Qualitative evaluation of thesaurus-based retrieval

D. Blocks, C. Binding, D. Cunliffe, D. Tudhope

Hypermedia Research Unit, School of Computing, University of Glamorgan,
Pontypridd CF37 1DL, UK
Email: {dblocks; cbinding; djcunlif; dstudhope}@glam.ac.uk
Telephone: +44 1443 482271
Fax: +44 1443 482715

Abstract. This paper reports on a formative evaluation of a prototype thesaurus-based retrieval system, which involved qualitative investigation of user search behaviour. The work is part of the ongoing 'FACET' project in collaboration with the National Museum of Science and Industry and its collections database. The main thesaurus employed in the project is the Getty Art and Architecture Thesaurus. The aim of the evaluation is to analyse at a micro level the user's interaction with interface elements in order to illuminate problems and inform interface design decisions. Data gathered included transcripts of think-aloud sessions, screen capture movie files, user action logs and observator notes. Key incidents from the sessions are analysed and the qualitative methodology is discussed. The evaluation analysis informs design issues concerning the allocation of search functionality to sub-windows, the appropriate role of thesaurus browsing in the search process, the formation of faceted queries and query reformulation. The analysis suggests that, although the prototype interface supports basic level operations, it does not provide non-expert searchers with sufficient guidance on query structure and when to use the thesaurus. Conclusions are drawn that future work should further support and suggest models of the search process to the user.

Forthcoming, *Proceedings (ECDL 2002) 6th European Conference on Research and Advanced Technology for Digital Libraries, (M. Agosti, C. Thanos eds.),* Lecture Notes in Computer Science, Berlin: Springer, 2000. http://www.springer.de/comp/lncs/index.html

Qualitative evaluation of thesaurus-based retrieval

D. Blocks, C. Binding, D. Cunliffe, D. Tudhope

Hypermedia Research Unit, School of Computing, University of Glamorgan,
Pontypridd CF37 1DL, UK
Email: {dblocks; cbinding; djcunlif; dstudhope}@glam.ac.uk
Telephone: +44 1443 482271
Fax: +44 1443 482715

Abstract. This paper reports on a formative evaluation of a prototype thesaurus-based retrieval system, which involved qualitative investigation of user search behaviour. The work is part of the ongoing 'FACET' project in collaboration with the National Museum of Science and Industry and its collections database. The main thesaurus employed in the project is the Getty Art and Architecture Thesaurus. The aim of the evaluation is to analyse at a micro level the user's interaction with interface elements in order to illuminate problems and inform interface design decisions. Data gathered included transcripts of think-aloud sessions, screen capture movie files, user action logs and observator notes. Key incidents from the sessions are analysed and the qualitative methodology is discussed. The evaluation analysis informs design issues concerning the allocation of search functionality to sub-windows, the appropriate role of thesaurus browsing in the search process, the formation of faceted queries and query reformulation. The analysis suggests that, although the prototype interface supports basic level operations, it does not provide non-expert searchers with sufficient guidance on query structure and when to use the thesaurus. Conclusions are drawn that future work should further support and suggest models of the search process to the user.

1. Introduction

As Digital Libraries and subject gateways grow, it becomes increasingly important to structure access for end-users to avoid the problems associated with Web search engines and their lack of terminology control. One useful tool in this quest is the thesaurus [1]. Thesauri are used in Library, Museum and Archive contexts as both indexing and search tools. While thesauri can also be used to expand free text search queries, in this paper we are concerned with a controlled vocabulary application. Users select search terms from a thesaurus which has been used to index the collection. This approach avoids false hits and provides users with a model of the domain which can facilitate term selection [2, 3], although they may need assistance to identify terms which are relevant and effective in retrieval [4]. Various studies have investigated issues concerning the role of the thesaurus in information seeking but there have been few empirical evaluations of thesaurus use in search systems by end-users. This paper reports on a formative evaluation of a prototype thesaurus-based retrieval system with semantic term expansion and the paper also discusses the qualitative methodology employed to study user search behaviour.

Fidel has investigated thesaurus use by professional searchers to construct knowledge bases for IR systems which could inform term selection of less experienced searchers [5-7]. Iivonen and Sonnenwald [8] conducted investigations into term selection in end-user-intermediary searching. They postulated that the processes should not simply be seen as a substitution of user terms for thesaurus terms, but that different ways of viewing and representing topics contribute to the selection of terms. Pollitt's HIBROWSE [9] assists users with faceted thesauri and other knowledge patterns in medicine. The user selects terms from hierarchical menus which correspond to various facets, e.g. "Physical disease" and "Therapy". Postings of documents that are indexed with all these terms are shown. For each facet, a new window with a list of narrower terms is opened, so that the number of results can progressively be reduced. The semantic links in a thesaurus can also be used in query expansion. Previous work at Glamorgan has investigated the potential of associative and spatial relationships [10]. The Okapi project has looked at the balance between system and user as to term selection. After automatic and interactive versions, they designed the Enquire interface which combines the two approaches [11]. Selecting from suggested terms was found to be difficult for users as they lacked an understanding of the impact terms might have on their query (see also[12]).

In this paper, we present key results from the formative evaluation and discuss their impact on design decisions. Quantitative evaluation methods can prove useful in analysing system performance and user attitudes identified via questionnaires [13]. However, the general aim of this study was to use qualitative methods to pinpoint problems at different search stages in order to better integrate the thesaurus into the search process. This is important as many potential users have little or no training in online searching and the use of complex tools. Making these tools available will however become more important as Digital Library collections grow and demand for more precise searching increases. The study was part of a larger project (FACET), being a formative, evaluation of a working prototype of the system. The qualitative evaluation methodology was informed by a previous study of the work practices of commercial software developers which followed an ethnographic approach [14]. Ethnographic techniques can also be employed in studies of information seeking behaviour [e.g. 15]. Although the time scale of our formative evaluation did not permit a longitudinal study, we wished to adapt elements of an ethnographic approach, working with potential users in their own settings, at realistic tasks (within the constraints of an experimental prototype) and, crucially, capturing rich detail of the fleeting events of user sessions. Data gathered included transcripts of think-aloud sessions, screen capture movie files, user action logs and participant observer notes. Combining several data gathering methods allows triangulation between them. The aim was to analyse at a micro level the user's interaction with interface elements and reasoning in order to illuminate tacit sources of problems and inform iterative interface design decisions. This requires representations of session events that combine the different modalities of data capture.

