

General interface

Do you feel that there are any points where people might require more assistance in order to help them searching for information?

How did you feel about the results you retrieved?

Do you think any feature of this interface is beneficial to information searching/seeking?

How do you think it would be possible to support users further in their searching for information?

Is there anything that could have influenced your search positively?

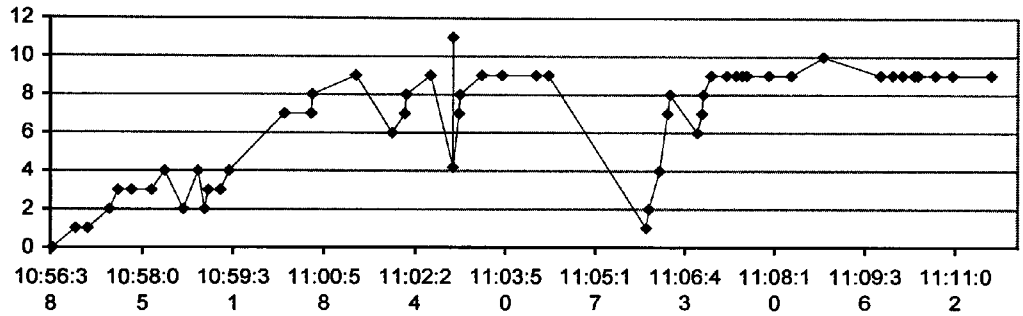
Appendix 8.6 - Brief descriptions and graphs of search sessions

Legend for interactions in figures

0 – creating new query
0.5 – "more like this option"
1 – Term Finder search
2 – double-clicking candidate term
3 – double-clicking thesaurus term
4 – dragging a term into query
4.2 – adding a term via menu
5 – changing the expansion
6 – deleting a term
7 – running the query
8- results are returned
9- opening a record
10 – displaying an indexing term in the thesaurus
11 – adding a term from a record to the current query

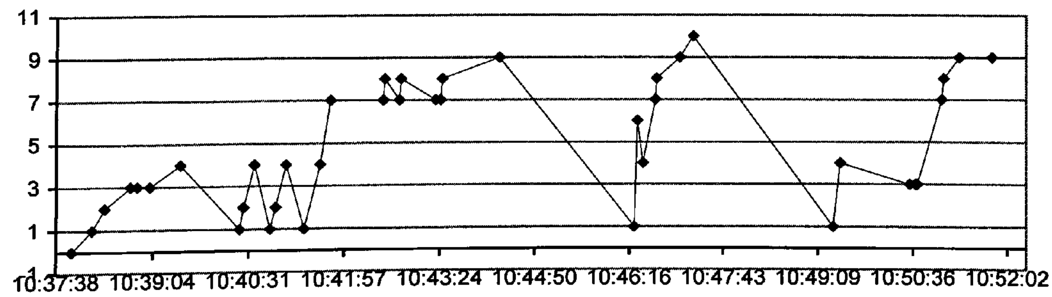
To maintain the participants' anonymity, the pronouns "he, his, him" have been used uniformly in these descriptions.

Participant A (clock)



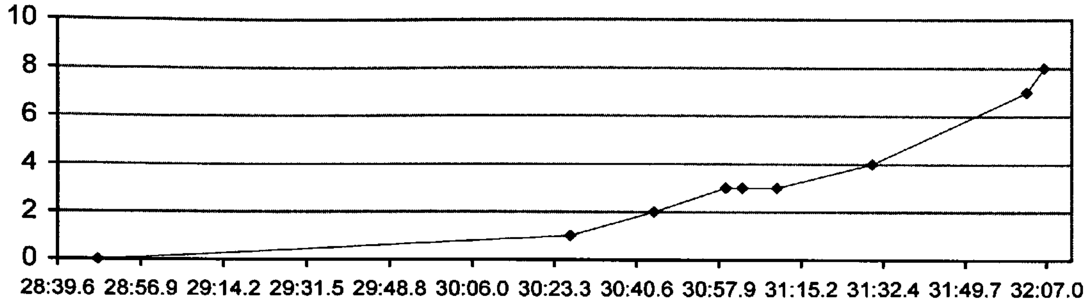
The user searches the thesaurus for "pendulum clocks". He adds "clocks", "pendulum" and "weight-driven clocks" to the query and sets the expansion to 28%. He then runs the query and deletes "<weight-driven clocks>" from the query and re-runs it. He adds "roman numerals" from a record and runs the query again. He browses some records and decides to search the thesaurus for "rectangular" and adds the term "rectangles" to the query. After running the query, the searchers deletes the term again though. He runs the query one last time and looks at some records. He identifies suitable ones, but also explains that the indexing does not make it worthwhile to continue searching.

Participant B (clock)



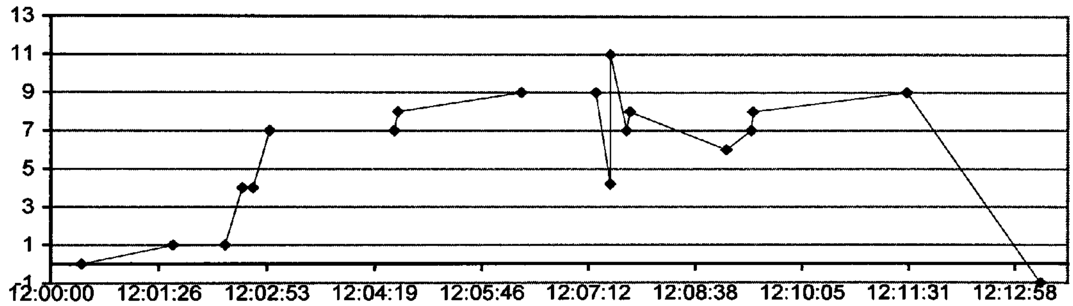
The user searches the thesaurus for "clock" and then browses "clocks" in the local context. He drags "wall clocks" into the query. He then searches one-by-one for "roman numerals", "pendulum" and "carving" and drags each of the terms after some browsing into the query. He runs the query but it does not retrieve any results. The searcher changes the theme from "Furnishings" to "Timepieces". The user re-runs the query. He then looks at some records and searches for "carving". He deletes "carving" and drags "carvings" into the query. After re-running the query, he then looks at the indexing term "cases (containers)" in its local context in the thesaurus and searches the thesaurus for "wooden case". He drags "wood" and "cases (containers)" into the query and then browses the "Properties" hierarchy a bit. He runs the query and looks at the results.

Participant C (clock)



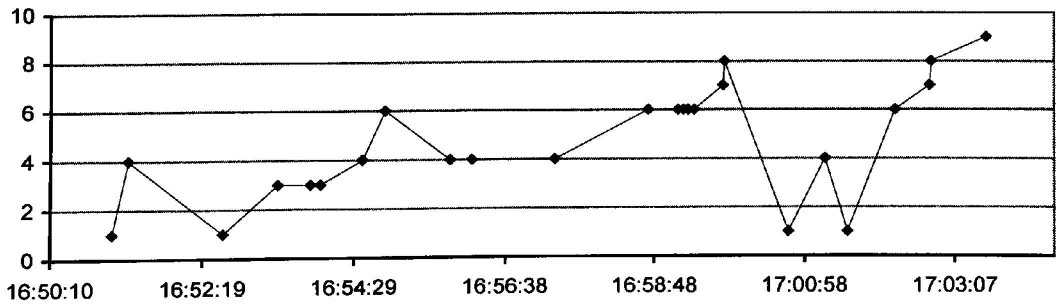
The user searches the thesaurus for "clock roman face drop case". He then drags the term "wall clocks" into his query and runs it. Scrolling the records, he identifies one which contains suitable indexing terms.

Participant D (clock)



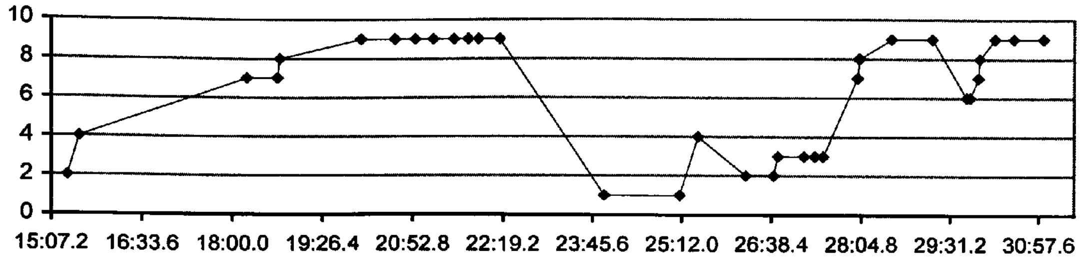
The user creates a new query and searches the thesaurus for "wall hanging clock" and then "wall hanging clock wooden". He drags "clocks" and "wood" into the thesaurus. He runs this query and looks at some records. He adds the term "wall clocks" from one of the records. He runs the query, and then takes "clocks" out and makes "wall clocks" the focus term. He re-runs the query and looks at a record.

Participant E (clock)



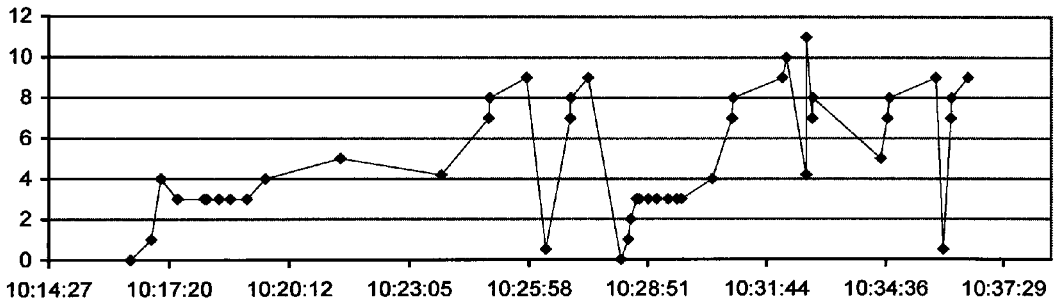
The user opens an existing query, searches the thesaurus for "pendulum clocks" and drags "<clocks by movement>" into the query. He searches the thesaurus for "equilibrium" and browses "<clocks by movement>". He drags "atomic clocks" into the query and deletes "pocket watches". He then browses "atomic clocks" and drags "self-winding clocks", "water clocks" and "mechanical clocks" into the query. After a discussion with the evaluator, he deletes all terms apart from "mechanical clocks" and runs the query. After looking through the list of records, he searches the thesaurus for "pendulum clocks" and drags "pendulums" into the query. He then searches the thesaurus for "pendulums" and deletes "mechanical clocks". He runs the query containing only the term "pendulum" and reading through the list of results, he decides on a suitable record which he opens.

Participant F (clock)



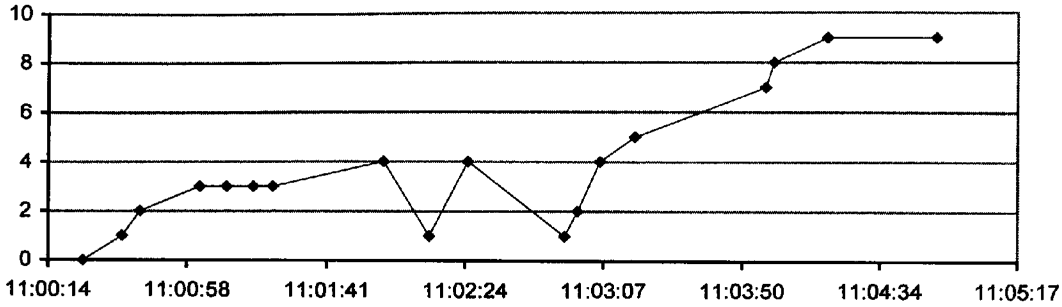
The user searches the thesaurus for "wooden clock pendulum roman numeral" and drags "clocks" into the query. While browsing he also drags "pendulums", "roman numerals" and wood" into the query. He sets the query expansion to 100% and then back to 33%. He runs the query and after browsing some records, he decides to use the term "wall clocks". He thus searches the thesaurus for "wooden clock roman numeral wall", and tries again with "wall clocks" when the appropriate term is not retrieved. He drags the terms into the query and browses it in the thesaurus. He runs the query and after browsing some records, he deletes "clocks" from the query and re-runs it. He identifies a suitable record from the list of results.

