

---

# User evaluation of a pilot terminologies server for a distributed multi-scheme environment

---

*Ali Shiri*

*Dennis Nicholson and*

*Emma McCulloch*

---

## The authors

Ali Shiri is a Senior Researcher, Dennis Nicholson is the Director and Emma McCulloch is a Researcher, all at the Centre for Digital Library Research, University of Strathclyde, Glasgow, UK.

---

## Keywords

Knowledge organizations, Collections management

---

## Abstract

The present paper reports on a user-centred evaluation of a pilot terminology service developed as part of the High Level Thesaurus (HILT) project at the Centre for Digital Library Research (CDLR) in the University of Strathclyde in Glasgow. The pilot terminology service was developed as an experimental platform to investigate issues relating to mapping between various subject schemes, namely Dewey Decimal Classification (DDC), Library of Congress Subject Headings (LCSH), the Unesco thesaurus, and the MeSH thesaurus, in order to cater for cross-browsing and cross-searching across distributed digital collections and services. The aim of the evaluation reported here was to investigate users' thought processes, perceptions, and attitudes towards the pilot terminology service and to identify user requirements for developing a full-blown pilot terminology service.

---

## Electronic access

The Emerald Research Register for this journal is available at  
[www.emeraldinsight.com/researchregister](http://www.emeraldinsight.com/researchregister)

The current issue and full text archive of this journal is available at  
[www.emeraldinsight.com/1468-4527.htm](http://www.emeraldinsight.com/1468-4527.htm)

## Introduction and overview

This paper deals with a workshop-based user evaluation of a pilot terminologies server developed as part of the Joint Information Systems Committee (JISC) funded High Level Thesaurus (HILT) Phase II project. Details of the evaluation process are provided, and workshop results and their significance for HILT are summarised. An illustrated description of the pilot server is also presented. Key workshop results are included in three appendices at the end of this paper. The complete results set is available in Appendix D.3 of HILT (2003). The paper does not cover the general problems of attaining subject interoperability in a distributed, multi-scheme environment. Further information on these can be found in Nicholson and Shiri (2003), Nicholson *et al.* (2002), and in section 3 of the addendum to HILT (2003).

HILT Phase II was funded to set up a pilot terminologies service for the JISC information environment (IE), aiming to:

- provide a practical experimental focus within which to investigate and establish subject terminology service requirements for the JISC IE; and
- make recommendations as regards a possible future service.

The workshop furnished user feedback that influenced the subsequent development of the terminologies server specification, providing feedback on the usability of the pilot interface, the need for a UK oriented scheme modifications registry (see Nicholson *et al.*, 2002), the need for mapping to fairly specific levels of granularity, the effects of training, and other matters. However, its outcomes are also of interest beyond HILT.

Specific workshop aims were:

- to find out what students, researchers, lecturers and intermediaries thought of the pilot interface and its features and facilities (how could they be improved?) (primary aim);
- to discover something about their subject retrieval behaviour and associated thought processes;
- to compare the terms they use with terms in the HILT database;
- to compare terms used by students, researchers, lecturers and intermediaries to describe some documents by subject (URLs);
- to see whether there was any evidence in the results to suggest that learning or experience improves performance in using the interface; and
- to utilise the data obtained to learn about the efficacy of the general approach.

---

Refereed article received 18 March 2004  
Accepted for publication 8 April 2004



## The pilot terminologies server

### The primary purpose of the server

At an early stage, a view was taken on the primary purpose of the server. It was agreed with the project management and steering groups that the function of the proposed terminologies server should be: to optimise the ability of users to carry out successful subject searches by providing a process that would, in time, permit JISC and JISC services to:

- (1) Achieve and maintain as high a level of interoperability as possible between:
  - the different standard subject schemes and versions of standard schemes in use in different services, both within and beyond JISC;
  - amendments, additions, and extensions made to standard schemes across the services; and
  - terms employed by users when composing search strategies.
- (2) Optimise both the consistency with which staff across various services apply schemes in the subject description of materials, and the ability of users to formulate successful search queries, through the provision of information on descriptive term usage, appropriate training, and helpful feedback mechanisms (e.g. a “disambiguation” facility to help clarify the subject of a user search).

Note that the aim is neither to improve precision at the expense of recall, nor to improve recall at the expense of precision, but rather to provide users with the information they require to do either of these things depending on their needs at a given time.

### Assumptions

Two assumptions were made at the outset:

- (1) That the terminologies server would be the basis of a community process that would develop, maintain, and gradually improve interoperability of subject descriptions by mapping terminology sets to a Dewey Decimal Classification (DDC) spine and that the aim of the project was to determine specific design requirements based on this approach. Not only was the focus on mapping in line with the community consensus in HILT I (see <http://hilt.cdrl.strath.ac.uk/Reports/Documents/HILTfinalreport.doc>), which strongly favored this approach over the adoption of a single scheme and other options (HILT, 2002), it was also felt that mapping schemes together was probably the approach most likely to be compatible internationally. Even if JISC were to adopt a single subject or

class scheme across all services, it is unlikely that the same scheme would be widely adopted elsewhere. Mapping is a widely used approach to providing interoperability (see for example, CARMEN Project – [www.bibliothek.uni-regensburg.de/projects/carmen12/](http://www.bibliothek.uni-regensburg.de/projects/carmen12/); MACS Project – <http://infolab.kub.nl/prj/mac/s/>; LIMBER Project – [www.limber.rl.ac.uk](http://www.limber.rl.ac.uk); RENARDUS [www.renardus.org/](http://www.renardus.org/)). Work that focused on DDC as the basis of the mapping is described in Heery *et al.* (2001), Koch *et al.* (2001), and Saeed and Chaudhury (2002).