2. The Facet Project

FACET is an experimental system [16] which is being developed at the University of Glamorgan in collaboration with the National Museum of Science and Industry (NMSI) which includes the National Railway Museum [17]. The research project investigates the possibilities of term expansion in faceted thesauri based on measures of semantic closeness. The main thesaurus employed in the project is the Art and Architecture Thesaurus (AAT) developed by the J. Paul Getty Trust [18]. With a view to making their collections more accessible to the public, NMSI is indexing parts of its collections database using the AAT. An export of these records is used as the underlying dataset for the system (some 400,000 records). The evaluation focused on the Railway Furniture and Railway Timepieces collections, although other collections are available.

The AAT is a faceted thesaurus; concepts are organised into a small number of high level, exclusive, fundamental categories or 'facets': *Associated Concepts, Physical Attributes, Styles & Periods, Agents, Activities, Materials, Objects* (with optional facets for *Time* and *Place*). As opposed to attempting to include all possible multi-concept headings explicitly in the thesaurus (e.g. *painted oak furniture)*, they are synthesised as needed by combining terms from different facets [19]. This can be performed either by the cataloguer when indexing or the searcher when constructing queries. One of the goals of the FACET project is to investigate matching between multi-concept headings that takes advantage of semantic term expansion. The semantic relationships between thesaurus terms provide a system with information on how similar two terms are by calculating the traversal cost from the number and type of traversals necessary to move from one term to the other [e.g. 10, 20, 21, 22]. Various possibilities need to be considered, e.g. the absence of a query term from the indexing terms of a record, the presence of further indexing terms, exact matches of terms and partial matches of a query terms with semantically close indexing terms. For a more detailed discussion of these issues and how FACET deals with them, refer to [19]. Thus, it is not necessary for the users to construct a query that exactly matches a record's set of index terms or to exhaustively browse trying different combinations of terms to achieve the exact match. FACET calculates a match value between a query term and an index term depending on traversal cost. This in turn feeds into a matching function that produces ranked results, including partial matches, from a multi-term query. The system is implemented in C++ with a Visual Basic interface using a SQL Server database. An in-memory representation of the network of thesaurus relationships permits an efficient term expansion algorithm. The focus of this paper, however, is on the formative evaluation of the prototype interface and the role of the thesaurus in that interface.

Figures 1 and 2 illustrate the initial prototype of the interface, which was modified as a result of the first evaluation session conducted in this study as discussed below. This initial interface contains a number of elements, assigned to separate windows or subpanes including a Thesaurus Browser for users to browse through the hierarchies of the thesaurus using a mouse and a string search facility (Term Finder) that attempts to map an initial string to vocabulary terms (Figure 1). All terms can be dragged into the Query window. A ranked list of summaries appears in a results pane and details can be viewed. The record opens in a new window in the foreground and shows the full description and the indexing terms (Figure 2). Double-clicking on any term will open the Thesaurus Browser, centred on that term.

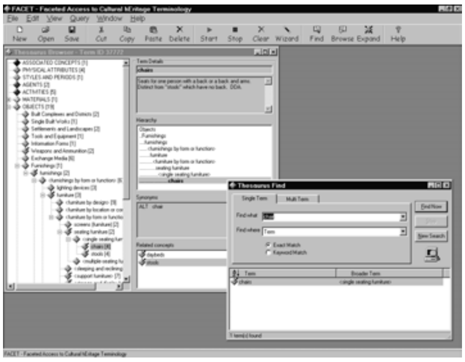


Figure 1 Thesaurus Browser and Term Finder

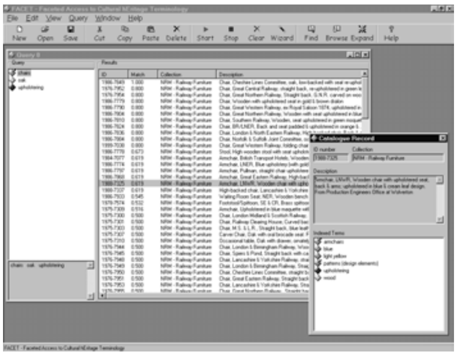

FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Window Help
New Open Save Cut Copy Paste Delete Start Stop Clear Wizard Find Browse Expand Help
Query
chairs
oak
upholstering
Results
ID
Match
Collection
Description
chairs oak upholstering
Catalogue Record
ID number
Collection
NRM - Railway Furniture
Description
Armchair, LNWR, Wooden chair with upholstered seat, back & arms; upholstered in blue & cream leaf design. From Production Engineers Office at Wolverton.
Indexed Terms
armchairs
blue
light yellow
patterns (design elements)
upholstering
wood
FACET - Faceted Access to Cultural hEritage Terminology

Figure 2 Query window and results record

3. Methodology

Researchers in the fields of HCI, library and information science use a number of qualitative and quantitative data collection and analysis methods (see [23-26] for overviews). In this study, we were interested in users' reasoning in addition to the interaction sequences, in order to have a better insight into motivation and to reveal potential problems. This requires a dialog between evaluator and participants during and/or after the search sessions. This dialog and other interactions need to be recorded in detail for thorough analysis. One drawback of working with people's thoughts is that users need to be encouraged to "think aloud" so their thoughts can be recorded. Some people find this difficult and if they are silent it is not always easy for the evaluator to ask them about their actions without giving the impression of criticism (e.g.[27]).

Although a recording of conversation itself can normally be followed without problems, it is not always possible to identify the incident or precise interface element the speakers refer to. It is therefore desirable to have a comprehensive history of interactions. The FACET system records critical user and system actions in a log, which is further processed to be readable. Screen capture tools proved useful to remind the evaluator of the events of the session. We used Camtasia [28] which creates movie files as a visual record of the session. In future evaluations, playback of incidents may also be used as a basis for discussion with participants. The main evaluator took notes after sessions to record any information not otherwise captured, e.g. uncertainty on the part of the user or evaluator or observations of particularly notable events for later analysis. For some sessions, notes from additional observers are also available.