Participant G (clock)



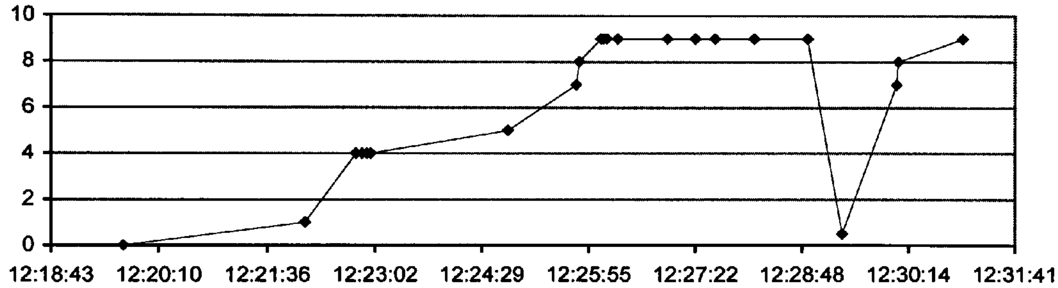
The user searches the thesaurus for "clock" and after browsing, drags "clocks" and "bracket clocks" into the query. He increases the term expansion to 75% and adds "chiming clocks" from the list of expanded terms to the query. He runs the query and opens a results record. He uses the "More like this" option and executes the resulting search. He creates a new query and searches the thesaurus again for "clock" and adds the term to the query. After running the query, he opens at some records and adds "pendulum" from one of them to the query. He runs the query and then increases the expansion rate to 93%. He re-runs the query and finds a record where he again uses the "More like this" option. He runs the resulting query and selects a matching record without opening it.

Participant H (clock)



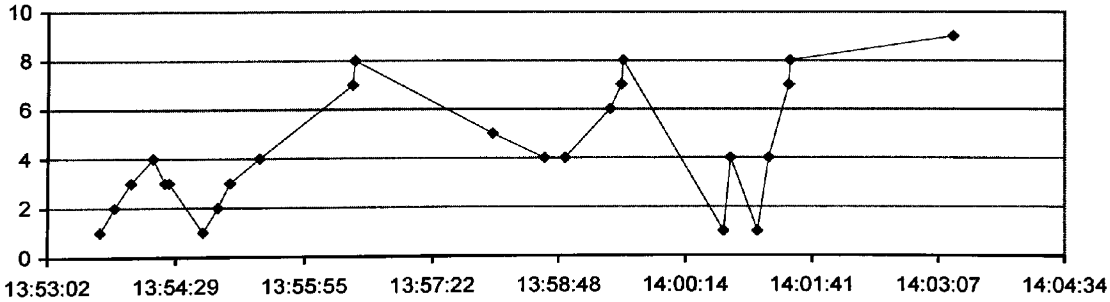
The user searches the thesaurus for "clock" and browses "clocks" in the local context. He drags "clocks" into the query. He then searches the thesaurus for "wood" and "pendulum" and after each search, drags the term into the query. He increases the term expansion to 20% and runs the query. He opens some records and identifies some of the higher ranked ones as suitable results.

Participant I (clock)



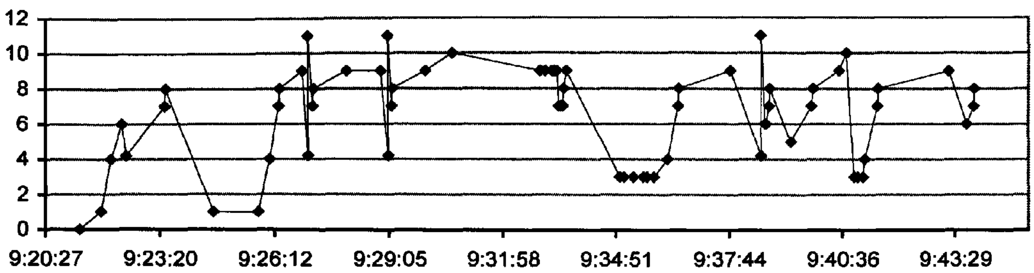
The user searches the thesaurus for "clock wooden case pendulum glass front wall". The then drags the terms "clock", "wood", "pendulum" and "glass" into the query. He increases the expansion to 12. He runs the query and opens some results. He then uses the "More like this option", runs this new query and opens another record.

Participant J (clock)



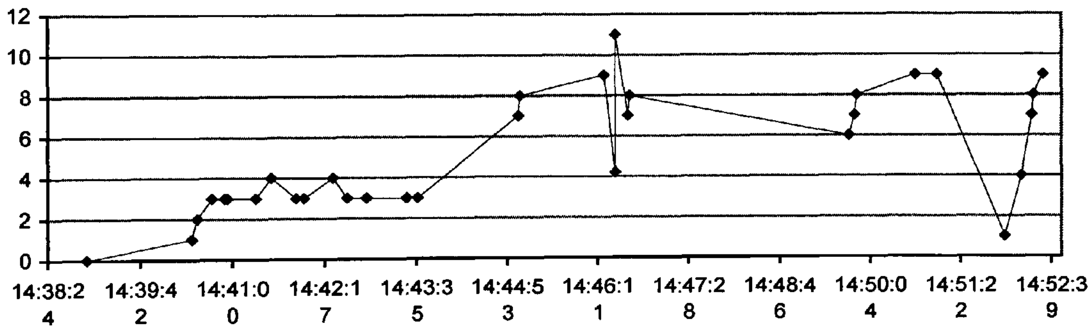
The user searches the thesaurus for "wooden clock", restricting the search to the Objects category. He browses the term "clocks" in its local context and drags the term into the query. He then starts browsing the "Materials" hierarchy from the top level, but prompted by the author, he searches for "wooden clock" in the Materials category. He browses "wood" and drags the term into the query. The user then runs the query and afterwards, remembers to set the expansion. He sets it first to 100% and then to 22%. He browses the term "wall clocks" in the expanded display, adds it to the query and deletes "clocks". He runs the query and looks at the list of records. He searches the thesaurus for "pendulum" and then also for "roman numerals" and drags the terms each time into the query. He re-runs the query, identifies records that fulfil the task requirements and opens one.

Participant K (clock)



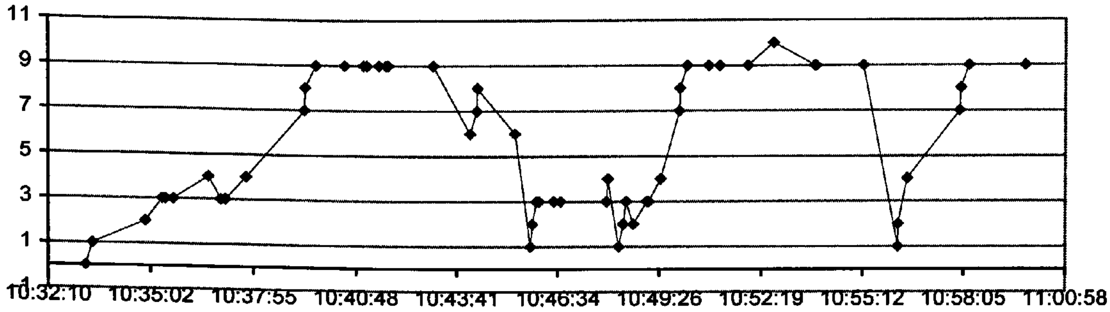
The user searches the thesaurus for "clocks". He browses the term in its local context and drags "<clocks by form or function>" into the query. He deletes the terms straight away and adds "clocks" using the menu. He runs the query and makes an attempt to find "clocks" in the Materials category, which does not retrieve thesaurus terms. He then searches for "wood" in Materials. He drags "<wood and wood products>" into the query and runs it. He adds "pendulum" and "roman numerals" from records to the query and after each, re-runs it. He looks at "stations", an indexing term in a record, in its local context and then re-runs the query. The user browses the thesaurus and adds "panel (wood)" to the query. After re-running the query, he adds "wood" and deletes "panel (wood)" and "<wood and wood products>". He runs the query and changes the term expansion to 22%. After re-running the query, the user browses "clocks" in the thesaurus and drags "wall clocks" into the query. After another query run, he deletes "clocks" and runs the query again.

Participant L (clock)



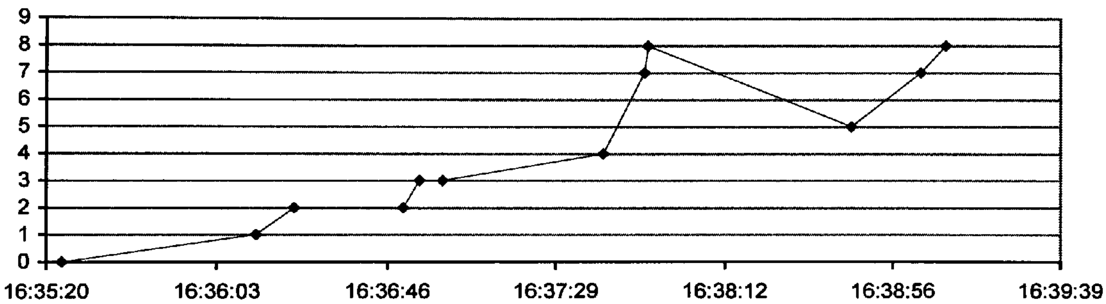
The user searches the thesaurus for "clock" and browses the term. He then drags "<weight-driven clocks>" into the query and after some more browsing, also selects "wall clocks". He runs the query, and looks at the list of records. He adds "pendulums" from one of the records to the query. He re-runs the query, deletes "<weight-driven clocks>" and runs it again. After opening more records, he searches the thesaurus for "roman numerals" and drags the term into the query. The user identifies records which contain all three terms.

Participant M (clock)



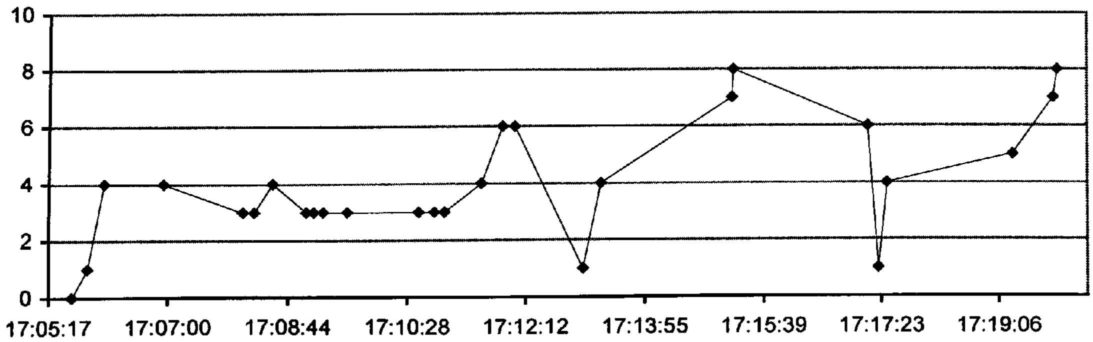
The user searches the thesaurus for "clock". He browses the term in its local context and drags "wall clocks" and "<weight-driven clocks>" into the query. He changes the term expansion, but sets it back to 0%. The user runs the query. After looking at some results, he deletes "wall clocks" from the query. He re-runs it, but no records are retrieved. The user deletes "<weight-driven clocks>". He searches the thesaurus again for "clocks" and drags the term into his query. He then also searches the thesaurus for "wood" and drags this term into the query. He considers higher expansion values, but again sticks with 0%. He runs the query and after looking at some records, he inspects the term "wood" in its local context. He searches the thesaurus for "pendulum" and adds the term to the query. He now sets the expansion value to 21%, runs the query and looks at some more records.