- (2) That software produced by Wordmap (see [www.wordmap.com/](http://www.wordmap.com/)) would be the basis of the pilot service. This is built around an Oracle database of terminology mappings, a windows “drag and drop” type staff interface to enable the creation and amendment of mappings, and a Web-based user interface. The staff interface was used “as is” but other elements (the database and, more particularly, the user interface) were significantly altered to suit project purposes.

### Illustrated description of the pilot server

At the time of writing, the pilot terminologies server is accessible on the Web at <http://hilt.cdrl.strath.ac.uk/pilot2/top.php>. Whether it will be available long-term will depend on whether or not a HILT Phase III is funded as recommended by the Phase II Final Report and on the basis of any Phase III. Details of its operation can be found in HILT (2003) and in the illustrative description provided below.

The following is an illustrated description of the pilot as evaluated at the user workshop.

#### Subject schemes

Four subject schemes have been incorporated into the pilot terminologies server. These are:

- (1) DDC.
- (2) Library of Congress Subject Headings (LCSH).
- (3) Unesco.
- (4) Medical Subject Headings (MeSH).

The DDC to LCSH mapping was provided by OCLC and is extensive but not complete. The Unesco and MeSH mappings were added by the project team and are illustrative only.

#### JISC collections database

One of the components of the server is a database covering JISC collections and services, consisting of URL links and brief descriptions of each collection or service, or service collection strength. Each collection, service, or service collection strength is classified by DDC. The database also records which subject scheme collections and

services use to describe resources. The database is used within HILT to identify which collections are relevant to a user’s search and what schemes the collections use locally.

*Mapping functions*

A literature review was conducted to investigate the problems and issues in integrating and mapping thesauri and classification schemes, and to study the different types of mapping in existence. A list of 19 match types identified by Chaplan (1995) was noted as a design issue. In order to explore further the problems and issues of mapping in practice some mapping exercises were carried out and reported in HILT (2003). Based on the literature review and the mapping exercise some simple match types were selected to build into the pilot server. The match types are shown in Table I with examples.

*User and staff interfaces*

The end-user interface of the pilot server is illustrated by Figures 1-3, covering the home page, the disambiguation page, and the collection identification page. Aspects of their design were influenced by a review of thesaurus-enhanced search interfaces conducted by Shiri *et al.* (2002a).

The staff interface allows the librarians and indexers to create, amend or manipulate different versions of subject schemes held within the pilot terminologies server. For instance, staff can make changes to local versions of DDC or Unesco through the interface. This is illustrated by Figure 4.

Figure 1 shows the homepage of the HILT pilot interface. This consists of a search bar, a brief description of the service, a link to search tips and DDC specific subject categories for browsing. The user enters a search term and the system responds with an appropriate disambiguation screen.

Figure 2 depicts the disambiguation stage of the HILT pilot service with possible options retrieved from DDC. At this stage the user contextualises their search term(s) and decides which option to choose. This page consists of options set in context by displaying search terms within the appropriate DDC hierarchy. A button labelled “more results”

is available should users desire to retrieve further options. The user selects their choice and the system responds with the collection selection screen.

Figure 3 illustrates the collection selection stage. The features on this page include: a search bar, the DDC hierarchy for the selected term, a browsable list of JISC collections retrieved, information on the subject scheme used with a mapping from the unprocessed DDC number to the appropriate term in the subject scheme used by the highlighted collection, the DDC number for the collection, and a link to the host service’s search screen option. If the user clicks on the highlighted term (e.g. Tooth in Figure 3), the collection is searched automatically using the selected term and sample retrieval presented to the user.

Figure 4 shows the HILT pilot terminologies staff interface where different terminology instances can be created, modified, manipulated and mapped.

**HILT methodologies and studies of user subject searching requirements**

Research work in HILT Phase II was coordinated via a methodologies document. A section of this document dealt with “methodologies to ensure the investigation examines representative user types, tasks, and associated retrieval requirements, and strategies” and outlined work designed to ensure that project staff were well-informed on user-associated issues as they developed views on the HILT model, mapping, functionality and interface features, cost-benefit analysis requirements, and so on. This work had two main products. The first was a survey conducted by interviewing a range of users on their views on, and approaches to, subject searching. The second was the workshop-based user evaluation covered in the present paper. This focused on managed user interactions with the pilot terminologies service described above and drew out information on a range of issues relevant to its design and on some of the assumptions that underpinned the design. Further information on the approach used is provided in the next two sections.

**Summary of workshop conditions and processes**

**Participants**

A total of 41 further and higher education users, covering a range of subject areas, and including

Table I Examples of match types implemented in the pilot server

Match type	First scheme	Second scheme
Type 1: singular plural	DDC	MeSH
	Teeth	Tooth
Type 2: exact match	DDC	LCSH
	Teeth	Teeth
Type 3: concept match	DDC	Unesco
	Persons in late adulthood	Elderly