Over the summer of 2001, six museum professionals from various institutions, one IT and one library professional from University of Glamorgan participated in the study. The evaluators travelled to participants, taking a laptop with the experimental system. Thus each session took place in an environment familiar to the participants. Before the sessions, they were given a demonstration of the interface including an explanation of the term expansion mechanism. The main evaluator then sat with each participant at a table so she could observe the events on the laptop screen. The participants were given a training scenario with step-by-step instructions. It covered all aspects of the interface that they needed to know to complete the tasks. After the demonstration, three users decided they did not need the training.

A second handout outlined the tasks. The session started with a focused warm-up task that required users to *search for a record similar to the one printed out.* They could thus simply search the thesaurus for terms from this record and drag them into the query. The second task required users to *find objects decorated with text*. The main challenge here was to move from terms in the question to more suitable thesaurus terms. The third task was more open and was designed to require the user to identify thesaurus terms which were not included in the task description. Participants were given a picture of a chair, and asked to *find a matching record*. One problem with this task was that the records did not include images, and the chair in the picture was not from the collection. However, matching records existed in the collection, and participants were successful in their search. The last task asked users to *generally explore the interface*. Not all users completed this task due to time restrictions.

The set-up was similar but not exactly the same for each session. Minor modifications were made to logging and the prompt strategy to encourage think-aloud. Occasionally, users felt uncomfortable with the audio recording, so none was taken during these sessions. Travelling to different locations meant the environment varied, e.g. it was not always possible to conduct the evaluation session in a separate room. Sometimes additional evaluators observed the session. The participants also had different levels of knowledge of the project. However, the differences were noted and considered in the data analysis.

The sessions totalled eight hours. Data for analysis include transcripts of audio tapes, screen capture files, the log created by the application and a number of sets of observer notes. The log files were post-processed both automatically and manually for clarity of presentation. The tapes were transcribed word for word in the large part, although sections not relating to the evaluation or interface were summarised. The evaluator watched the screen capture files and generated descriptions of events. Transcripts of tapes and screen capture files were collated together with log data and comments from the notes. This resulted in a rich set of data for each session. The notation used to describe the incidents selected for discussion clearly identifies the contribution of each data source:

Descriptive summary of screen capture files	Log file (post-processed)
`Transcript of audio recording`	Post-hoc evaluator notes

Initially, the tape transcripts were the primary source of interest, as they explicitly and implicitly revealed problematic situations. These were identified and analysed in more detail, e.g. by examining the log and the screen recording. Other incidents analysed included those the evaluator noted during the sessions and when looking at screen capture files. For each important or problematic incident, the evaluator considered the different data sources from the session. Juxtaposing incidents allowed further identification of possible reasons behind problems. Lower level interface issues were identified, however this paper focuses on issues particularly relevant to the search process.

4. Selected Key Incidents from the Evaluation

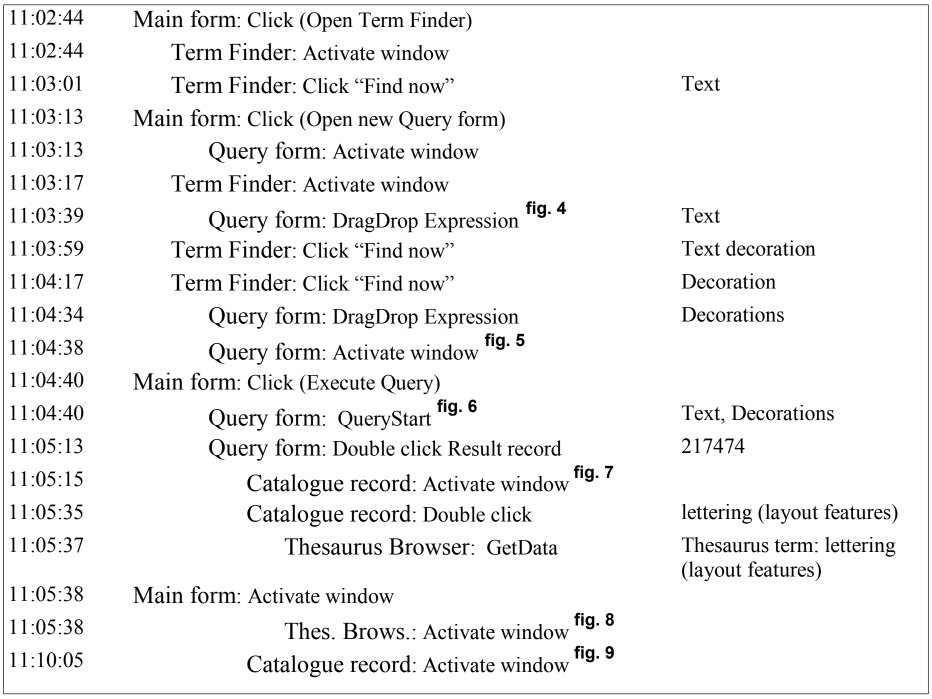
4.1 Window Switching

At the interface level, one initial issue was switching between windows. The initial prototype comprised a number of individual windows for searching the thesaurus, browsing it, constructing the query and viewing the results (Figures 1 & 2). This profusion of separate windows caused problems in the first evaluation session. In the example shown, the participant is trying to find objects decorated with text (Task 2). Figure 3 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 user interacts very little with forms before changing to another form. Note that the user has to return to the background window or main form to execute the query (e.g. 11:04:40). The user activates windows (e.g. 11:04:38) simply to move them. Screen dumps of the original screen capture file provide a visual representation of the events (Figures 4-9).

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. 4). 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. 5) and executes the query by clicking the "Start button" on the main form. The results appear (fig. 6) and the user double-clicks one to bring up the record (fig. 7). On the record form, he then double-clicks the term "Lettering (layout features)" which brings up the Thesaurus Browser (fig. 8). Note that 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. 9 and around 11:10:05). (Descriptive summary of screen capture)

As a consequence of this first evaluation, the interface was modified to attempt to reduce window context switching. Underlying these problems appears to be a lack of intrinsic order in the user's progress through the search. The user has 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 different windows. The user cannot anticipate easily which windows will follow, continually having to move/adjust existing windows as new ones appear.