Participant C (upholstered chair)



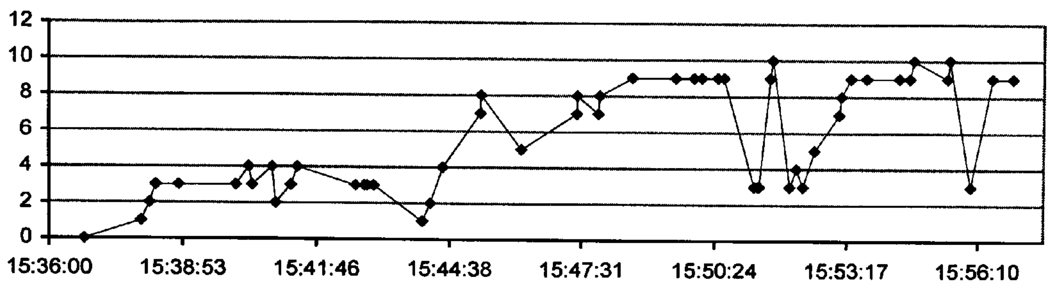
The user searches the thesaurus for "chair upholstered". He browses first "<upholstered seat>", then "chairs" in their local context and drags "<chairs by form>" into the query. He runs it but no results are returned. He increases the expansion rate to 39% and re-runs the query. Looking at the records, he states that apart from those chairs which are wooden, all records are relevant.

Participant E (upholstered chair)



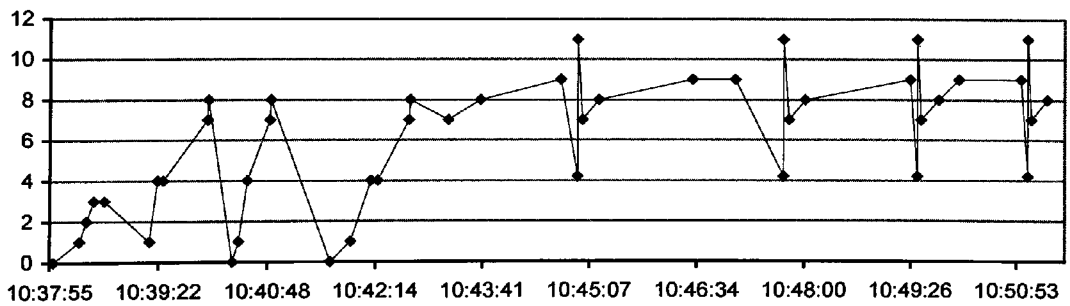
The user searches the thesaurus for "moquettes". He drags this term into the query, browses the thesaurus starting from "furniture" and adds more terms to the query, some of which he deletes again. He then searches the thesaurus for "upholstering" drags this term into the query. He runs the query, and then deletes "moquettes". He searches the thesaurus for "leather" and adds the term to the query. He changes the expansion value to 70 and runs the query again.

Participant F (upholstered chair)



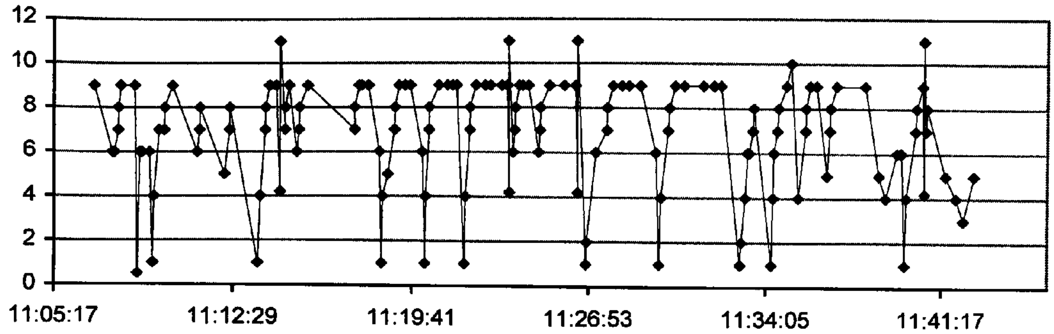
The user searches the thesaurus for "upholstered chair". He browses the term "<upholstery foundation components>" in its local context and drags "webbing (material)" and "chairs" into the query. He also browses "upholstery" in its local context and drags the term into the query. He then starts browsing the "Materials" hierarchy from the top, but then prompted by the evaluator, he searches the thesaurus for "upholstery cotton chair" and browses "cotton (textile)". He drags "<textile materials by composition or origin>" into the query. He runs the query, but no results are returned. He changes the expansion value to 100% and then down to 27%. The query still does not return any results. The user then changes the focus term, now records are retrieved and he opens some of them. Seeing "upholstering" as an indexing term, he starts browsing the "Processes" hierarchy from the top. Prompted by the evaluator, he opens the records and looks at the local context of the term. He drags "upholstering" into the query, increases the expansion rate to 49% and re-runs it. Looking at the records, he identifies some materials and looks at "imitation leather" and "padding (process)" in their local context before opening more records.

Participant G (upholstered chair)



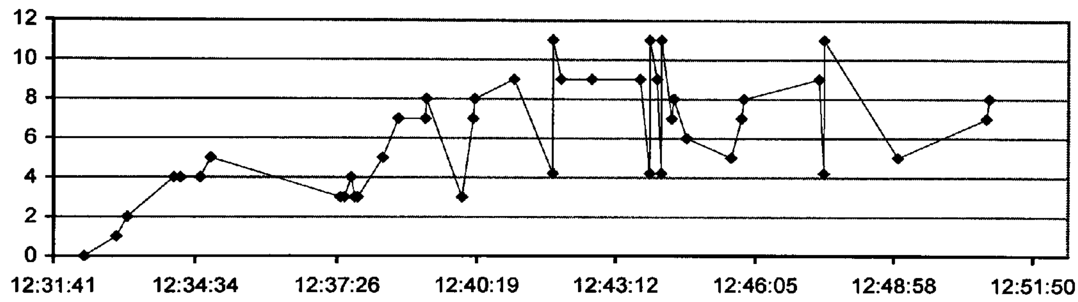
The user searches the thesaurus for "upholstery". He closes, re-opens the query and repeats the search. He drags the terms "upholstery" and "<upholstery components>" into the query. When executed, it does not return any results. He creates a new query, searches the thesaurus for "upholstery" and drags it into the query. Again, there are no results. He then creates another new query and searches the thesaurus for "materials". From the candidate list, he drags "materials" and "<materials by form or function>" into the query. This query also does not retrieve any results. The user then increases the term expansion to 96%. Now, results are retrieved and the user adds "upholstering" to the query. He re-runs it and also adds "padding (process)", "cushions" and "cloth" from records and runs the query again each time. The evaluator then terminates the session.

Participant H (upholstered chair)



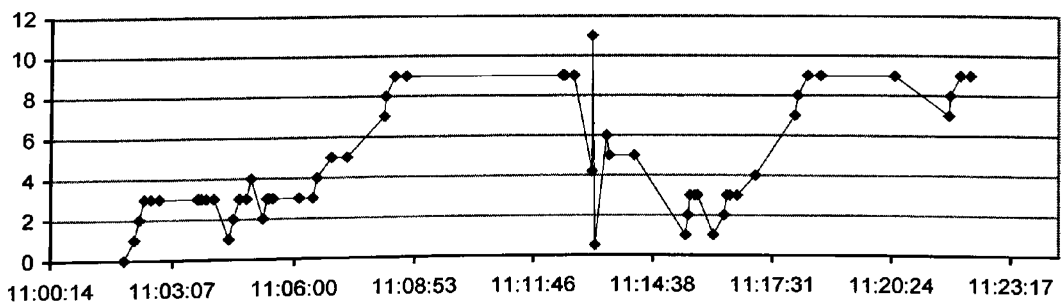
The user clicks on an existing query for "Edwardian armchairs" and browses some records. He deletes terms from the query and runs it with the terms "armchairs, Edwardian" remaining. He opens some records and uses the "More like this option". After deleting several of the terms, the user searches the thesaurus for "armchairs" and adds this term to the query. He runs the resulting query and compares it to the previous one. He then realises that he has not changed the query expansion and sets it to 20%. In the following interactions, the user runs the query, opens records and uses different combinations of adding terms from records or from the thesaurus and deleting terms before re-running it.

Participant I (upholstered chair)



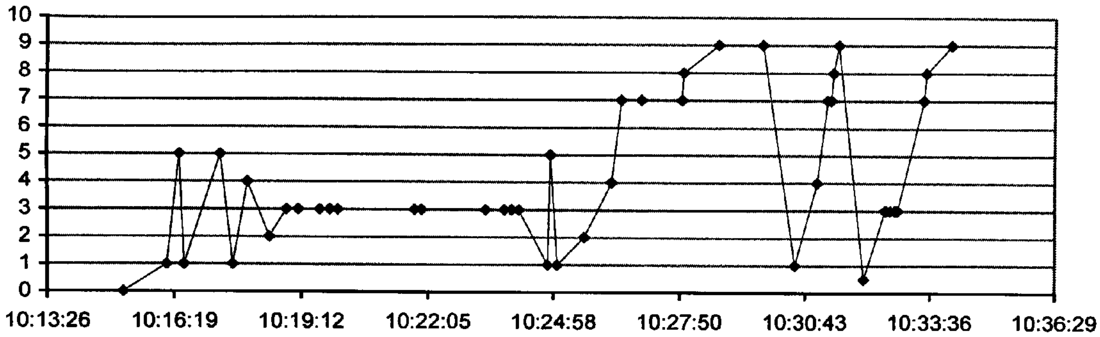
The user searches the thesaurus for "upholstered chair". He then adds the terms "chairs" and "<upholstered seat>" to the query. He tries out different term expansion values and finally sets it to 91%. He browses more terms and accidentally adds "paper-making". He runs the query, but no results are retrieved. He changes the focus term and after running the query again, he adds one by one "moquettes", leather" and "imitation leather" from records to the query. He re-runs the query, deletes "paper-making" and sets the expansion down to 35%. After running the query, he adds "cloth" from a record, increases the term expansion to 81% and re-runs the query.

Participant M (upholstered chair)



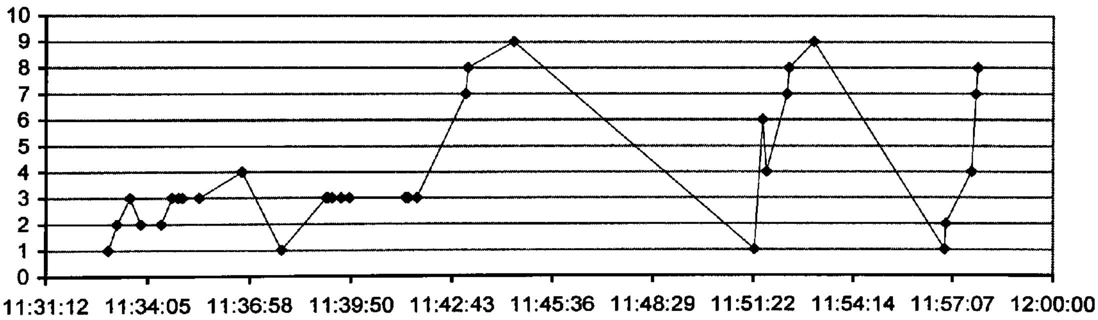
The user first searches the thesaurus for "upholstery" and after some browsing of the thesaurus for "chair upholstery". He then drags the term "chairs" and a little later "<upholstery and upholstery components>" into the query, changes the query expansion first to 28% and then to 20% and runs the query. He opens some records and adds "upholstering" from a record to the query and uses the "More like this" option. He deletes "wood" from the new query and sets the term expansion to 91%. He then again searches the thesaurus for "upholstery" and after some browsing of the results, for "upholstery materials". He drags the term "upholstery" into the query and runs it. He then looks at the previous query to compare the results.

Participant B (19th century chair)



The user searches the thesaurus for "Chair carvings". He then searches for "Chair carvings late 19th century". He drags the term "chairs" into the query and clicks the "Times" query heading. He browses the term "<Late Indian styles and periods>" in its local context. He searches the thesaurus for "late 19th century" and browses the term "Queen Anne" in its local context. He drags the term into the query. He sets the theme to "Furnishings" and runs the query containing "Queen Anne, Chairs". He browses records and searches the thesaurus for "carving". He drags "Carving" into the query and runs it. Looking at the results, he uses the "More like this" option and browses "Times" in the thesaurus. He re-runs the query and discusses some options with the evaluator.