Figure 1 Homepage of the HILT pilot terminologies service

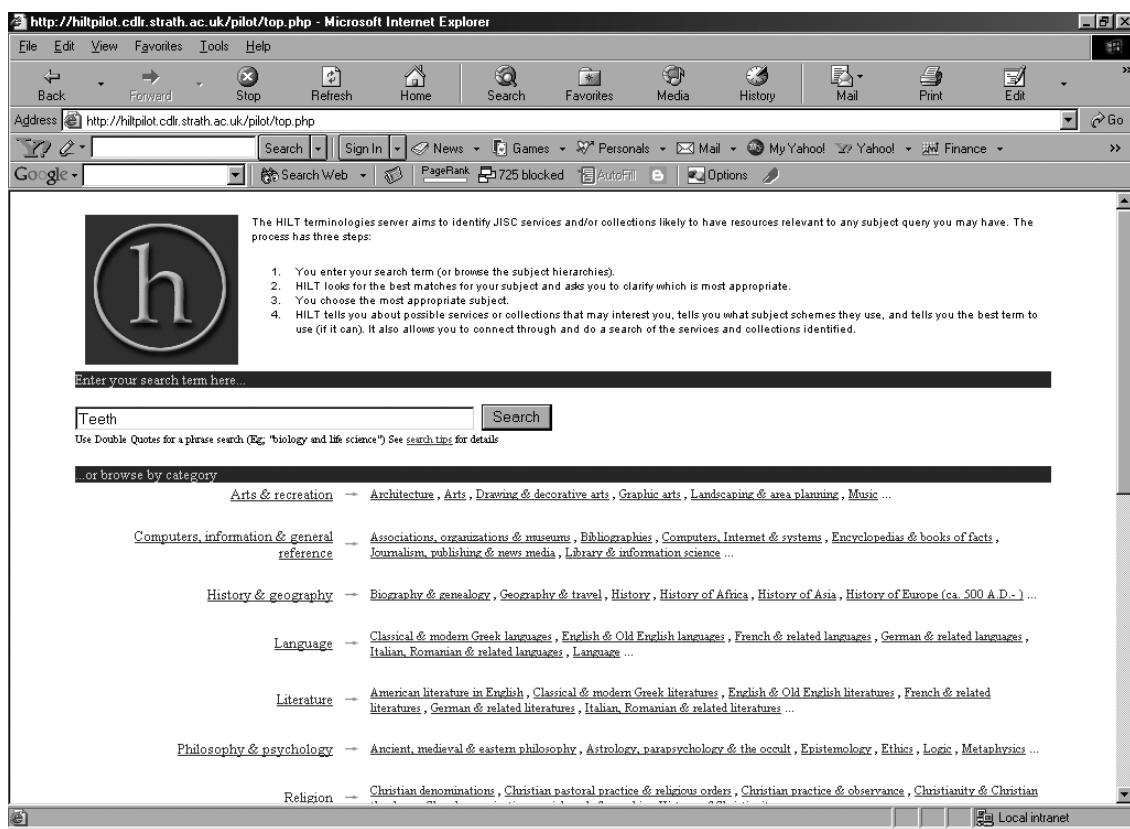


Figure 2 Disambiguation page of the HILT pilot terminologies service