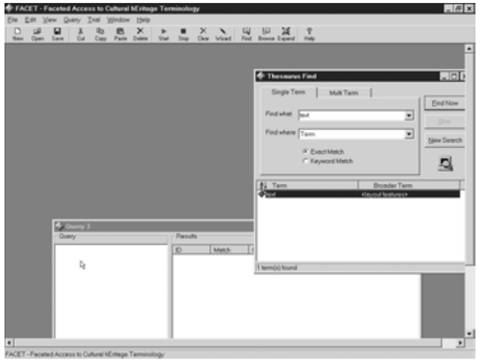
In the revised interface used for the subsequent sessions in this evaluation, Term Finder and Thesaurus Browser are integrated into one window using tabs. The Browser can be seen in Figure 10. The query is no longer in a separate window but forms the background of the interface (with query terms on the left and results on the right). These modifications were an attempt to reflect the logic of users' search behaviour better and to reinforce stages of search behaviour as identified e.g. by Kuhlthau [29]. The use of overlapping windows remained an issue though, e.g. if the Thesaurus Browser window was not minimised before executing the search, it covers the results.



```
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. 4      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. 5
11:04:40   Main form: Click (Execute Query)
11:04:40         Query form:  QueryStart fig. 6              Text, Decorations
11:05:13         Query form: Double click Result record      217474
11:05:15            Catalogue record: Activate window fig. 7
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. 8
11:10:05            Catalogue record: Activate window fig. 9
```

Figure 3 Extract from the post-processed log

Figures 4 – 9 Windowing issues

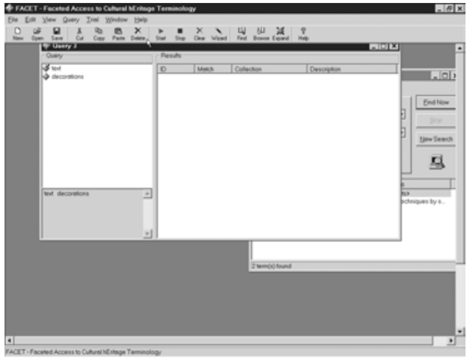

FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Trial Window Help
New Open Save Cut Copy Paste Delete Start Stop Clear Wizard Find Browse Expand Help
Thesaurus Find
Single Term
Multi Term
Find Now
Stop
New Search
Query 3
Query
Results
ID
Match
FACET - Faceted Access to Cultural hEritage Terminology

Find what: text
Find where: Term
● Exact Match
○ Keyword Match

Term	Broader Term
text	<layout features>

1 term(s) found

Figure 4

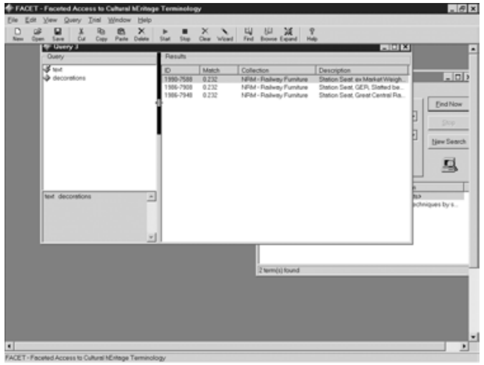

FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Trial Window Help
New Open Save Cut Copy Paste Delete Start Stop Clear Wizard Find Browse Expand Help
Query 3
Query
Results
ID
Match
Collection
Description
Find Now
Stop
New Search
FACET - Faceted Access to Cultural hEritage Terminology

text
decorations

text decorations

2 term(s) found

Figure 5


FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Trial Window Help
New Open Save Cut Copy Paste Delete Start Stop Clear Wizard Find Browse Expand Help
Query 3
Query
Results
Find Now
Stop
New Search
FACET - Faceted Access to Cultural hEritage Terminology

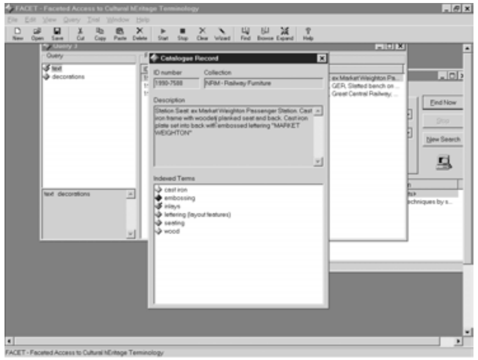
text
decorations

text decorations

ID	Match	Collection	Description
1990-7588	0.232	NRM - Railway Furniture	Station Seat, ex Market Weigh...
1986-7908	0.232	NRM - Railway Furniture	Station Seat, GER, Slatted be...
1986-7948	0.232	NRM - Railway Furniture	Station Seat, Great Central Ra...

2 term(s) found

Figure 6


FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Trial Window Help
New Open Save Cut Copy Paste Delete Start Stop Clear Wizard Find Browse Expand Help
Query 3
Query
Catalogue Record
Find Now
Stop
New Search
FACET - Faceted Access to Cultural hEritage Terminology

text
decorations

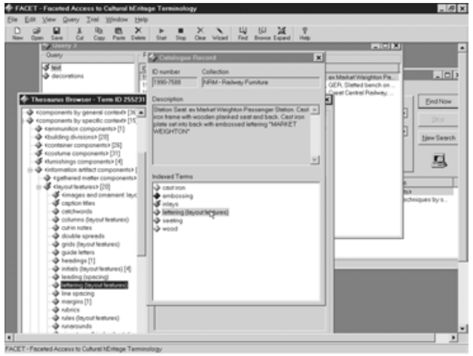
text decorations

ID number: 1990-7588
Collection: NRM - Railway Furniture

Description: Station Seat ex Market Weighton Passenger Station. Cast iron frame with wooden planked seat and back. Cast iron plate set into back with embossed lettering "MARKET WEIGHTON"

Indexed Terms
- cast iron
- embossing
- inlays
- lettering (layout features)
- seating
- wood

Figure 7


FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Trial Window Help
New Open Save Cut Copy Paste Delete Start Stop Clear Wizard Find Browse Expand Help
Query 3
Query
Catalogue Record
Thesaurus Browser - Term ID 255231
Find Now
Stop
New Search
FACET - Faceted Access to Cultural hEritage Terminology

text
decorations

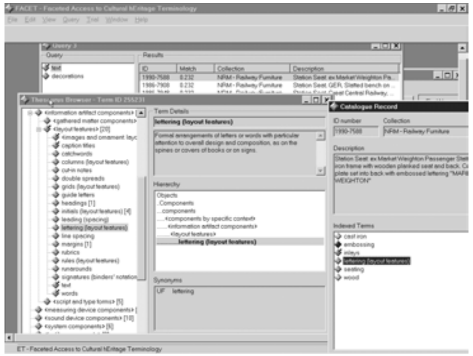
<components by general context> [36
<components by specific context> [15
<ammunition components> [1]
<building divisions> [20]
<container components> [26]
<costume components> [31]
<furnishings components> [4]
<information artifact components> [
<gathered matter components>
<layout features> [20]
<images and ornament: layo
caption titles
catchwords
columns (layout features)
cut-in notes
double spreads
grids (layout features)
guide letters
headings [1]
initials (layout features) [4]
leading (spacing)
lettering (layout features)
line spacing
margins [1]
rubrics
rules (layout features)
runarounds