Participant D (19th century chair)



The user opens an existing query. He searches the thesaurus for "wooden chair carving 19th century". He browses the terms "carving", "carvings" and "chairs" in their local context in the thesaurus and drags "carvings" into the query. He then searches the thesaurus for "wooden chair carving period" and after clicking the "Browse" tab, he starts drilling down the "Time" hierarchy. The evaluator stops him and the user runs his query. After opening a results record, he searches the thesaurus again for the same terms as before. He deletes "carvings" and drags "carving" into the query. After re-running it, he searches the thesaurus for "Victorian" and browses the local context of the term. He then drags the term into his query and re-runs it.

Appendix 8.7 – Definitions of moves and tactics

The following moves and tactics stem from Bates (1979b) and Fidel (1985) and have been paraphrased in one single document in order to make them easier to work with for the purpose of the analysis.

Add1 - Add synonyms and variant spellings

Add2 - Add descriptors as free-text terms

Add3 - Add descriptors assigned to relevant citations retrieved.

Add4 - Add terms from the database's index that have a high number of postings

Cancel - Eliminate restrictions previously imposed.

Cut - Submit only part of the retrieved answer set, arbitrarily selected

Expand1 - Enter a broader descriptor

Expand2 - Group together search terms to broaden the meaning of a set

Expand3 - Group together descriptor with an equivalent role indicator

Expand4 - Represent a query component explicitly only by qualifying another component with role indicators

Exclude - Exclude from a formulation concepts present in most documents of the database

Expand5 - supplement a specific answer set with sets representing broader concepts

F1 Bibble - check for an existing bibliography or similar work which would be useful

F2 Select - Break complex request down and work on each component separately

F3 Survey - To review, at each decision point, the available options before selecting

F4 Cut - To choose the option that leaves out a large part of the domain

F5 Stretch - Use a source for other than its intended purpose

F6 Scaffold - To design an auxiliary, indirect route through the information files ad resources to reach the desired information

F7 Cleave - employ binary searching in locating an item in an ordered file

Include - Group together descriptor with all the descriptors that are narrower terms. This is the equivalent to "Explode" in Ovid.

Intersect1 - Intersecting a set with a set representing another query component

Intersect2 - Intersect sets with role indicators

Limit1 - Limit to documents written in a particular language

Limit2 - Limit to documents indexed or published in a particular period of time

Limit3 - Limit to documents retrieved from a specific portion of the database

Limit4 - Limit to sources that have, or not have, certain terms in their titles

M1 Check - Make sure original request and current search topic are the same

M2 Weigh - Make cost-benefit assessment of action to come

M3 Pattern - Reassess search pattern

M4 Correct - To watch for and correct spelling and actual errors in one's search topic
M5 Record - To keep track of trails one has followed and of desirable trails not followed up or completed
Narrow1 - Intersect a descriptor set with a set created by more specific free-text terms
Narrow2 - Qualify descriptors with role indicators
Negate - Eliminate unwanted elements by using the AND NOT operator
Probe1 - Construct an indexing probe set
Probe2 - Use the difference among the number of postings for a search term in various databases to decide how to represent the components in each database
Refine - Find a "better" descriptor
S1 Specify - To search on terms that are as specific as the information desired
S2 Exhaust - To include most/all elements of the query in the initial search formulation; to add one/more of the query elements to an already prepared search formulation
S3 Reduce - Reduce the number of elements in the search formulation
S4 Parallel - Broaden the search formulation with synonyms or otherwise parallel terms
S5 Pinpoint - To make the formulation precise by minimising (or reducing) the number of synonyms/parallel terms, while keeping the best suited one
S6 Block - To reject in the search formulation items containing or indexed by certain term(s) even if that means loosing some document sections of relevance
T1 Super - move to a hierarchically super ordinate term
T2 Sub - move to a hierarchically subordinate term
T3 Relate - Move hierarchically sidewise
T4 Neighbor - Look at neighbouring terms, e.g. alphabetically or by subject similarity to find more terms
T5 Trace - Examine results to find more terms to include
T6 Vary - To alter or substitute search terms
T7 Fix - Try alternate affixes (prefix, suffix, infix)
T8 Rearrange - To rearrange the order of search terms
T9 Contrary - To search for the term logically opposite from that describing the desired information
T10 Respell - Try a different spelling
T11 Respace - Try spacing variants
Weight1 - Limit a descriptor to be a major descriptor
Weight2 - Intersect free-texts with a broader descriptor
Weight3 - Limit free-text terms to occur in a pre-defined field
Weight4 - Require that free-text terms occur closer to one another in the searched field
Weight5 - Limit to documents of a certain format

Appendix 8.8 – Moves and tactics identified in the sessions

Abbreviation	Assigned value	Used during all sessions	Adapted definition
Adding terms to the query			
s2	1.3	3 (+ 39 before 1st query run)	Including another concept in the query
add3	1.2	1	Adding terms from records, which a user might or might not have thought about
s2/add3	1.1	12	Both apply.
include	1.4	9 (+ 24 before 1st query)	"Explode", include narrower terms. Here, this does not apply exactly
s4	1.5	(6 before 1st query run)	Including a term that describes a concept that is already included in the query
Replacing terms in the query			
t1	3.1	2	Using a BT
t2	3.2	3	Using an NT
t3	3.3	0	Using a sibling term
t3/t4		0	not sure which one applies
t3/[add3]		0	Refining the query using terms from records which have a relationship to existing query terms
t4	3.4	1	Using a related or alphabetically close term - this is biased towards systems that do not have a hierarchical structure?
t6	3.5	2	Replacing a term by another term - most general option
t6/add3			Refining the query using terms from records which have a relationship to existing query terms
t9	3.6	0	Using the opposite term. FACET does not actually provide functionality for this, but depending on the thesaurus, this is thinkable.
refine			Should not need to be going from BT to NT, but apply to any choice of a better term, so this could apply to all "T" moves
Removing terms from the query			
s3/s5	-1.1		These both refer to reducing the number of query terms.
s5	-1.2	(6 before 1st query)	Reducing parallel terms, which does not strictly speaking exist, but using term from different guide term hierarchies might fall under that. Extend to terms that describe the same concept, as there is no simply "remove".
s3	-1.3	(4 before 1st query)	Reducing the number of elements in the query
Exclude	-1.4	0	Opposite of increasing the expansion – this move has been created in order to express the opposite of "Increase"
Other moves/tactics			
More like this	4.5	3	No equivalent
Focus	0	2	Change in focus term
Limit	0.1	1	Limitations on the dataset, here, applied to selection of database.

Appendix 10.1 Tabular model representation

	MODEL	1	2	3	4	5	6	7	8	9	10	11	12	13	14
		1a) Concept/free text term	2) Candidate term list	Number of candidate terms [0]	Suitability of candidate terms [1]	3) Selected controlled term	Enough query terms [2]	1c/4) Query	Suitability of query terms [6]	5) Results	Number of results [3]	Suitability of results [4]	1b/6) Record	Aspects of record [5]	7) Collection of re
A	1a) Concept/free text term		1a-1												
B	2) Candidate term list			2-1	2-2										
C	Number of candidate terms [0]	2-4			2-3, 2-4										
D	Suitability of candidate terms [1]	2-12			2-6,2-7,2-9	2-8,2-10,2-11									
E	3) Selected controlled term		3-2				3-2	3-1							
F	Enough query terms [2]	4-7			4-6			4-1/4-4/1c-4				4-4	4-5, 6-9		
G	1c/4) Query						4-3		1c-1	4-2					
H	Suitability of query terms [6]					1c-3		1c-2							
I	5) Results										5-1				
J	Number of results [3]	5-3					5-6	5-3		5-5		5-4	5-3, 5-5		
K	Suitability of results [4]	5-7					5-11	5-7					5-7, 5-10		5-9
L	1b/6) Record						4-5						6-x	1b-1, 6-1	
M	Aspects of record [5]					6-6	6-7					6-2		5-9	6-3
N	7) Collection of records											7-2	7-3	7-1	

Appendix 10.2 Definition of processes from the model

Processes have been numbered depending on the entity from which they start, apart from interactions related to query formulation. Here, it was unclear which other entity these would relate to, for example the decision to return to a starting point/task definition in order to select more concepts (4-7). The searchers could have used any of the sources of query terms before making this move.

Several processes describe the termination of the search sessions. Searchers have at this point potentially identified useful records and the process encompasses activities necessary to make use of these, for example saving or printing them, as well as activities such as saving the query or logging out of the system. As these are do not directly impact the information search processes, they are not presented separately in this model.

1a-1 **Entity (1a) to entity (2)**

The steps required to map free text terms to candidate terms have not been modelled separately. The searchers normally have to enter the word or phrase into a mechanism provided by the interface. The system will then search for candidate terms and return these as an entity (2) Set of Candidate Terms. It is possible that the system itself selects terms which are deemed appropriate according to specific rules which form part of the matching mechanism.

1b-1 **Entity (1b) to decision [5]**

Once a record has been selected, the searchers will inspect it in detail. This leads to decision [5] the consideration of whether and how to use information from the record. If the searchers have chosen a relevant record to start a search, it is likely that the they select an indexing term for the query (6-6).

1c-1 **Entity (1c) to decision [6]**

The searchers assess an existing query on whether to work with it or whether to select individual terms. As it is assumed that the existing query has been identified as relevant, no alternative process is represented.

1c-2 **Entity (1c) to entity (4)**

An existing query can be modified in a similar way to a new or current query. Once the searchers decide to make modifications, the query can be treated as entity 4) Query, which represents a current query, which can either be reformulated further or executed.

1c-3 **Decision [6] to entity (3)**

The searchers select query terms from the existing query to include in the current query.

2-1 **Entity (2) to decision [0]**

The number of Candidate terms returned by the mapping mechanism is assessed. This leads to decision [0] the assessment of the number of retrieved terms.

2-2 **Entity (2) to decision [1]**

Some systems present searchers with a selection of Controlled Terms which searchers can browse and select without having to map concepts to the controlled vocabulary. This functionality leads from these Candidate Terms straight to decision [1a], the assessment of whether any of the available terms conform to the searchers' expectations.

2-3 **Decision [0] to decision [1a]**

The searchers look for the expected term amongst the candidate terms. This leads to decision [1c] Assessment of suitability of other Candidate Terms.

2-4 **Decision [0] to entity (1a)**

The searchers return to the starting point, (1a) Concept or Free Text term.

2-5 **The outcome of decision [0] leads to the end of the search.**

The searchers terminate the search.

2-6 **Decision [1a] to decision [1c]**

The searchers evaluate the relevance/usefulness of the (other) Candidate Terms. In some systems, searchers might be able to select the free text term initially entered instead of a Controlled Term. This leads to decision [1c] Assessment of suitability of other Candidate Terms.

2-7 **Decision [1a] to decision [1b]**

The searchers check whether the term found has the intended meaning. This leads to decision [1b] Assessment of the scope and meaning of a Candidate Term.

2-8 **Decision [1a] to entity (3)**

The searchers find the expected terms and decides to use it in the query.

2-9 **Decision [1b] to decision [1c]**

The term which appears to be appropriate does not have the expected scope and meaning. This leads to decision [1c] Assessment of suitability of other Candidate Terms.

2-10 **Decision [1b] to entity (3)**

The term has the expected scope and meaning and the searchers include it in the query.

2-11 **Decision [1c] to entity (3)**

The searchers select appropriate Candidate Terms for the query.

2-12 **Decision [1c] to entity (1a)**

The searchers return to the starting point for the session, (1a) Concept or Free Text term. The searchers could modify the entered phrase or think of a new one and recommence the mapping process.

3-1 **Entity (3) to entity (4)**

Some systems, especially those based on the Internet, allow searchers to execute queries by clicking a term, for example implemented as a hyperlink. Searchers are not involved in the query set-up and any potential modifications have to be applied after the execution of the query. In these systems, the search continues with entity (4) Query.

3-2 **Entity (3) to decision [2]**

The searchers add a controlled term to the query, for example by checking a box or dragging the term across.

4-1 **Decision [2] to entity (4)**

The searchers set up the query. Intermediate steps are not modelled individually in the model. Potential options are:

- Arranging terms – applying syntax
- Selecting a collection
- Setting query options such as language, document type....
- Setting term options, for example weights, expansion, etc.

All steps are interface and system dependent, i.e. not all are required, the system can take some of the decisions and impose a sequence in which they have to be selected.