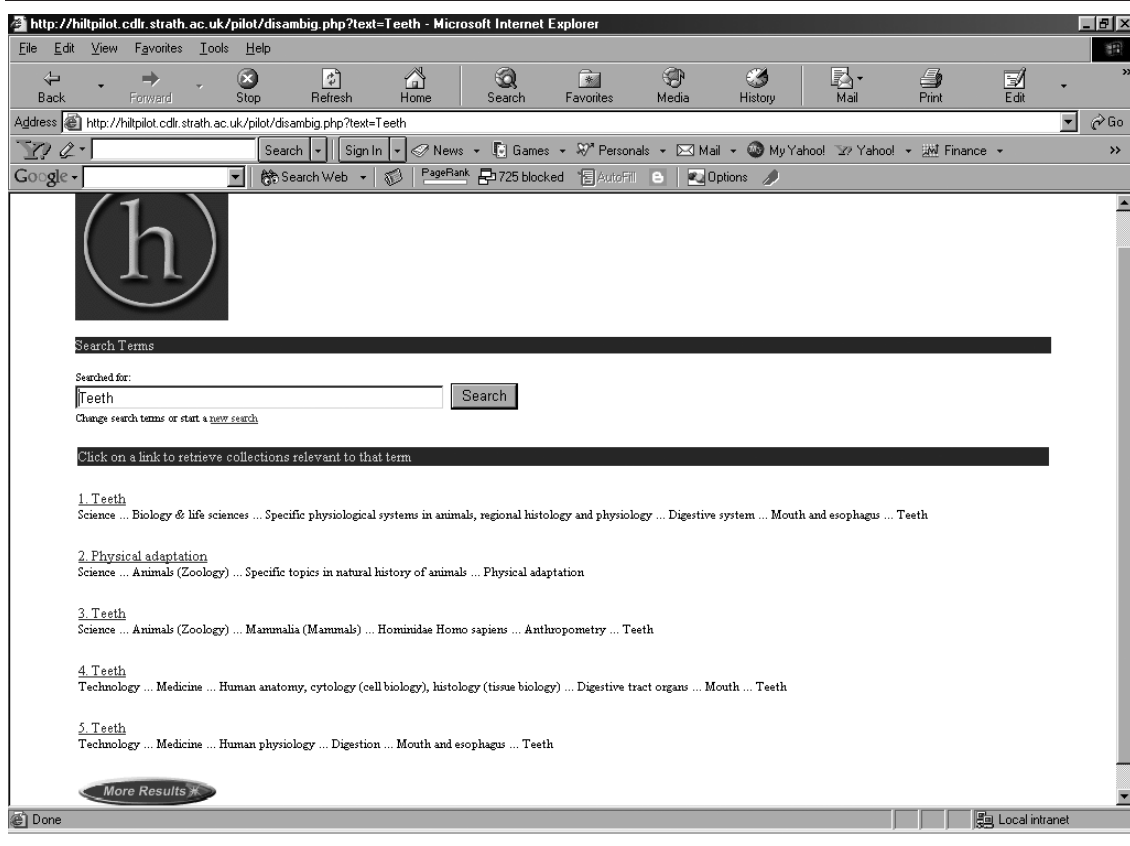


Figure 3 Collection selection page of the HILT pilot terminologies service

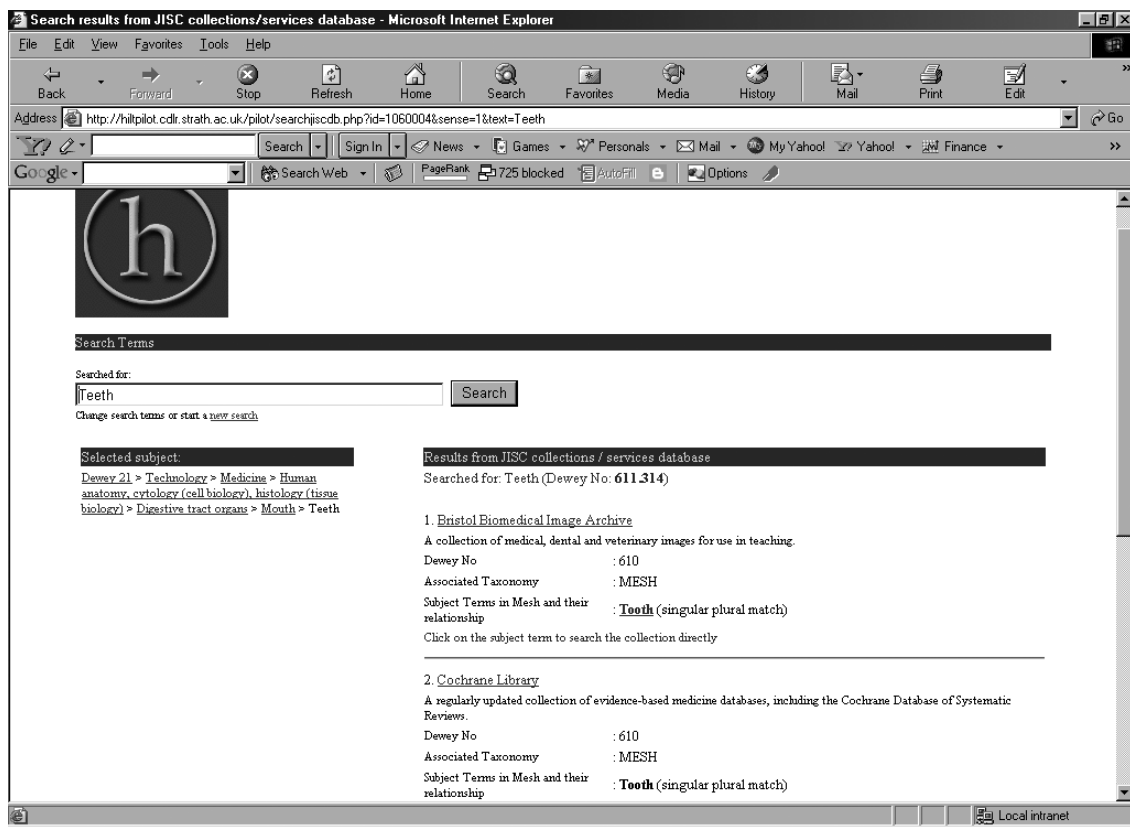
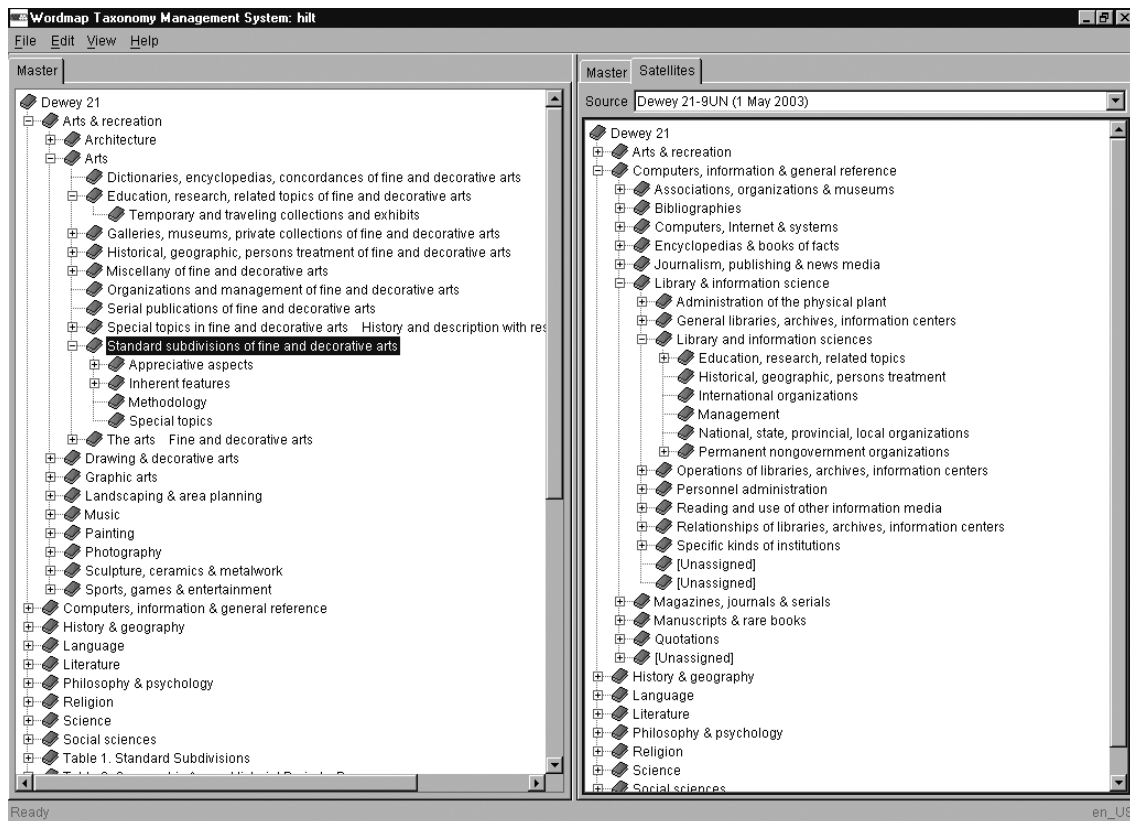