ID number: 1990-7588
Collection: NRM - Railway Furniture

Description: Station Seat ex Market Weighton Passenger Station. Cast iron frame with wooden planked seat and back. Cast iron plate set into back with embossed lettering "MARKET WEIGHTON"

Indexed Terms
- cast iron
- embossing
- inlays
- lettering (layout features)
- seating
- wood

Figure 8


FACET - Faceted Access to Cultural hEritage Terminology
File Edit View Query Trial Window Help
Query 3
Query
Results
Thesaurus Browser - Term ID 255231
Term Details
Hierarchy
Synonyms
Catalogue Record
ET - Faceted Access to Cultural hEritage Terminology

text
decorations

ID	Match	Collection	Description
1990-7588	0.232	NRM - Railway Furniture	Station Seat ex Market Weighton Pa...
1986-7908	0.232	NRM - Railway Furniture	Station Seat, GER, Slatted bench on ...
[illegible]	[illegible]	[illegible]	Station Foot, Great Central Railway, ...

<information artifact components> [
<gathered matter components>
<layout features> [20]
<images and ornament: layo
caption titles
catchwords
columns (layout features)
cut-in notes
double spreads
grids (layout features)
guide letters
headings [1]
initials (layout features) [4]
leading (spacing)
lettering (layout features)
line spacing
margins [1]
rubrics
rules (layout features)
runarounds
signatures (binders' notation
text
words
<script and type forms> [5]
<measuring device components> [
<sound device components> [10]
<system components> [6]

Term Details: lettering (layout features)

Formal arrangements of letters or words with particular attention to overall design and composition, as on the spines or covers of books or on signs.

Hierarchy
Objects
..Components
....components
......<components by specific context>
........<information artifact components>
..........<layout features>
............lettering (layout features)

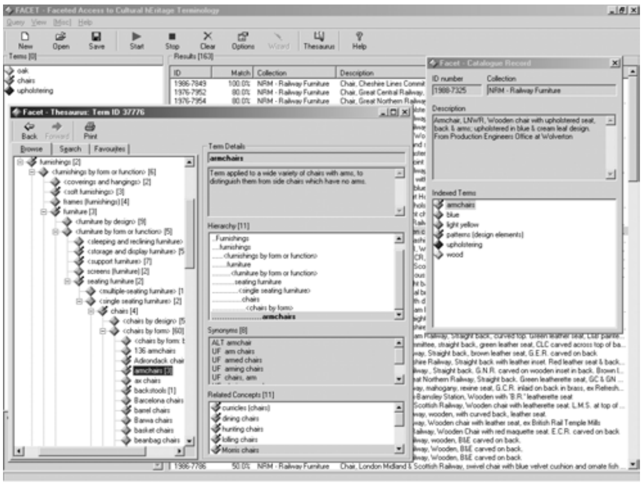
Synonyms
UF lettering

ID number: 1990-7588
Collection: NRM - Railway Furniture

Description: Station Seat ex Market Weighton Passenger Stati iron frame with wooden planked seat and back. Ca plate set into back with embossed lettering "MARI WEIGHTON"

Indexed Terms
- cast iron
- embossing
- inlays
- lettering (layout features)
- seating
- wood

Figure 9



FACET - Faceted Access to Cultural hEritage Terminology

Terms [0]

- oak
- chairs
- upholstering

Results [163]

ID	Match	Collection	Description
1986-7849	100.0%	NRM - Railway Furniture	Chair, Cheshire Lines Commi
1976-7952	80.0%	NRM - Railway Furniture	Chair, Great Central Railway,
1976-7954	80.0%	NRM - Railway Furniture	Chair, Great Northern Railwa
1986-7786	50.0%	NRM - Railway Furniture	Chair, London Midland & Scottish Railway, swivel chair with blue velvet cushion and ornate fish ...

```
am Railway, Straight back, curved top. Green leather seat, LSB painte...
mittee, straight back, green leather seat, CLC carved across top of ba...
way, Straight back, brown leather seat, G.E.R. carved on back
shire Railway, Straight back with leather inset. Red leather seat & back...
ilway., Straight back, G.N.R. carved on wooden inset in back, Brown l...
at Northern Railway, Straight back. Green leatherette seat, GC & GN ...
way, mahogany, rexine seat, G.C.R. inlaid on back in brass, ex Refresh...
-Barnsley Station, Wooden with 'B.R.' leatherette seat
Scottish Railway, Wooden chair with leatherette seat. L.M.S. at top of ...
way, wooden, with curved back, leather seat
way, Wooden chair with leather seat, ex British Rail Temple Mills
ailway, Wooden Chair with red maquette seat. E.C.R. carved on back
lway, wooden, B&E carved on back
lway, Wooden, B&E carved on back
lway, Wooden, B&E carved on back
```

Facet - Thesaurus: Term ID 37776

```
furnishings [2]
  <furnishings by form or function> [6]
    <coverings and hangings> [2]
    <soft furnishings> [3]
    frames (furnishings) [4]
    furniture [3]
      <furniture by design> [9]
      <furniture by form or function> [5]
        <sleeping and reclining furniture>
        <storage and display furniture> [5
        <support furniture> [7]
        screens (furniture) [2]
        seating furniture [2]
          <multiple-seating furniture> [1
          <single seating furniture> [2]
            chairs [4]
              <chairs by design> [5
              <chairs by form> [60]
                <chairs by form: t
                136 armchairs
                Adirondack chair
                armchairs [3]
                ax chairs
                backstools [1]
                Barcelona chairs
                barrel chairs
                Barwa chairs
                basket chairs
                beanbag chairs
```

Term Details

armchairs

Term applied to a wide variety of chairs with arms, to distinguish them from side chairs which have no arms.