4-2 **Entity (4) to entity (5)**

The searchers submit the query after the set-up. In a dynamic system, the execution is triggered by the modifications.

4-3 **Entity (4) to decision [2]**

This process allows for the fact that several options might have to be set.

4-4 **Decision [2] to decision [4]**

After making modifications to the query, for example by adding a term, the searchers return to the results set.

4-5 **Decision [2] to entity (6)**

If the searchers have taken a Controlled Term from a record, they can return to the record they were looking at after modifying the query by adding the Controlled Term.

4-6 **Decision [2] to decision [1a]**

After considering the query and its set-up, the searchers return to the Set of Candidate Terms to select more terms for the query. This does not necessarily have to happen immediately after the searchers have decided on a Candidate Term to use in the query.

4-7 **Decision [2] to entity (1)**

After considering the query and its set-up, the searchers return to entity (1) Starting point or reconsiders the task definition.

5-1 **Entity (5) to decision [3]**

The system retrieves records according to the submitted query and presents them to the searchers. This leads to decision [3] Assessment of the number of results retrieved.

5-2 **The outcome of decision [3] leads to the end of the search.**

The searchers terminate the search.

5-3 **Decision [3] to entity (1)**

The searchers return to entity (1) Starting point or reconsider the problem definition.

5-4 **Decision [3] to decision [4]**

The searchers consider the results retrieved by the query as a whole. This leads to decision [4] Assessment of the results retrieved.

5-5 **Decision [3] to entity (6)**

The searchers open a retrieved record in a random manner.

5-6 **Decision [3] to decision [2]**

After assessing the number of results, the searchers return to make modification to the query set-up (decision [2]).

5-7 **Decision [4] to entity (1)**

The assessment of the query results leads the searchers to return to entity (1) Starting point or reconsider the problem definition.

5-8 **The outcome of decision [4] leads to the end of the search.**

The searchers terminate the search.

5-9 **Decision [4] to entity (7)**

The searchers add useful or suitable records to the Collection of relevant records (entity (7).

5-10 **Decision [4] to entity (6)**

The assessment of the query results leads the searchers to decide on a particular record to inspect. This process might involve selecting or opening the record. In cases where complete records are shown, this process is understood as a shift in focus from the results set as a whole to the individual record.

5-11 Decision [4] to decision [2]

The assessment of the query results leads the searchers to make modification to the query set-up (decision [2]).

6-1 Entity (6) to decision [5]

Once a record has been selected, the searchers inspect this record in detail. Records can function as a means for generating Controlled Terms. Inspecting the record in detail leads to decision [5] Assessment of whether and how to use information from the record.

6-2 Decision [5] to decision [4]

The assessment of the aspects of the records leads the searchers to return to the record set to select a different record or to take any of the other options possible at decision [4] Assessment of the results retrieved.

6-3 Decision [5] to entity (7)

The assessment of the aspects of the records leads the searchers to add the record to the collection of relevant or useful records.

6-4 The outcome of decision [5] leads to the end of the search.

The searchers terminate the search.

6-5 Decision [5] to entity (8)

The searchers return to entity (1), Starting point or reconsider the problem definition.

6-6 Decision [5] to entity (3)

The searchers select a Controlled Term from the record.

6-7 Decision [5] to decision [2]

The assessment of the aspects of the records leads the searchers to return to make modification to the query set-up.

6-8 One instance of entity (6) to another

In some interfaces, searchers have the option to move from one record to the next without returning to the records set. This move does not require the searchers to make a decision on which record to view, as the order is generally predetermined by the system.

7-1 Entity (7) to decision [5]

From entity (7) Collection of relevant records, the searchers can return to the record they were inspecting in order to interact with other aspects, for example to select Controlled Terms, or to gather more information.

7-2 **Entity (7) to decision [4]**

From entity (7) Collection of relevant records, the searchers can return to decision [4], Assessment of the results retrieved, for example to choose another record to inspect.

7-3 **Entity (7) to entity (1b)**

The searchers can take a record from entity (7), their Collection of relevant records and use it as the basis of a search.

8-1 **Entity (8) to entity [5]**

After considering current search information extracted from a record, the searchers focus again on the aspects of the record in order to interact with other aspects, for example selecting Controlled Terms or gathering more information.

II)

Through consideration of current search information extracted from a record, the searchers can use this information generate new free text terms or concepts and effectively recommence at the beginning of the search cycle by mapping these new terms to the controlled vocabulary.

Appendix 11.1 Risk Points tables

RPIII	Type: Support/Instruction
Broad context	
General approaches to searching	
Model reference point	III – Permanent influence on searching behaviour
Condition - N/a	
Potential causes	
Lack of search training or practise leads to limited notions of how to search.	
Risks	
Searchers encounter difficulties in progressing with their search.	
Solutions/support options	
- Training (see 10.11.2) - Provision of information and support for the basic search process. Prompting can help novices to remember and benefit from functionality. - Integration of search stages into the interface so that following the correct sequence becomes the "default" behaviour. This can be reinforced through visual clues for the current stage, as they are used in online banking systems. Provision of functionality in itself does not ensure that the search is conducted efficiently, even if the potential is there, so the best general and other specific strategies to use the system should be reflected in the implementation. - Implementation of search strategies which correspond to particular tasks that are common in a particular environment will help users less familiar with searching to complete these tasks efficiently. Particular tools can also be developed to make completion of specific tasks easier, especially when they require more than one search to be executed. - The interface should as far as possible reflect domain-specific ways of expression but also allow non-experts to express themselves accordingly. Dates can for example be expressed in arrange of years or periods/eras depending on the subject area. The system can provide functionality that helps non-experts to identify the relevant categories and indexing terms.	
Limitations to implementation	
- The inclusion of many tools requires searchers to be aware of the options if they are to benefit from them. They can initially feel overwhelmed by the amount of information and the apparent complexity of the system. If information on functionality is to be introduced slowly, issues like identifying users and taking into account how much of what they have been told they still remember arise. - Computational limitations - Difficulties due to the complexity of the task	
Minimum information required	
Some basic signposts of how to progress through the search.	

Other potentially useful information
More detailed information on how to proceed and possible approaches to solving tasks.
Other Behaviour – N/a
Other factors which influence behaviour at this point
- Search experience and knowledge. - Similarities between other interfaces the searchers know and the interface which is used as well as transferability of familiar syntax, search strategies, tools etc. - Experience with the current interface. - Familiarity with the topic influences knowledge of potential terms and synonyms.

RP(1a)/II)	Type: Support/Instruction
Broad context	
General approaches to searching	
Model reference point	
(1a)/II) – Developing concepts to map to the controlled vocabulary	
Condition – N/a	
Potential causes – N/a	
Risks	
- Transfer from other interfaces and knowledge of the topic lead searchers to expect to find particular terms. - The searchers execute a search which is too specific for their vague Information Need. They might for example need to clarify the task before starting to search or execute a more general search to inform themselves about their topic first. - The searchers do not break the task down into appropriate chunks. - Searchers employ terms in unsuitable ways. In a hybrid system, some terms are more suitable for controlled vocabulary, others for free text searching. The later applies for example for personal names, if no authority file exists. - Searchers make assumptions about how candidate terms were generated and do not question them under the assumption that they must be appropriate.	
Solutions/support options	
- Wizard - Extracting information from data in the system - Searchers who are not familiar with indexing practises benefit from information on how to select terms accordingly, for example by the presentation of sample records.	
Limitations to implementation	
- Task complexity - Integration of sources of information used to clarify Information Need might not be possible. Help could be sought for example from an expert, somebody working in the environment of the search system (e.g. a library) or in reference material that is not available or too expensive in electronic form. - Searchers might not extract the appropriate information from sample records, as this	

also requires a certain amount of experience.
Minimum information required
Information on how to start searching, e.g. to access the mapping mechanism.
Other potentially useful information
Information on how the search system works.
Other Behaviour
- The searchers select terms from an existing record or query, etc. - The searchers browse the collection, for example starting from an example record. - The searchers browse the controlled vocabulary or knowledge structures to explore the representation of the domain in the particular system.
Other factors which influence behaviour at this point
Linked to the search stage, the extent to which searchers have thought about the query.

RP[0].1	**Type: Error recovery**
Broad context	
Identifying suitable controlled vocabulary terms	
Model reference point	[0] Number of candidate terms retrieved
Condition	
The number of candidate terms is 0 or insufficient.	
Potential causes	
- The searchers made spelling mistakes. - The searchers did not use the correct syntax, for example inserting separators between terms, incorrect spacing or have entered a phrase when only one word is allowed. The later is particularly an issue in systems which work with a high degree of post-co-ordination of terms. - Searchers have imposed (inappropriate) limitations on the matching mechanism.	
Risks	
The searchers become frustrated and no longer approach the task logically. They thus might not reach the point of actually executing a query.	
Solutions/support options	
- Searchers are referred back to the stage of identifying suitable concepts and free text terms. - Searchers are made aware of any imposed limitations which might have resulted in the exclusion of possible terms from the Set of Candidate Terms. - Automated modifications - Wizard	
Limitations to implementation	
- Computational limitations - Certain types of mistakes are difficult to detect, for example when searchers have confused two words or misspelled one so that it equals another one. - It can be frustrating for the searcher when the system retrieves irrelevant terms without an explanation.	

- If searchers are looking for concepts which cannot be presented using the controlled vocabulary or which are not in the database, no support is possible. - Some mechanisms might fail under certain circumstances, for example concepts which occur in the free text fields of a record are not necessarily translated into controlled vocabulary terms, so that examining these records does not necessarily lead to the identification of controlled terms. - If terms that do not retrieve suitable Candidate Terms are used as free text terms, searchers might not feel motivated to identify controlled terms.
Minimum information required
- The mechanism has finished searching for Candidate terms. - Candidate Terms retrieved (if the amount is larger than 0).
Other potentially useful information
- Additional information on the Candidate Terms retrieved. - Information on how Candidate Terms are retrieved. - Information on potentially available options. - Additional resources could potentially be integrated into the interface so that searchers can identify other options for terms. A reverse dictionary for example enables searchers to find words which describe their concepts appropriately.
Other Behaviour
- The searchers change the phrase to correct a mistake. - The searchers think of a synonym or another way of expressing a concept and search for that instead of a phrase which retrieved no results. - The searchers decide to deal with a different concept. - The searchers select a different access mechanism such as browsing the controlled vocabulary or the collection.
Other factors which influence behaviour at this point
- If the searchers have a relatively good knowledge of the topic, it is potentially easier for them to try alternative concepts or phrases. - If searchers are flexible in their task execution, they are possibly able to vary terms more than if they have stringent instructions.

RP[0].2	**Type: Support**
Broad context	
Identifying suitable controlled vocabulary terms	
Model reference point	[0] Number of candidate terms retrieved
Condition	
The number of candidate terms is too large.	
Potential causes	
The searchers have searched for words with are very common within the fields accessed by the search mechanism. During the FACET studies, this case occurred	

usually when searchers first did not retrieve enough terms and then decided to search scope notes as well. Likewise, the system might apply algorithms which use for example stemming, and which result in common character strings.
Risks
- Searchers impose limitations on the mechanism which reduce the amount of Candidate Terms retrieved but could prevent them from finding appropriate terms. - Selection difficulties
Solutions/support options
- Searchers can be informed about available limitations, but should be made aware that using them can result in no Candidate Terms to be retrieved or relevant Controlled Terms to be excluded. - Access to items - Wizard
Limitations to implementation
Computational limitations
Minimum information required
The Candidate Terms.
Other potentially useful information
- The Candidate Terms' local context in the controlled vocabulary. - Other information on the Candidate Terms such as scope notes. - An explanation on why the terms are retrieved. In the studies, searchers were particularly interested to know why there were so many results. - Postings or other information on which terms will actually retrieve results.
Other Behaviour
- Searchers go through the set until they have found an appropriate term in spite of the effort involved. - Searchers modify their entry into the search mechanism in a way that still retrieves the suitable terms, but leaves out many of the rest, for example by using synonyms or by expressing the concept in a different way.
Other factors which influence behaviour at this point
- If searchers have already tried other options for retrieving terms, they might be more inclined in giving up on either the particular concept or the task altogether. - If searchers are flexible in their task execution, they are possibly able to vary terms more than if they have stringent instructions.