Figure 4 HILT pilot terminologies service staff interface



undergraduate and postgraduate students, researchers, teachers, lecturers and intermediaries such as librarians and electronic information service staff, were recruited. The workshop was run twice in a single day. A total of 20 participants attended the morning session and 21 the afternoon session. A review by Shiri *et al.* (2002b) of studies involving different types of users interacting with a thesaurus-aided search environment helped inform the workshop design.

### Search-tasks

Search tasks used in the study comprised a combination of assigned search tasks and real search requests elicited from participants. Since medical terminology was a primary focus of the pilot, users were provided with an assigned medical term to search for. This allowed participants to familiarise themselves with the interface and the way it deals with a query. Users were also asked to search for a medical term of their choice. In addition, users were asked to bring a recently set exam, assignment or research question to use as the basis for a search task.

### Data gathering tools

A number of data gathering tools have been utilised in interface evaluation and usability studies. For instance, the RENARDUS project made use of workshops, online questionnaires and interviews for gathering data on the usability of the service (RENARDUS, n.d.). The main data-gathering tool used in the HILT study was a questionnaire presented in a workshop setting. It was divided into three sections. Section one elicited users' personal information. Section two aimed to identify users' perceptions of the interface in terms of ease of understanding, instruction given, language used and layout and organisation of the screen at different stages of the search process. For these issues, users searched the assigned medical search task as well as their own medical term. A seven-point Likert scale was presented for users to rate the ease of: finding the relevant term(s) at the disambiguation stage, the relevance of collections and services to the topic selected and the relevance of results retrieved from the collections database. In section three users were presented with two Web-based resources and asked to write down three subject terms to describe each. The aim was to investigate the way in which users describe a resource using subject terms and how consistency/inconsistency in the description of the same Web resource can be dealt with (the full questionnaire is reproduced in Appendix D.2 of HILT, 2003).

### Experimental procedure

The workshop was conducted as follows:

- Participants were asked to attend the workshop venue for either the morning or afternoon session (times were pre-assigned).
- Each received a brief description of the workshop aim and of the steps to be taken to complete the session.
- Each received a copy of the eight page workshop questionnaire.
- A randomly chosen selection of 15 of the 41 participants received a demonstration of the pilot terminologies server and a brief talk on its function. This lasted for five minutes and questions were then invited. The objective was investigate whether or not training would have an effect on user performance when operating the pilot server.
- In completing the workshop task, participants were asked to provide as much detailed information as possible on their thought processes during the course of their interaction.
- Participants were also told they could ask questions at any time if they needed clarification on tasks described in the questionnaire. In the event that they did ask, HILT staff expanded on the printed instructions but avoided telling them anything new that might skew results.
- Following completion, each participant's questionnaire was checked to ensure that HILT researchers had no difficulty in reading and/or understanding the responses.

### Summary results related to workshop aims

#### Primary aim

To find out what students, researchers, lecturers and intermediaries think of the interface and its features and facilities (how could they be improved?).

The results here were generally very positive; surprisingly so, given that both the interface and the type of service were entirely new to the users in question. A selection of key results are provided in appendices 1-3 of this paper.

Seven elements were examined. The first four were: ease of understanding the screen, language used, instruction given, and layout and organisation. For each of these elements, the ratings given by users were well above average. Additional comments and suggestions collected from users gave specific data that will be of value in

improving the interface in an operational server. The remaining three elements examined were ease of choosing the most appropriate term (disambiguation stage), relevance of the collections found and relevance of the results retrieved from the collections database. Ratings given by users for these elements were good for the medical searches where users chose their own terms and topics, and less so (but still mostly above average) for the topics from real exam or essay questions which provided more general subject coverage.

*Usability of the interface*

The HILT workshop questionnaire asked a set of questions relating to four main search tasks. Detailed information elicited from participants is summarised below under the following headings:

- A medical search selected by HILT (Homepage (Table II), Disambiguation (Table III), Collection Selection stages (Table IV)).
- A medical search selected by the users (Table V).
- Two user selected searches in user selected subject areas (Tables VI and VII).

Ratings shown are on a seven-point scale from 1 (very unclear) to 7 (very clear).

**Secondary aims**

To discover something about users' subject retrieval behaviour and associated thought processes.