Hierarchy [11]

```
..Furnishings
....furnishings
......<furnishings by form or function>
........furniture
..........<furniture by form or function>
............seating furniture
..............<single seating furniture>
................chairs
..................<chairs by form>
....................armchairs
```

Synonyms [8]

```
ALT armchair
UF  arm chairs
UF  armed chairs
UF  arming chairs
UF  chairs, arm
```

Related Concepts [11]

- curricles (chairs)
- dining chairs
- hunting chairs
- lolling chairs
- Morris chairs

Facet - Catalogue Record

ID number: 1988-7325

Collection: NRM - Railway Furniture

Description: Armchair, LNWR, Wooden chair with upholstered seat, back & arms; upholstered in blue & cream leaf design. From Production Engineers Office at Wolverton

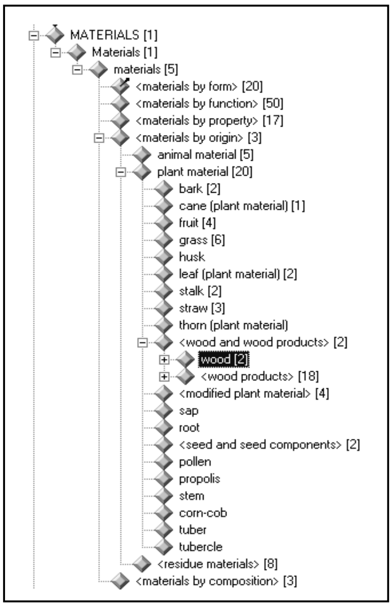
Indexed Terms

- armchairs
- blue
- light yellow
- patterns (design elements)
- upholstering
- wood

Figure 10 Revised interface: Thesaurus Browser (left) and record in front of query

4.2 Browsing behaviour

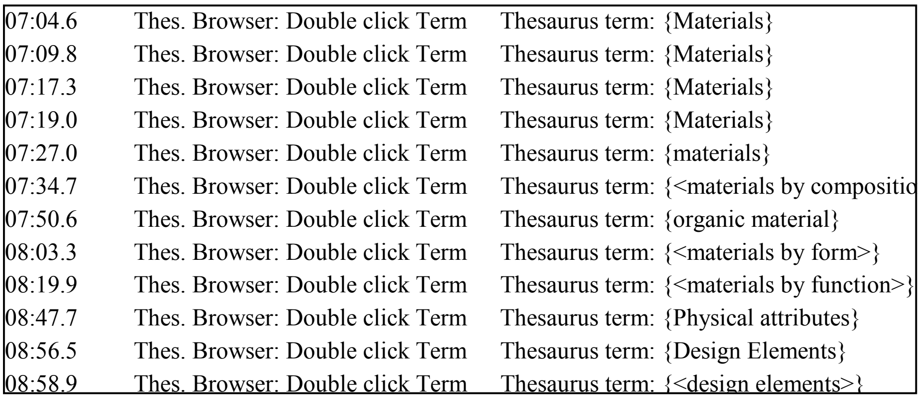
This example from a later session demonstrates use of the Thesaurus Browser. While searching for a chair matching the picture (Task 3), the participant browsed the thesaurus from the root attempting to find “Wood”. It is unclear precisely what motivated the user to opt for this approach rather than initially 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 to a search for “Wood” in the thesaurus, was hidden. The participant was unable to locate the term because it was unclear 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 the terms, the more abstract they are. This makes it difficult for users, especially those who are not familiar with the AAT, to distinguish between them. Guide terms can also be difficult for the non-expert to interpret. Figure 11 shows the path (six levels deep) that leads to “Wood” in the Thesaurus Browser.



```
MATERIALS [1]
  Materials [1]
    materials [5]
      <materials by form> [20]
      <materials by function> [50]
      <materials by property> [17]
      <materials by origin> [3]
        animal material [5]
        plant material [20]
          bark [2]
          cane (plant material) [1]
          fruit [4]
          grass [6]
          husk
          leaf (plant material) [2]
          stalk [2]
          straw [3]
          thorn (plant material)
          <wood and wood products> [2]
            wood [2]
            <wood products> [18]
          <modified plant material> [4]
          sap
          root
          <seed and seed components> [2]
          pollen
          propolis
          stem
          corn-cob
          tuber
          tubercle
        <residue materials> [8]
      <materials by composition> [3]
```

Figure 11 “Wood” in the hierarchy

Figure 12 shows a log excerpt of this incident. The first three clicks are actually on the same term. The user initially clicked twice so that the hierarchy contracted and had to be expanded again. Note the user 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 user thought of searching the thesaurus with Term Finder and instantly found “Wood”.



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

Figure 12 Browsing from the root to find the term “Wood”

The following example stands in contrast to this unprofitable thesaurus browsing. For the final open task (exploration of the interface), the participant first searched for a term and then inspected its local context in the Thesaurus Browser.

The user brings up two results for “Trains“ in the Term Finder – “Trains (vehicle groupings)” and “Trains (costume components)”. The user double-clicks “Trains (vehicle groupings)” and clicks the “Browse” tab to look at the term’s local context. The user scrolls down a bit, then double-clicks “Trains (vehicle groupings)” and then one of its more specific terms, “Passenger trains”. It has one more specific term, “High speed trains”. The user then drags “Passenger trains” into the query, deletes the terms from the previous query and executes it.

The main difference between this browsing and the incident previously described is that the user’s entry point into the thesaurus is much lower. Within three clicks, the user reaches the bottom of the hierarchy. The term is six levels away from the Facet (root) level – as is “Wood”. Finding this seemingly obvious term by browsing down from the root would probably have been just as difficult. For example, the hierarchy to select after “Objects” would have been “Object groupings and systems”. While a user may sometimes browse to explore the structure of a thesaurus, this incident suggests an insight into the role of thesaurus browsing when part of a specific query. It seems fair to assume that thesaurus browsing to find a particular term is useful as long as the entry point is low enough so that users navigate within the term’s local context and do not face highly abstract categories near the root. Not only is the cognitive load for the user reduced, but the process is faster (particularly so for large thesauri such as the AAT). While the browsing in the first example took almost 2 minutes and did not result in a term selection, the user in the latter example found a term within 20 seconds. These examples show that browsing the local context of a term can be useful to fine-tune the query and reassure the user that they are on the right track. The contrast between the incidents also suggests that the interface provides the right functionality, but perhaps still does not sufficiently model the query process, given that the problematic browsing occurred.