RP[1]	**Type: Support/Facilitation**
Broad context	
Identifying suitable controlled vocabulary terms	
Model reference point	[1] Locating terms for the query amongst candidate terms.
Condition	**Ideal/Routine outcome**
The number of Candidate Terms is acceptable and the searchers attempt to identify	

suitable terms.
Risks
Selection difficulties
Solutions/support options
- Access to items. Quick access to additional information about the term, such as the scope note, can encourage searchers to check the meaning and scope before using a term. - Integration of additional resources, for example a dictionary or images, in the interface can help searchers identify other options for terms and distinguish between Candidate Terms. - Controlled vocabulary expansion can counterbalance non-optimal term selection. - Automated modifications - Wizard
Limitations to implementation
Difficulty in calculating postings in systems with expansion
Minimum information required
The candidate terms retrieved.
Other potentially useful information
- The candidate terms' local context in the controlled vocabulary. - Other information on the Candidate Terms such as scope notes. - Postings or other information on which terms will actually retrieve results.
Other Behaviour
- The searchers find the term they were looking for, check its meaning and add it to the query. - The searchers find suitable terms for their query, even if they are not similar to the ones they initially entered into the mechanism. - The searchers browse starting from Candidate Terms and identify more suitable terms in their proximity.
Other factors which influence behaviour at this point
- If searchers have used the system before, they possibly recognise useful terms while scanning the set of Candidate Terms and use those instead of the terms they initially thought of. - If searchers have some subject knowledge, they possibly recognise other terms that are alternatives to the ones they were initially looking for.

RP[2]	**Type: Support/Instruction**
Broad context	
Setting up the query	
Model reference point	[2] –Modifications required before first query execution
Condition	**Ideal/Routine outcome**
Searchers have selected at least one Controlled Term for the query.	
Risks	

- Selection of sub-optimal options - Searchers preconceptions about the way in which results are retrieved can influence their term selection. Assuming for example that the system uses Boolean operators, it is sensible not to use too many terms in case no results are retrieved, but the effect of using many query terms is not the same in all types of systems.
Solutions/support options
Default settings are such that the maximum amount of results is retrieved. If searchers require fewer results, the settings can be adjusted accordingly later. Experienced searchers who know the implications can change the settings immediately.
Limitations to implementation
At the point of query reformulation, it is possible to give more specific information regarding the query set-up, than when first constructing the query when all eventualities would have to be covered. (See RP[7]/[8])
Minimum information required
- Information on syntax, if any is required. - Information on how the query is submitted.
Other potentially useful information
- Information on options available. - Additional information on query execution which is relevant to the query set-up.
Other Behaviour
- The searchers set up their query in an appropriate way. - The searchers employ default options and maybe decide to change the parameters later.
Other factors which influence behaviour at this point – N/a

RP[7]/[8]	Type: Support/Facilitation
Broad context	
Query reformulation	
Model reference point	[7]/[8] Reflection on concepts/query terms/query options.
Condition	
The searchers have either executed a query or they are in the process of setting up the first query.	
Potential causes	
- The results of the current query do not fulfil the requirements of the task. - The searchers have come to think about whether the query as it is will fulfil the task requirements.	
Risks	
- Selection of sub-optimal options - The searchers believe that no appropriate records are in the database. - In some situations where few results are retrieved, lack of information prevents searchers or support mechanisms from identifying query modifications which could	

result in retrieval of different or additional records. - Selection difficulties of both query options and terms - The searchers narrow their topic down even though they are at a search stage where they have not enough knowledge about the topic to make these decisions. - Searchers worry about not having retrieved all relevant records.
Solutions/support options
- Extracting information from data in the system - Provision of general guidance on query reformulation and the available options. - Automated modifications
Limitations to implementation
Difficulties in choosing suggestions due to the complexity of the problem
Minimum information required
- Controlled vocabulary and other terms (if applicable) which are already in the query. - Results retrieved by the last-executed version of the query.
Other potentially useful information – N/a
Other Behaviour
The searchers identify the appropriate changes and improve their query.
Other factors which influence behaviour at this point
- Number of times searchers have already executed a query similar to the existing one. - The number of options searchers have already tried in order to improve their query. If they feel they have tried "everything", they probably feel less motivated to continue. - Flexibility the searchers have in selecting terms and concepts or in modifying them.

RP[3].1	**Type: Error recovery**
Broad context	
Assessing results retrieved	
Model reference point	[3] Number of result records retrieved
Condition	
No or too few results are retrieved.	
Potential causes	
- Limitations to the query set-up, for example on the collection searched, prevent searchers from retrieving more records. - No result records are associated with the query terms.	
Risks	
- The searchers believe that no appropriate records are in the database. - In some situations where few results are retrieved, lack of information prevents searchers or support mechanisms from identifying query modifications which could result in retrieval of different or additional records.	
Solutions/support options	
Automated modifications	
Minimum information required	
- The query has finished executing.	

- The query that was executed, as a visual check can sometimes reveal (spelling) mistakes or the fact that a different query had accidentally been executed.

Other potentially useful information

- Pointing out that the query can be reformulated.
- Options for changing the query.
- Information on how the query was executed, for example in a system using Boolean operators, showing how many records would be retrieved by each individual term.

Limitations to implementation

Computational limitations

Other Behaviour

The searchers make appropriate modifications, for example relaxing query limitations, and improve their query.

Other factors which influence behaviour at this point

- The number of times the searchers have previously executed the current query or a similar one.
- Certainty with which searchers have selected the query terms.
- Level of easy with which searchers were able to select query terms. If the selected controlled vocabulary terms were options amongst others, the searchers possibly find it easy to replace terms in the query. Otherwise, they are maybe not aware of other options.
- The searchers re-run a saved or existing query to monitor the new additions to the database collection and accept the fact that no new items have been added.

RP[3].2	**Type: Support**
Broad context	
Assessing results retrieved	
Model reference point	[3] Number of result records retrieved
Condition	
Too many or a very large number of results are retrieved.	
Potential causes	
- The query terms are common in the particular database. - High-level, broad terms are used with a large expansion rate so that a large amount of query terms is used. - The query settings include many records in the database or several databases.	
Risks	
- Selection difficulties - The searchers narrow their topic down even though they are at a search stage where they have not enough knowledge about the topic to make these decisions.	
Solutions/support options	
Access to items	
Limitations to implementation – N/a	
Minimum information required	

The query that was executed because a visual check can sometimes reveal (spelling) mistakes or the fact that a different query had accidentally been executed.
Other potentially useful information
- Pointing out that the query can be reformulated. - Options for changing the query. - Information on how the query was executed, for example in a system using Boolean operators, showing how many records would be retrieved by each individual term.
Other Behaviour
The searchers make appropriate modifications, for example changing query settings or terms, and improve their query.
Other factors which influence behaviour at this point
- The number of times the searchers have previously executed the current query or a similar one. - Generality/specificity of query terms. If the terms used are broad, searcher can potentially replace them with more specific terms. - The extent to which the searchers can make their queries more specific. If they require broad terms, other operational limitations such as imposing a publication date might be more appropriate than conceptual changes to the topic. - Detail of the controlled vocabulary. If the next-specific term is too specific or cannot be used for searching because it is a guide term, the searchers have limited possibilities to modify the terms.

RP[4].1	**Type: Error recovery**
Broad context	
Assessing the result records retrieved	
Model reference point	[4] Relevance of retrieved result records
Condition	
An appropriate amount of records has been retrieved, but the searchers judge the records as generally not suitable.	
Potential causes	
- Selection of an inappropriate term - Inappropriate indexing - The combination of query terms represents concepts other than those intended by the searcher.	
Risks	
The searcher makes (reasonable) modifications to the search which do not lead to better or different results.	
Solutions/support options	
- Searchers need to be made aware of the reasons for which unsuitable records have been retrieved. This could be done through displaying a list of suggestions, or be	

specifically based on the results retrieved (e.g. through an analysis of co-existing indexing terms).
- Access to items
- Automated modifications

Limitations to implementation

The information available in the controlled vocabulary or records can be too limited to be useful.

Minimum information required

Displaying information on the retrieved documents, document surrogates or records.

Other potentially useful information

- The match values that the system has calculated for the record, especially where this is part of the retrieval process (for example in a ranking system)
- Information on why the record is considered to be a match for the query, for example highlighting relevant terms in the record or representing the record in a visual way, for example using bars etc.
- Information on the scope of the query terms used if they stem from a controlled vocabulary to make searchers aware of inappropriate terms they might have used.

Other Behaviour

The searchers inspect results records and find reasons why their query has not retrieved more appropriate records. They reformulate the query by changing query settings or terms according to those reasons.

Other factors which influence behaviour at this point

- Searchers' ability to understand why some records are considered more relevant than others and how the match with the query comes about will make it easier for them to reformulate their queries.
- Quality of indexing, i.e. the detail and the correctness.
- Use of other systems might have lead searchers to establish what they consider to be a "good" match, for example in cases where a percentage is given as an indication of closeness to the query.
- Search experience or experience with the particular system.

RP[4].2	Type: Support
Broad context	
Assessing the result records retrieved	
Model reference point	[4] Relevance of retrieved result records
Condition	
An appropriate amount of records has been retrieved, but the searchers judge the records as not quite suitable or satisfactory.	
Potential causes	
- The searchers have taken an overall appropriate approach to formulating the query but it still needs to be adjusted to fulfil the task requirements more exactly.	

- The searchers think that more suitable records exist in the database.
Risks
Selection difficulties
Solutions/support options
- Provision of general guidance on query reformulation and the available options - Extracting information from data in the system - Automated modifications
Limitations to implementation – N/a
Minimum information required
Displaying information on the retrieved documents, document surrogates or records.
Other potentially useful information
- The match values that the system has calculated for the record, especially where this is part of the retrieval process (for example in a ranking system) - Information on why the record is considered to be a match for the query, for example highlighting relevant terms in the record or representing the record in a visual way, for example using bars etc. - Information on the scope of the query terms used if they stem from a controlled vocabulary to make searchers aware of inappropriate terms they might have used.
Other Behaviour
- The searchers inspect results records and find reasons why their query has not retrieved more appropriate records. They reformulate the query by changing query settings or terms according to those reasons. - The searchers select more suitable records for the Collection of relevant records.
Other factors which influence behaviour at this point – N/a

RP[4].3	**Type: Support**
Broad context	
Assessing the result records retrieved	
Model reference point	[4] Relevance of retrieved result records
Condition	**Ideal/Routine outcome**
An appropriate amount of records has been retrieved, and the searchers judge the records as generally suitable.	
Risks	
- Selection difficulties - Searchers worry about not having retrieved all relevant records.	
Solutions/support options	
- Access to items - Being able to save the query allows searchers to monitor new records when added to the database or similar functionality. Execution of the query could also be done automatically when the database is updated and the searchers are informed when new records are found.	
Limitations to implementation	

- The information available in the controlled vocabulary or records can be too limited to be useful.
- Computational limitations
- Difficulties in choosing suggestions due to the complexity of the problem

Other Behaviour

The searchers reformulate the query in order to make sure they have found all relevant records.

Minimum information required

Displaying information on the retrieved documents, document surrogates or records.

Other potentially useful information

- The match values that the system has calculated for the record, especially where this is part of the retrieval process (for example in a ranking system)
- Information on why the record is considered to be a match for the query, for example highlighting relevant terms in the record or representing the record in a visual way, for example using bars etc.
- Information on the scope of the query terms used if they stem from a controlled vocabulary to make searchers aware of inappropriate terms they might have used.
- Records already included in the Collection of relevant results can be highlighted or identified in some form, possibly even if they were selected during a previous search.
- Information on when or how often the database is updated allows searchers to execute the search again when new records are available.

Other Behaviour

- The searchers select more suitable records for the Collection of relevant records.
- The searchers reformulate the query by changing either terms or settings.
- The searchers decide that they have found enough records and stop searching.

Other factors which influence behaviour at this point

- Searchers might have executed several searches and identified selected records as the most suitable ones they will be able to retrieve in the current data collection.
- If the same search has been executed before the last database update, searchers are likely to be particularly interested in the new records.
- If a similar query has been executed before, the searchers might already have added some of the records to the Collection of relevant results.