Some data collected of relevance to this topic has been analysed and is reported under the other headings in this overview. However, a good deal of it could not be tackled fully within the timescale of

**Table II** Users' rating of the homepage

Users' impression of the homepage	Mean rating
Ease of understanding of the screen	5.9
Language used	5.8
Instruction given	6.1
Layout and organisation	5.4

**Table III** Users' ratings for disambiguation screen

Users' impression of the disambiguation screen	Mean rating
Ease of understanding of the screen (term teeth)	5.7
Language used	5.4
Instruction given	6.7
Layout and organisation	5.8

**Note:** "Enter the term 'teeth' into the search box and click on 'search'. Look at the new screen."

**Table IV** Users' rating for the results screen

Users' impression of the results screen	Mean rating
Ease of understanding of the screen (results for teeth)	5.3
Language used	5.2
Instruction given	5.7
Layout and organisation	5.3

**Note:** "Click on option 4 from the list of results (Teeth; Dewey: 611.314). Look at the new screen."

**Table V** Users' ratings for self-selected medical term

Users' impression of the search for their self-selected medical term	Mean rating
Ease of choosing the most appropriate term (disambiguation stage)	6.04
Relevance of collection/service to the medical term (result stage)	5.04
Relevance of the results retrieved from the collection/service	4.7

**Note:** "Please think of another medical term and write it down. Return to the homepage and carry out a search on the term you have written down."

**Table VI** Users' ratings for the first term from assignment

Users' impression of the search for their assignment or research interests (first term)	Mean rating
Ease of choosing the most appropriate term (disambiguation stage)	4.5
Relevance of collection/service	3.9
Relevance of the results retrieved from the collection/service	3.5

**Table VII** Users' ratings for the second term from assignment

Users' impression of the search for their assignment or research interests (second term)	Mean rating
Ease of choosing the most appropriate term (disambiguation stage)	3.9
Relevance of collection/service	4.2
Relevance of the results retrieved from the collection/service	3.1

the project. In particular, the questionnaires filled in by participants provide detailed information on how searches were conducted and the participants' assessments of the responses received from the pilot service. Full analysis of this requires a repetition by project staff of each user search and a comparison of their comments with the responses of the system.

To compare the user terms with terms in the HILT database.

The self-selected terms employed by users are charted in Appendix 1 of this paper. Further information is available in HILT (2003), including:

- terms utilised in the user-selected medical search task; and
- a detailed matching of user terms against the HILT pilot terminologies pilot set (access to the whole of DDC 21, indexed on the DDC captions, standard sub-divisions, relative index, LCSH as mapped to DDC by OCLC, illustrative mappings to Unesco and MeSH).

Analysis showed that close to 80 percent of search terms used by participants also appeared in the HILT mappings database. Approximately 20 percent of terms were not in the database and will be studied in any HILT Phase III as possible candidates for inclusion in a UK oriented scheme modifications registry term set (“digital divide” might be one example of a candidate term).

The exercise produced some additional evidence of the need for a UK oriented scheme modifications registry term set; in particular, the attempt by a participant to search for GP (general practitioner).

To compare terms used by students, researchers, lecturers and intermediaries to describe some documents by subject (URLs).

Participants were given two Web sites and asked to provide three subject terms, for each, that they would use to describe them. It was hoped that the results would cast some light on any differences in terminologies utilised by students and staff (researchers, lecturers and intermediaries) describing resources. In the event, the results from the two groups were not very different as far as appropriate responses were concerned (many students may have misunderstood the task. They often used terms like “colourful”, “crowded”, “friendly”, “appealing”). This may have been because the subjects covered by the sites they were asked to describe were too obviously specified on their homepages. A better test might have been to ask staff to use standard schemes and to also cover areas where UK oriented scheme modifications were thought to be needed. However, this would not have been practical in the workshop context. The full results of the exercise are available in HILT (2003).

To see whether there is any evidence in the results to suggest that learning or experience improves performance in using the interface.

A small test was carried out to test this idea. Some participants were given a short presentation and demonstration of the service prior to completing the questionnaire; the remainder were left to their own devices. Results provided some evidence that

training improved success, as presented in Appendix 2. In addition, 30 of the 41 participants stated that they felt more confident in the use of the interface as the tasks progressed (the questionnaire itself was designed to introduce participants to the various screens of the pilot in a gradual and progressive way and then required them to repeat their use of it for other topics. In short, the questionnaire itself was a training programme to some extent).

To utilise the data obtained to learn about the efficacy of the general approach (taken with service design).

See above under the effects of training, participants search terms, and participant reactions to the interface.

### Other results

*Results relating to the likely efficacy of limited granularity mapping*

The project collected information on the kinds of exam, essay and tutorial questions set for or by the group of students, researchers, lecturers and intermediaries participating in the workshop (see Appendix 3). This, together with the search terms used by the group at the workshop (see Appendix 1), and those used in the earlier user interviews, lend support to the view that searches by users tended, in many cases, to cover fairly specific terms, suggesting that mapping only at higher levels of granularity was not sufficient to meet user needs.

*The need for a more complex disambiguation interface*

It became clear from project work undertaken just before, during, and immediately after the user workshop that an interface permitting only one disambiguation choice was insufficient to deal with the complexity of user queries likely to be encountered, a conclusion also supported by the data on user assignments and search terms collected by the project.

### Lessons learned

As is clear from the above summaries, the outcomes of the exercise were mixed in respect of the value of the data obtained. However, the following are worth noting as positive outcomes from the process:

#### Usability of the pilot interface

Whilst it is clear there is scope for further improvement in respect of interface design, the usability results from this early pilot were encouraging. User evaluation of the various server



screens, and their content, in the experiments conducted suggested that the approach to a range of design issues was at least on the right track. In addition, a good deal of further information was collected that will help identify design features regarded as either positive or negative by workshop participants and so help inform future improvements. It was also encouraging to note that users generally understood the purpose of the terminologies server and that there was also some evidence that training had a positive effect on user understanding and performance.

### Subject retrieval behaviour

Although some of the data collected by the project on this front has yet to be analysed, useful information was gleaned from the exercise (see under the headings below).