4.3 Breaking a Task down into Concepts

For faceted retrieval, users have to identify individual query components corresponding to thesaurus terms. One participant wanted to find a wooden table in the database (Task 1). He first searched the Term Finder for "Tables" and then browsed the local context of this term in the thesaurus in an attempt to locate a compound term for the query.

> Participant: So you want a table to go with the chair… right… so you know that… okay, take this. So … you can just go through and if you see … Tables by … Just trying to think. It obviously tells you … "Tables by design". But I was looking for "Tables by materials". Because you know you are trying to match it with an oak chair.

We see that he does not fully break the task down into concepts. He should look for "Wood" (or even "Oak") and "Tables" separately as these are two different concepts represented by different terms in different facets ("Materials" and "Objects" respectively). This misunderstanding also occurred in some other sessions. Following this study, the next version of the interface will provide more support in faceted query formulation by matching terms entered into the Term Finder individually and in combination and by a more structured Query Builder.

4.4 Reformulation of Query

Task 2 required the users to find items decorated with text. The challenge was to move away from the terms in the task description to more suitable thesaurus terms. One participant had a particularly persistent difficulty in that he reformulated the query repeatedly, but the same results were returned each time. A collated extract from different data sources is shown in Figure 13. It is not completely clear why this session was unsuccessful in reformulating the query, although it partly depends on the terms employed. Two closely related terms (e.g. "Text" and "Words") cover each other through the expansion algorithm, so there is no need to include them both 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 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, 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? … Participant: So I can now try that and see if it will get me anything more.

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

54:40.9 Query form: Toolbar Start Query: Text, Words, Lettering

54:46.5 Query form: Results: 3

The same results as above now have a 50.0% match. The user deletes "text" and re-runs query. He then also removes "words" The results remain the same with 66.7% and 100% matches respectively.

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] 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 – now "decorations" and "lettering".

P: Right, so I could try that …

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 term 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 13 Collated data of unsuccessful query reformulation

The term “Decorations” did not improve the query due to a misunderstanding. The user 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 (Figure 13; 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 applied this knowledge in other situations. Providing more context in Term Finder, e.g. the broader term, might have helped the user 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 user employed reasonable, advisable strategies to modify the query. It was not the strategies themselves that led to unsuccessful results. The results would indeed have fulfilled the task requirements. That the user *wanted* to find more results and was unable to do so constitutes the problem. This situation is especially problematic as the local context of the terms did not lead to other appropriate terms, such as “Inscriptions”, which was very effective in retrieval. Other users made the leap to this term, or searched for objects, such as plaques or posters which they imagined could be decorated with text.

Mechanisms to support moves similar to that from “Text” to “Inscriptions” would be complex, but users could generally speaking be encouraged by appropriate prompts to search for objects that may have useful terms. The nature of the data collections implies this approach, and users commented accordingly when asked about the authenticity of the tasks. However, it cannot be denied that users could still commence a search with inappropriate terms from which they cannot easily move on to more suitable ones without additional help.

5. Conclusions

Collecting data from various sources was valuable, despite the fact that transcribing the tape and analysing the screen capture files was very time consuming. The collated representation (e.g. Figure 13) together with screen dumps combine the different sources. To some extent, such representations also serve to provide a record of the basis for reported findings. (We hope to include relevant screen capture video files with the electronic version of the proceedings). Triangulation of different data sources provides a fuller picture of user’s reasoning and interactions. Users’ interactions on a micro level could be put in the context of their reasoning and the specific features of the interface. As discussed in Section 4, the rich picture provided by the qualitative evaluation data of user sessions served to inform analysis of important issues, including the allocation of search functionality to sub-windows, the appropriate role of thesaurus browsing in the search process, the formation of faceted queries and query reformulation. Additional lower level interface issues were noted but lack of space prevents discussion here. The findings discussed are currently contributing to a further iteration of the user interface.

Certain aspects of the methodology still need to be resolved however. Not all participants are comfortable with audio recording, so an alternative method of collecting data needs to be developed for these cases. A type of shorthand notation might be suitable to capture as much detailed information as possible while conducting the evaluation and observing the user. The evaluator's prompts to encourage thinking-aloud need to be developed further. Alternatively, a session set-up with two collaborating users might encourage them to confer verbally on their steps and the individual might feel less under observation. The evaluator has to reassure users and give them appropriate pointers when they are unsure of how to proceed.

Looking at the problems that occurred during the sessions, it becomes apparent that despite some minor issues, interaction techniques used by FACET are successful in allowing a person with little knowledge of the interface to use the functionality. However, constructing a good query that returns satisfactory results is more challenging. It is during the search process itself that most users encounter problems of a conceptual nature. These range from breaking down the query into concepts that match thesaurus terms, to improving results through repeated reformulation. The analysis suggests that the prototype interface does not provide non-expert searchers with sufficient guidance as to query structure and when to use the thesaurus in the search process. The potential of the thesaurus and term expansion mechanism is also not explicit enough. We are currently working on a model of thesaurus informed searching which includes potential methods to resolve common problems. Users often have a choice of approaches, e.g. searching or browsing the thesaurus. At times, both options might be equally valid, but at other times use of the thesaurus should be channeled so that users employ techniques at search stages where they benefit most from them. This would for example mean discouraging browsing the thesaurus from the root level to find a term on a relatively low level or including a number of closely related terms in a query. Approaches which could assist users include more initial training and providing initial template queries and results. This would provide information on indexing practice. Users could modify the given queries according to their requirements so that they do not start with a blank sheet. A search wizard would be another method of channeling search activities in (likely) productive ways. An initial user profile might help determine which search technique would be the most appropriate to follow. In the long run, more information on the stages and options of the search process could be integrated into the interface.

However, problems in achieving results can arise even when people demonstrate textbook searching behaviour. In these cases, users might require more background information on how and why results are retrieved. In one example described above, the user performed reasonable interactions, but the results would not expand. The user was not aware of other approaches such as modifying the term expansion ratio because this feature was not intended for the participants' use in this interface version. Particularly in versions for non-specialists, the system might provide more active support and possibly automatic query refinement. Further options for thesaurus integration and support thus remain to be explored. Query histories would allow comparisons of query versions and give an opportunity to easily return to the best result set. Users could have the opportunity to modify query options, e.g. the term expansion ratio. This would be helpful in fine-tuning queries, although the cognitive load would also increase.