RP[5]	Type: Support
Broad context	
Inspecting individual records	
Model reference point	[5] Aspects of a record to interact with
Condition	**Ideal/Routine outcome**
The searchers have selected a record from the results set which is of interest to them.	
Risks	

- Selection of sub-optimal options - Especially in interfaces where browsing from records is facilitated, searchers might take up a new lead without resolving the initial task.
Solutions/support options – N/a
Limitations to implementation – N/a
Minimum information required
Minimum information on the record.
Other potentially useful information
- The more information is accessible, the easier it will be for the searchers to make an informed decision on which records to select. Terms can for example be hyperlinked so that the searchers can easily return to the controlled vocabulary. - Within the records, more detailed information on how the record matches the query can be displayed. - Information in records can also help searchers in understanding the indexing and how to modify their query accordingly.
Other Behaviour
- The searchers add terms from the records to the query. - By inspecting the records, the searchers learn about the topic, which in turn influences the Current Search Information. - The searchers select more suitable records for the Collection of relevant records. - The searchers decide that they have found enough records and stop searching.
Other factors which influence behaviour at this point
If the searchers have seen the record previously, they might require less time to assess aspects of the record.

References

References which are also available online have been last accessed on the 21/06/2004.

AITCHISON, J., A. GILCHRIST and D. BAWDEN (2000), *Thesaurus construction: a practical manual.* AsLib, London.

BARRETT, B. and M. MATICKA (1989), An analysis of user failure in subject searching in an online catalogue. *Proceedings of Garbage in, Garbage out - The Need for Quality in the Age of Automation. 8th National Cataloguing Conference*, Adelaide, Australia, Auslib Press, 38-49.

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

BATES, M. J. (1977), System meets user: problems in matching subject search terms. *Information Processing & Management.* 13(6): 367-375.

BATES, M. J. (1979a), Idea tactics. *Journal of the American Society for Information Science.* 30(5): 280-289.

BATES, M. J. (1979b), Information search tactics. *Journal of the American Society for Information Science.* 30(4): 205-214.

BATES, M. J. (1986), Subject access in Online Catalogs - a design-model. *Journal of the American Society for Information Science.* 37(6): 357-376.

BATES, M. J. (1994), The design of databases and other information resources for humanities scholars: The Getty Online Searching Project Report No. 4. *Online & CDROM review.* 18(6): 331-340.

BATES, M. J. (2002), After the dot-bomb: Getting Web Information Retrieval right this time. *First Monday.* 7(7). http://firstmonday.dk/issues/issue7_7/bates/index.html.

BATES, M. J., D. N. WILDE and S. SIEGFRIED (1993), An analysis of search terminology used by Humanities scholars: The Getty Online Searching Project Report No 1. *Library Quarterly.* 63(1): 1-39.

BATES, M. J., D. N. WILDE and S. SIEGFRIED (1995), Research practices of Humanities scholars in on online environment: The Getty Online Searching Project Report No 3. *Library & Information Science Research.* 17(1): 5-40.

BATTY, D. (1998), WWW - Wealth, weariness and waste. *D-Lib magazine.* 4(11). http://www.dlib.org/dlib/november98/11batty.html.

BEAULIEU, M. (1997), Experiments on interfaces to support query expansion. *Journal of Documentation.* 53(1): 8-19.

BEAULIEU, M., T. DO, A. PAYNE and S. JONES (1997), Enquire OKAPI Project. *British Library and Innovation Report 17.*

BELKIN, N. J., C. COOL, A. STEIN and U. THIEL (1995), Cases, Scripts and information seeking strategies - on the design of interactive IR systems. *Expert systems with applications.* 9(3): 379-395.

BEYNON-DAVIES, P., D. TUDHOPE, C. TAYLOR and C. JONES (1994), A semantic database approach to knowledge-based hypermedia systems. *Information and Software Technology.* 36(6): 323-329.

BINDING, C. AND D. TUDHOPE (2004), KOS at your service: Programmatic access to Knowledge Organisation Systems. *Journal of Digital Information.* 4(4) Article No. 265, 2004-02-05. http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/.

BLOCKS D., C. BINDING, D. CUNLIFFE and D. TUDHOPE (2002), Qualitative evaluation of a thesaurus-based retrieval system. *Proceedings 6th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2002)*, Rome. Eds: M. Agosti, C. Thanos. Lecture Notes in Computer Science, Berlin: Springer. 346-361.

BORGMAN, C. L. (1996), Why are online catalogs still hard to use? *Journal of the American Society for Information Science.* 47(7): 493-503.

BORGMAN, C. L., S. G. HIRSH and J. HILLER (1996), Rethinking online monitoring methods for Information Retrieval systems: From search product to search process. *Journal of the American Society for Information Science.* 47(7): 568-583.

BORLUND, P. and P. INGWERSEN (1997), The development of a method for the evaluation of interactive Information Retrieval systems. *Journal of Documentation.* 53(3): 225-250.

BOYCE, B. R., C. T. MEADOW and D. H. KRAFT (1994). *Measurement in Information Retrieval.* San Diego, Academic Press.

BRANCH, J. L. (2000), The trouble with think alouds: Generating data using concurrent Verbal Protocols. *Proceedings of CAIS 2000: Dimensions of a global information science,* School of Library & Information Studies, University of Alberta, Edmonton, Canada. Canadian Association for Information Science, 28. http://www.slis.ualberta.ca/cais2000/branch.htm.

BRAJNIK, G., S. MIZZARO and C. TASSO (1996), Evaluating user interfaces to Information Retrieval systems: A case study on user support. *Proceedings of SIGIR '96,* Zurich, Switzerland, ACM, 8: 128-136.

CARROLL, J. M. (2000). *Making use - scenario-based design of human-computer interactions.* Cambridge, MA, USA, MIT Press.

CHAMIS, A. Y. (1991), *Vocabulary control and search strategies in online searching.* New York, Greenwood Press.

CHAPMAN, J. L. (1981), A state transition analysis of online information-seeking behavior. *Journal of the American Society for Information Science.* 32(5): 325-333.

CHIN, J. P., V. A. DIEHL and K. L. NORMAN (1988), Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of CHI '88,* ACM, 213-218.

CHOO, C. W., B. DETLOR and D. TURNBULL (2000), Information seeking on the web: an integrated model of browsing and searching. *First Monday.* 5(2). http://www.firstmonday.dk/issues/issue5_2/choo/index.html#author.

CUNLIFFE, D., C. TAYLOR and D. TUDHOPE (1997), Query-based navigation in semantically indexed hypermedia. *Proceedings of the 5th ACM conference on Hypertext (Hypertext '97),* Southampton, UK, ACM, 8: 87-95.

DERVIN, B. (1999), On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing & Management.* 35(6): 727-750.

DICKSON, J. (1984), An analysis of user errors in searching an online catalog. *Cataloging and Classification Quarterly.* 4(3): 19-38.

DIX, A., J. FINLAY, G. ABOWD and R. BEALE (1993). *Human-computer interaction.* New York, Prentice-Hall.

DOERR, M. (2001), Semantic Problems of Thesaurus Mapping. *Journal of Digital Information.* 1(8). http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/.

ELLIS, D. (1989), A behavioural model for Information Retrieval systems design. *Journal of Information Science.* 15(4&5): 237-247.

ELLIS, D. (1996), The dilemma of measurement in Information Retrieval research. *Journal of the American Society for Information Science.* 47(1): 23-36.

FIDEL, R. (1985), Moves in online searching. Online Review. 9(1):61-74

FIDEL, R. (1987), What is missing in research about online search behaviour? *Canadian Journal of Information Science.* 12(3): 54-61.

FIDEL, R. (1990), Online searching styles. *Proceedings of the 53rd American Society for Information Science (ASIS) Annual Meeting.* 27: 98-103.

FIDEL, R. (1991a), Searchers' selection of search keys: I. The selection Routine. *Journal of the American Society for Information Science.* 42(7): 490-500.

FIDEL, R. (1991b), Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science.* 42(7): 501-514.

FIDEL, R. (1991c), Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science.* 42(7): 515-527.

FIDEL, R. (1993), Qualitative Methods in Information-Retrieval Research. *Library & Information Science Research.* 15(3): 219-247.

FIDEL, R. (1994), User-centered indexing. *Journal of the American Society for Information Science.* 45(8): 572-576.

FIDEL, R. and E. N. EFTHIMIADIS (1995), Terminological Knowledge Structure for intermediary expert systems. *Information Processing & Management.* 31(1): 15-27.

FORD, N., F. WOOD and C. WALSH (1994), Cognitive Styles and Searching. *Online & CDROM Review.* 18(2): 79-86.

FOWKES, H. and M. BEAULIEU (2000), Interactive Searching behaviour: OKAPI experiment for TREC-8. *Proceedings of IRSG 2000 - 22nd Annual Colloquium on Information Retrieval Research*, Cambridge, UK.

GONÇALVES, M. A., M. LUO, R. SHEN, M. F. ALI and E. A. FOX (2002), An XML log standard and tool for digital library logging analysis. *Proceedings of the 6th European Conference on Digital Libraries (ECDL 2002) (ECDL 2002)*, Rome. Eds: M. Agosti, C. Thanos. Lecture Notes in Computer Science, Berlin: Springer. 129-143.

HAAS, D. and D. KRAFT (1984), Experimental and quasi-experimental designs for research in information science. *Information Processing and Management.* 20(1-2): 229-237.

HANCOCK-BEAULIEU, M. (1990), Evaluating the impact of an Online Library Catalog on subject searching behavior at the catalog and at the shelves. *Journal of Documentation.* 46(4): 318-338.

HANCOCK-BEAULIEU, M., M. FIELDHOUSE and T. DO (1995), Evaluation of Interactive Query Expansion in an online library catalog with a graphical user interface. *Journal of Documentation.* 51(3): 225-243.

HANCOCK-BEAULIEU, M. and S. WALKER (1992), An evaluation of Automatic Query Expansion in an online library catalog. *Journal of Documentation.* 48(4): 406-421.

HANSEN, P. (1998), Evaluation of Information Retrieval user interface - implications for user interface design. *Human IT.* 2(2). http://www.hb.se/bhs/ith/2-98/ph.htm.

HARKNESS CONNELL, T. (1995), Subject searching in online catalogs: metaknowledge used by experienced searchers. *Journal of the American Society for Information Science. 46(7):506-518.*

HAWKINS, D. T. (1996), Hunting, grazing, browsing: a model for online Information Retrieval. *Online.* 20(1): 71-73. http://www.infotoday.com/online/JanOL/hawkins.html.

HEARST, M. A. (1999), User interfaces and visualization. *Modern Information Retrieval.* Eds.: R. Baeza-Yates and B. Ribeiro-Neto. New York, Addison Wesley/ACM: 257-323.

HEARST, M., A. ELLIOTT, J. ENGLISH, R. SINHA, K. SWEARINGEN and K.-P. YEE (2002), Finding the flow in web site search. *Communications of the ACM.* 45(9): 42-49.

HEINECKE, A. M. (1995), Evaluation of hypermedia systems in museums. *Proceedings of Hands on Hypermedia and Interactivity in Museums - 3rd International Conference on Hypermedia and Interaction in Museums (ICHIM'95- MCN '95)*, San Diego, CA, 67-78.

HERMAN, I., G. MELANCON and M. S. MARSHALL (2000), Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics.* 6(1): 24-43.

HILL, L. L. L. CARVER, M. LARSGAARD, R. DOLIN, T. R. SMITH, J. FREW and M.-A. RAE (2000), Alexandria Digital Library: User evaluation studies and system design. *Journal of the American Society for Information Science.* 51(3): 246-259.

HODGE, G. (2000), Systems of Knowledge Organization for Digital Libraries: Beyond traditional authority files. Report number 1-887334-76-9. http://www.clir.org/pubs/abstract/pub91abst.html.

HUANG, M. L., P. EADES and J. WANG (1998), On-line animated visualization of huge graphs using a modified spring algorithm. *Journal of Visual Languages & Computing.* 9(6): 623-645.