### The effects of training

Results provided some confirmation that, as assumed when the project outlined the primary purpose of a terminologies server (see paragraph above on the primary purpose of the server), training had a role to play in improving interoperability in respect of the terms employed by users and those featured in standard schemes.

### The need for a UK-oriented scheme modifications registry

During the workshop, one user verbally commented that he had entered "GP" for general practitioner and had inappropriate results. This, together with possible examples (e.g. "digital divide") from the user search terms detailed in Appendix 1) provided some confirmation of the need for an UK oriented scheme modifications registry first identified in HILT Phase I. HILT Phase I surveyed staff at JISC services on whether they amended or extended standard subject schemes and found that they did. The main reasons given for doing this were:

- to accommodate new concepts or areas of knowledge (31 percent);
- subject scheme is too broad (27 percent); and
- to reflect user needs and demands (27 percent).

### Inclusion of subject terms used by participants in the HILT mappings database

The results were encouraging in that they showed that four out of five terms input by users were in the pilot database already. Of more potential value, however, is the fact that the project was able to gather illustrative examples of terms not included in the database. This data will help inform the development of any future operational service.

### More complex disambiguation facility

The pilot interface allowed users to choose only one term when disambiguating a search topic they had input. An comparison of system responses with the topics being researched by participants (see Appendix 3) and the terms used in their searches (see Appendix 1) showed that this would not always be sufficient.

### Limited granularity mapping

Data from the workshop, together with supporting evidence from the earlier user interviews, also informed the project's view as regards the depth of mapping required in the terminologies server. The option of mapping between subject schemes, user terms, and DDC at less specific levels of granularity only was ruled out on the grounds that limiting mapping in this way would make it impossible to deal with the vast majority of most user subject queries. As indicated by the examples in Appendices 1 and 3 of the present paper, these tend at times to be more, rather than less, specific than the levels of granularity available in standard schemes. A particular point noted by the project is that there is no necessary connection between more general levels of granularity in subject description and "collection level requirements". The user need will most often be to map a subject search at a very specific level of granularity up to a collection classified at a higher level and then down again, within the local scheme used, to a level of granularity appropriate to the original query. Limited granularity mapping would not permit this. This is regarded as a significant outcome of the project, and suggests that the idea of a "high level thesaurus" which gives HILT its name gives a misleading perspective on the problem tackled by the project.

## References

- Chaplan, M.A. (1995), "Mapping Laborline thesaurus terms to Library of Congress subject headings: implications for vocabulary switching", *Library Quarterly*, Vol. 65 No. 1, pp. 39-61.
- Heery, R., Carpenter, L. and Day, M. (2001), "Renardus Project developments and the wider digital library context", *D-Lib Magazine*, Vol. 7 No. 4, available at: [www.dlib.org/dlib/april01/heery/04heery.html](http://www.dlib.org/dlib/april01/heery/04heery.html)
- HILT (2002), "HILT Phase I: final report", available at: <http://hilt.cdrl.strath.ac.uk/Reports/FinalReport.html>
- HILT (2003), "HILT Phase II: final report", available at: <http://hilt.cdrl.strath.ac.uk/hilt2web/finalreport.htm>
- Koch, T., Neuroth, H. and Day, M. (2001), "Renardus: cross-browsing European subject gateways via a common classification system (DDC)", *Proceedings of the 2001 IFLA Satellite Conference on Subject Retrieval in a Networked Environment*, Dublin, 14-16 August.

- Nicholson, D. and Shiri, A. (2003), "Interoperability in subject searching and browsing", *OCLC Systems and Services*, Vol. 19 No. 2, pp. 58-61.
- Nicholson, D., Dunsire, G. and Neill, S. (2002), "HILT: moving towards interoperability in subject terminologies", *Journal of Internet Cataloguing*, Vol. 5 No. 4, pp. 97-111.
- RENARDUS (n.d.), "RENARDUS Evaluation Report", available at: [www.renardus.org/about\\_us/deliverables/d5\\_2/D5\\_2summ.html](http://www.renardus.org/about_us/deliverables/d5_2/D5_2summ.html)
- Saeed, H. and Chaudhury, A.S. (2002), "Using Dewey decimal classification scheme (DDC) for building taxonomies for knowledge organisation", *Journal of Documentation*, Vol. 58 No. 5, pp. 575-83.
- Shiri, A.A., Revie, C. and Chowdhury, G. (2002a), "Thesaurus-enhanced search interfaces", *Journal of Information Science*, Vol. 28 No. 2, pp. 111-22.
- Shiri, A.A., Revie, C. and Chowdhury, G. (2002b), "Thesaurus-assisted search term selection and query expansion: a review of user-centred studies", *Knowledge Organization*, Vol. 29 No. 1, pp. 1-19.

## Appendix 1. Users' search terms

In total users provided 159 terms for their searches. The average number of terms provided per search was two. Table AI summarises the terms entered by all users and are categorised according to the number of terms in each search. The terms vary greatly in granularity level, from fairly general headings like "educational psychology" to fairly specific terms like "cleft palate" or "theatrical agent 'Ciaran Hinds'" or "angular modification".