The next version of the interface under development will facilitate query formulation. An integrated Query Builder tool, which combines searching/browsing with query formulation and maintains the top level facets, will better reflect the search process and the thesaurus as a source of terms for the query. This presentation should assist users in breaking down their query into faceted components and in forming a better mental model of the thesaurus. More feedback will be available on the effects of semantic term expansion on the query and on query results (why records were retrieved). This should allow users to establish a better understanding of how to construct and reformulate a query according to their priorities. Thus, it is intended that the next version of the system will further support and suggest models of the search process to the user.

6. Acknowledgements

We would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council (grant GR/M66233/01) and the support of HEFCW for the Internet Technologies Research Lab. We would like to thank Geoff Elliot and Traugott Koch for helpful comments; the J. Paul Getty Trust for provision of their vocabularies; the NMSI for provision of their collection data; and the participants in the study for their time, effort and input.

References

1. Aitchison, J., A. Gilchrist, and D. Bawden. *Thesaurus construction: a practical manual*. (4th ed.), London: AsLib. 2000.
2. Bates, M. Where Should the Person stop and the information search interface start. *Information Processing & Management*, 26(5), 1990, 575-591.
3. Fidel, R. and E.N. Efthimiadis. Terminological knowledge structure for intermediary expert-systems. *Information Processing & Management*, 31(1), 1995, 15-27.
4. Spink, A. and T. Saracevic. Interaction in information retrieval: selection and effectiveness of search terms. *Journal of the American Society Information Science*, 48(8),1997, 741-761.

5.,6.,7. Fidel, R. Searchers' selection of search keys: I The selection routine; II Controlled vocabulary or free-text searching; III Searching styles. *Journal of the American Society for Information Science*, 42(7), 1991, 490-500, 501-514, 515-527.

8. Iivonen, M. and D.H. Sonnenwald. 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), 1998, 312-326.
9. Pollitt, A.S., M.P. Smith, and P.A.J. Braekevelt. View-based searching systems - a new paradigm for IR based upon faceted classification and indexing using mutually constraining

knowledge-based views. in C. Johnson and M. Dunlop (Eds.) *Joint workshop of the Information Retrieval and Human Computer Interaction Specialist Groups of the British Computer Society*. Glasgow University. 1996. 73-77.

10. Alani H., C. Jones and D. Tudhope. Associative and spatial relationships in thesaurus-based retrieval. in J. Borbinha, T. Baker (eds.). *Proceedings (ECDL 2000) 4th European Conference on Research and Advanced Technology for Digital Libraries,* (J. Borbinha, T. Baker eds.), Lecture Notes in Computer Science, Berlin: Springer, 2000, 45-58.
11. Beaulieu, M. Experiments on interfaces to support query expansion*. Journal of Documentation*, 53(1), 1997, 8-19.
12. Oakes, M.P. and M.J. Taylor. Automated assistance in the formulation of search statements for bibliographic databases. *Information Processing & Management*, 34(6), 1998, 645-668.
13. Berenci, E., C. Carpineto, V. Giannini and S. Mizzaro. Effectiveness of keyword-based display and selection of retrieval results for interactive searches, *Proceedings (ECDL 1999) 3rd European Conference on Research and Advanced Technology for Digital Libraries,* (S. Abiteboul and A. Vercoustre eds.). Lecture Notes in Computer Science, Berlin: Springer, 1999, 106 - 125.
14. Tudhope D., P. Beynon-Davies, H. Mackay and R. Slack. Time and representational devices in Rapid Application Development.*Interacting with Computers*,13(4),2001,447-466
15. Kuhlthau C.C. and M.J. McNally. Information seeking for learning: a study of librarians' perceptions of learning in school libraries. *Proceedings 3rd International Conference on Information Seeking in Context (ISIC III), New Review of Information Behaviour Research*, Vol. 2, 2001, 167-177.
16. FACET Project, University of Glamorgan: Pontypridd, 2000. http://web.glam.ac.uk/schools/soc/research/hypermedia/facet_proj/index.php
17. National Museum for Science and Industry http://www.nmsi.ac.uk/
18. Art and Architecture Thesaurus, J. Paul Getty Trust, 2000. http://www.getty.edu/research/tools/vocabulary/aat
19. Tudhope D., C. Binding, D. Blocks and D. Cunliffe. Compound descriptors in context: a matching function for classifications and thesauri, *Proceedings Joint Conference on Digital Libraries (JCDL'02),* Portland, ACM Press, forthcoming, 2002.
20. Rada, R., H. Mili, E. Bicknell, and M. Blettner, Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, 19(1), 1989, 17-30.
21. Tudhope, D. and C. Taylor. Navigation via similarity: Automatic linking based on semantic closeness*. Information Processing & Management*, 33(2), 1997, 233-242.
22. Chen H. and Dhar V. 1991. Cognitive process as a basis for intelligent retrieval systems design. *Information Processing & Management*, 27(5), 405-432.
23. Harter, S.P. and C.A. Hert. Evaluation of information retrieval systems: Approaches, issues, and methods. *Annual Review of Information Science and Technology*, 32,1997, 3-94.
24. Savage-Knepshield, P.A. and N.J. Belkin. Interaction in information retrieval: Trends over time*. Journal of the American Society for Information Science*, 50(12), 1999, 1067-1082.
25. Seymour, S. Online public-access catalog user studies - a review of research methodologies, March 1986 - November 1989*. Library & Information Science Research*, 13(2), 1991, 89-102.
26. Tague-Sutcliffe, J. The pragmatics of information-retrieval experimentation, revisited*. Information Processing & Management*, 28(4),1992, 467-490.
27. Branch, J.L. The trouble with think alouds: Generating data using concurrent verbal protocols. in A. Kublik (ed.) *Proc. of CAIS 2000: Dimensions of a global information science*. Canadian Association for Information Science. 2000.
28. Camtasia, TechSmith Corporation, 2000, http://www.camtasia.com/products/camtasia/camtasia.asp
29. Kuhlthau, C.C. Inside the search process - Information seeking from the users perspective*. Journal of the American Society for Information Science*, 42(5),1991, 361-371.