IIVONEN, M. and D. H. SONNENWALD (1998), From translation to navigation of different discourses: A model of search term selection during the pre-online stage of the search process. *Journal of the American Society for Information Science.* 49(4): 312-326.

INGWERSEN, P. (1996), Cognitive perspectives of Information Retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation.* 52(1): 3-50.

JOHNSON, E. H. and P. A. COCHRANE (1995), A Hypertextual Interface for a Searcher's Thesaurus. *Proceedings of Digital Libraries '95*, Austin, Texas, USA. http://csdl.tamu.edu/DL95/papers/johncoch/johncoch.html.

JONES, S., M. GATFORD, T. DO and S. WALKER (1997), Transaction logging. *Journal of Documentation.* 53(1): 35-50.

KENNEDY, L., C. COLE and S. CARTER (1999), The false focus in online searching - The particular case of undergraduates seeking information for course assignments in the humanities and social sciences. *Reference & User Services Quarterly*. 38(3): 267-273.

KOENEMANN, J. and N. J. BELKIN (1996), A Case for Interaction: a study of interactive Information Retrieval behavior and effectiveness. *Proceedings of CHI 96 "Common Ground" - Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, Canada, ACM. http://www.acm.org/sigchi/chi96/proceedings/papers/Koenemann/jk1_txt.htm.

KONTOGIANNIS, T. (1999), User strategies in recovering from errors in man–machine systems. *Safety Science*. 32(1): 49-68.

KUHLTHAU, C. C. (1988), Developing a model of the library search process: cognitive and affective aspects. *RQ*. 28(2): 232-242.

KUHLTHAU, C. C. (1991), Inside the search process - Information seeking from the users perspective. *Journal of the American Society for Information Science*. 42(5): 361-371.

LAZONDER, A. W. and H. VAN DER MEIJ (1993), The minimal manual - is less really more? *International Journal of Man-Machine Studies*. 39(5): 729-752.

LECKIE, G., K. PETTIGREW and C. SYLVAIN (1996), Modelling the information seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers. *Library Quarterly*. 66(2): 161-193.

LYKKE NIELSEN, M. (2001), A framework for work task based thesaurus design. *Journal of Documentation*. 57(6): 774-797.

MANGLANO, V., M. BEAULIEU and S. ROBERTSON (1998), Evaluation of Interfaces for IRS: Modelling End-User Searching Behaviour. *Proceedings of the 20th Annual BCS-IRSG Colloquium on IR*, BCS. http://www.bcs.org/ewic/.

MARCHIONINI, G. (1995). *Information seeking in electronic environments*. Cambridge, Cambridge University Press.

MEADOW, C. T., G. MARCHIONINI and J. M. CHERRY (1994), Speculations on the measurement and use of user characteristics in information-retrieval experimentation. *Canadian Journal of Information and Library Science - Revue Canadienne des Sciences de l'Information et de Bibliotheconomie*. 19(4): 1-22.

NAHL, D. and C. TENOPIR (1996), Affective and cognitive searching behavior of novice end-users of a full-text database. *Journal of the American Society for Information Science*. 47(4): 276-286.

NICHOLAS, D. (1996), An assessment of the online searching behaviour of practitioner end-users. *Journal of Documentation*. 52(3): 227-251.

PATERNO, F. and C. MANCINI (1999), Designing web user interfaces adaptable to different types of use. *Proceedings of Museums and the Web 1999*, Minnesota. http://www.archimuse.com/mw99/abstracts/prg_57.html.

PENNIMAN, W. D. (1982), Modeling and evaluation of Online user behavior. *Proceedings of the Annual Meeting of the American Society for Information Science*. Asis. 19: 231-235.

POLLITT, A. S. (1997), Interactive Information Retrieval based on faceted classification using Views. *Proceedings of the 6th International Study Conference on Classification: Knowledge Organisation for Information Retrieval*, University College London. http://poseidon.hud.ac.uk/external/research/groups/cedar/dorking.htm.

POLLITT, A. S., M. P. SMITH and P. A. J. BRAEKEVELT (1996), View-based searching systems - a new paradigm for IR based upon faceted classification and indexing using mutually constraining knowledge-based views. *Proceedings of the Joint Workshop of the Information Retrieval and Human Computer Interaction Specialist Groups of the British Computer Society*, Glasgow University, 73-77. http://poseidon.hud.ac.uk/external/research/groups/cedar/bcshci.htm.

PORS, N. O. (2000), Information Retrieval, experimental models and statistical analysis. *Journal of Documentation*. 56(1): 55-70.

RADA, R., H. MILI, E. BICKNELL and M. BLETTNER (1989), Development and application of a metric on Semantic Nets. *IEEE Transactions on Systems, Man and Cybernetics*. 19(1): 17-30.

ROBERTSON, G. G., S.K. CARD and J. D. MACKINLAY (1993), Information Visualization using 3D interactive animation. *Communications of the ACM*. 36(4): 57-71.

ROBERTSON, S. E. and M. M. HANCOCK-BEAULIEU (1992), On the evaluation of Information Retrieval systems. *Information Processing and Management.* 28(4): 457-466.

ROBINS, D. (2000), Shifts of focus on various aspects of user information problems during interactive Information Retrieval. *Journal of the American Society for Information Science and Technology.* 51(10):913-928.

RUTHVEN, I. (2003a), Incorporating user search behavior into relevance feedback. *Journal of the American Society for Information Science and Technology.* 54(6): 529-549.

RUTHVEN, I. (2003b), Re-examining the potential effectiveness of Interactive Query Expansion. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in Information Retrieval*, Toronto, Canada, ACM, 213-220.

SALTON, G. and M. J. MCGILL (1983). *Introduction to modern Information Retrieval*, McGraw-Hill Inc.

SARACEVIC, T. (1996a), Modelling interaction in Information Retrieval (IR): A review and a proposal. *Proceedings of the Annual Meeting of the American Society for Information Science*, ASIS, 33: 3-9.

SARACEVIC, T. (1996b), Relevance reconsidered. *Proceedings of CoLIS 2. Second International Conference on Conceptions of Library and information Science: Integration in Perspective*, Copenhagen, Denmark, The Royal School of Librarianship, 201-218.

SARACEVIC, T. (1997), The stratified model of Information Retrieval interaction: Extension and applications. *Proceedings of the Annual Meeting of the American Society for Information Science*, ASIS, 34: 313-327.

SAVAGE-KNEPSHIELD, P. A. and N. J. BELKIN (1999), Interaction in Information Retrieval: Trends over time. *Journal of the American Society for Information Science.* 50(12): 1067-1082.

SCHAMBER, L. (2000), Time-line interviews and inductive content analysis: Their effectiveness for exploring cognitive behaviour. *Journal of the American Society for Information Science.* 51(8): 734-744.

SHAW, D. (1986), 9 sources of problems for novice online searchers. *Online Review.* 10(5): 295-303.

SHNEIDERMAN, B. (1998). *Designing the user interface - Strategies for effective human-computer interaction*, Addison-Wesley.

SEBRECHTS, M. M., J. VASILAKIS, M. S. MILLER, J. V. CUGINI and S. J. LASKOWSKI (1999), Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces. *Proceedings of 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, California, ACM, 3-10.

SHIRI, A. A. and C. REVIE (2000), Thesauri on the Web: current developments and trends. *Online Information Review.* 24(2): 273-280.

SHIRI, A. A. and C. REVIE (2001), User-thesaurus interaction in a web-based database: An evaluation of users' term selection behaviour. *Proceedings of the Infotech Oulu International Workshop on Information Retrieval*, Oulu, Finland. 23-32.

SHUTE, S. J. and P. J. SMITH (1993), Knowledge-based search tactics. *Information Processing & Management.* 29(1): 29-45.

SIATRI, R. (1999), The evolution of user studies. *Libri.* 49(3): 132-141.

SIEGFRIED, S., M. J. BATES and D. N. WILDE (1993), A profile of end-user searching behavior by humanities scholars - The Getty Online Searching Project Report No 2. *Journal of the American Society for Information Science.* 44(5): 273-291.

SMITHSON, S. (1994), Information-Retrieval evaluation in practice - a case-study approach. *Information Processing & Management.* 30(2): 205-221.

SPINK, A. (1997), Information science: A third feedback framework. *Journal of the American Society For Information Science.* 48(8): 728-740.

SPINK, A., N. FORD, T. D. WILSON, A. FOSTER and D. ELLIS (2002), Information seeking and mediated searching. Part 3: Successive searching. *Journal of the American Society for Information Science and Technology.* 54(9): 716–727.

SPINK, A., A. GOODRUM and D. ROBINS (1995), Search intermediary elicitations during mediated online searching. *Proceedings of CAIS/ACSI 95 Annual Conference of the Canadian Association for Information Science: Connectedness: Information, systems, people, organizations*, University of Alberta, Edmonton, Canada. http://www.ualberta.ca/dept/slis/cais/spink.htm.

SPINK, A. and T. SARACEVIC (1997), Interaction in Information Retrieval: selection and effectiveness of search terms. *Journal of the American Society for Information Science.* 48(8): 741-761.

STOREY, M.-A., F. D. FRACCIA and H. A. MUELLER (1999), Customizing a fisheye view algorithm to preserve the mental map. *Journal of Visual Languages and Computing.* 10: 245-267.

TAGUE, J. and R. SCHULTZ (1989), Evaluation of the user interface in an Information-Retrieval system - a model. *Information Processing & Management.* 25(4): 377-389.

TAGUE-SUTCLIFFE, J. (1992), The pragmatics of Information-Retrieval experimentation, revisited. *Information Processing & Management.* 28(4): 467-490.

TAMLER, H. (1998), How (much) to intervene in a usability testing session. *Common Ground.* 8(3): 11-15. http://www.htamler.com/papers/intervene/.

TAYLOR, R. S. (1968), Question-negotiation and information seeking in libraries. *College & Research Libraries.* 29: 178-194.

TER HOFSTEDE, A. H. M., H. A. PROPER and T. P. VAN DER WEIDE (1996), Query formulation as an Information Retrieval problem. *Computer Journal.* 39(4): 255-274.

TONTA, Y. (1992), Analysis of search failures in document retrieval systems: a review. *The Public-Access Computer Systems Review.* 3(1): 4-53. http://info.lib.uh.edu/pr/v3/n1/tonta.3n1.

TUDHOPE, D., C. BINDING, D. BLOCKS and D. CUNLIFFE (2002), Compound Descriptors in context: A matching function for search thesauri. *Proceedings of the Joint Conference on Digital Libraries 2002*, Portland, USA, ACM Press.

TUDHOPE, D. and C. TAYLOR (1997), "Navigation via similarity: Automatic linking based on semantic closeness." *Information Processing & Management.* 33(2): 233-242.

VAKKARI, P. (1999), Task complexity, problem structure and information actions - Integrating studies on information seeking and retrieval. *Information Processing & Management.* 35(6): 819-837.

WIEDENBECK, S., P. L. ZILA and D. S. MCCONNELL (1995), End-user training: An empirical study comparing on-line practice methods. *Proceedings of CHI'95*, ACM. http://www.acm.org/sigchi/chi95/proceedings/papers/sw_bdy.htm.

WILDEMUTH, B., R. DEBLIEK and C. P. FRIEDMAN (1993), Measures of searcher performance - A psychometric evaluation. *Information Processing & Management.* 29(5): 18

WILDEMUTH, B. M. and M. E. MOORE (1995), End-user search behaviors and their relationship to search effectiveness. *Bulletin of the Medical Library Association.* 83(3): 294-304.

WILSON, T. D. (1997), Information behaviour: An interdisciplinary perspective. *Information Processing & Management.* 33(4): 551-572.

WILSON, T. D. (1999), Models in information behaviour research. *Journal of Documentation.* 55(3): 249-270.

XIE, H. I. (2002), Patterns between interactive intentions and information-seeking strategies. *Information Processing & Management.* 38(1): 55-77.

YEE, K.-P., K. SWEARINGEN, K. LI and M. HEARST (2003), Faceted metadata for image search and browsing. *Proceedings of CHI 2003.* http://bailando.sims.berkeley.edu/papers/flamenco-chi03.pdf.

YUAN, W. J. (1997), End-user searching behavior in Information Retrieval: A longitudinal study. *Journal of the American Society for Information Science.* 48(3): 218-234.