Table AI Search terms entered by all users

	Search terms	No. of searches
Searches with one word	Scottish; engineers; food; WTO; copyright; depression; forecasting; e-commerce; finance; Internet; mobile; tuna; collectors; law; Taiwan; women; library; actors; legend	19
Searches with two words	Energy efficiency; Scottish parliament; cleft palate; archival description; local authority; charter flights; child development; theory of constraints; library digitisation; PPP and relative; facility lay-out; digital divide; child psychology; space toy; trucking industry; Internet advertising; protein function; canned tuna; nearsight vision; succession (Scotland); neural network; Internet regulation; business to employee; Internet law; share price; interstate succession; angular magnification; protein engineering; Soviet Union; advertising effectiveness; educational psychology; digital library; Web page design; financial engineering; commercial Internet; Box Jenkins; communication flow; neurotransmitters physiology; social inclusion; project management; food safety; developmental psychology; UK regions; construction in Scotland; information standards; breast feeding; Scottish Nationalism; domestic appliances	48
Searches with three words	Macgregor building industry; epidemiology of HIV; Aids in Uganda; information and library studies; British foreign policy; seasoned issues of equity; digital social inclusion; information literacy: Africa; radial basis function; business intelligence tools; design a children Web site; Internet service provision; PPP and medium term	12
Searches with four words	Theatrical agent "Ciaran Hinds"; Web design or Web site and evaluation; Aids treatment and care in Uganda	3
<b>Total</b>		<b>82</b>

**Notes:** Around 80 per cent of users' search terms matched terms in the HILT database. Around 70 per cent of users' search terms have been matched to the HILT pilot server either exactly or partially. Concept matches constituted around 6 per cent of search terms. Over 3 per cent of users' search terms were matched to either broader or narrower terms in the HILT database and around 20 per cent of terms were not matched. An examination of these may provide useful data on the likely content of a UK-oriented scheme modifications registry term set

## Appendix 2. The effect of demonstration on users' judgement of the interface

A number of statistical tests were carried out to establish whether those who attended the demonstration session judged the interface differently from those who did not. In the following we report the results of the statistical tests.

A *t*-test on the effect of the demonstration session on participants' ratings of "instruction given on the HILT homepage" resulted in a *p*-value of 0.099. Table AII summarises the results.

A *t*-test was run to assess the difference between the demonstration and non-demonstration groups in relation to their rating of ease of understanding the "teeth" screen. The *t*-test was found to be significant (*p* = 0.018), indicating that those who attended the demonstration session found the "teeth" screen easier to understand than those who did not attend the session. Table AIII summarises the result that indicates the importance of training for the search process.

A *t*-test was conducted to evaluate the difference between the two groups in terms of rating the ease of finding the subject category relevant to the user's search. The *p*-value was found to be 0.086 reflecting that those who attended the

Table AII The effect of demonstration on rating the instruction given on the homepage

Groups	<i>n</i>	Mean	Std
Demonstration attended	15	6.3	0.61
No demonstration	26	5.9	0.93

**Table AIII** *t*-test results for ease of understanding the "teeth" screen

Groups	<i>n</i>	Mean	Std
Demonstration attended	15	6.3	1.2
No demonstration	26	5.4	0.90

**Table AIV** *t*-test results for ease of finding the subject category on the homepage

Groups	<i>n</i>	Mean	Std
Demonstration attended	15	4.8	0.62
No demonstration	26	3.3	0.51

demonstration session may have found the subject category relevant to their search more easily than those who did not attend the session. Table AIV shows details.

As was shown, clarity of the information given on the homepage, ease of understanding the disambiguation stage for the assigned search task and ease of finding the subject category on the homepage during browsing were all influenced to some extent by the demonstration effect. These findings suggest that training was a factor that affected some aspects of user interaction with the terminology service.

### Appendix 3. Assignment topics provided by participants

- Outline and discuss the theories of absolute and relative purchasing power parity. Does the empirical record suggest that PPP holds in the medium term.
- Produce a project report no longer than 3,500 words outlining the business case for developing an e-commerce; commercial Internet site.
- Investigate business intelligence tools applied to a specific sector of your choice.
- Is there a Web site that holds information about names; addresses; contact details of all the public libraries in Scotland?
- How can society use psychological knowledge to promote the wellbeing of children?
- The First British Copyright Act of 1710 and its struggle against monopolies.
- Discuss the management philosophy of The Theory of Constraints (TOC) as an approach to project management, first described by Dr Eliyahu M. Goldratt.
- Write an essay on the regulation on Internet content.
- How do different methodologies converge to generate psychological knowledge?
- Find information on holiday charter flights from UK airports, to provide a breakdown on market share by geographical region.
- Investigate the impacts of WTO on macro environment.
- Digitisation should provide opportunities for opening up a library's holdings and being more socially inclusive.
- Find information on female engineers from Kelvin's time.
- Are there any Web sites or articles dealing with flaws and biases in probationer training for police officers?
- Choose a food production or processing company, and briefly describe it and its activities and products, including if possible the number of employees and the yearly value of its trade. Describe the quality problems encountered by the company, directly from contact with the company or indirectly from what you know or find out about its products. Describe the solutions to the problems actually adopted by the company, or suggest solutions to the problems you have identified.
- To determine the type of problems that occur in telecommunication business which require the application of business intelligence tools.
- Investigate how to use SPSS to analyse time series.
- Canned tuna: quality problems and the solutions.
- Is there a list of all the local authority libraries in Scotland and how many of them are there?
- Chemistry with IT and instruments.
- Where are the active volcanoes, how are they monitored, what about inactive volcanoes, how long are they inactive before they are no longer a danger?
- Write an essay on private finance initiative.
- LLB.
- Business administration.
- Describe and outline the key differences between radial basis function and back propagation neural networks.
- Prepare a short paper on the open problems related to automatic indexing.
- Outline how protein functionality has been altered in food proteins by genetic modification of the primary structure through manipulation of the genes, or a form of genetic selection or plant breeding.