Revealing Interdisciplinarity in Nanoscience and Technology Queries: A Transaction Log Analysis Approach

Ali Shiri*

School of Library and Information Studies, University of Alberta,
3-09 Rutherford South, Edmonton, T6G 2J4, Alberta, Canada, <ashiri@ualberta.ca>

Ali Shiri is an Associate Professor in the School of Library and Information Studies at the University of Alberta, Edmonton, Canada. Ali's research areas and interests include: user search and interaction behavior, digital library user interfaces, knowledge organization systems in digital libraries, institutional repositories, and subject gateways and social tagging trends and interfaces.

Shiri, Ali. **Revealing Interdisciplinarity in Nanoscience and Technology Queries: A Transaction Log Analysis Approach.** *Knowledge Organization, 38(2)*, 135-153. 86 references.

ABSTRACT: The study reported here investigated the search behaviour patterns of nanoscience and nanotechnology searchers as revealed by transaction log analysis of the NANOnetBASE electronic book digital library. This paper examines the patterns and strategies of nano searchers' query formulation and reformulation, then explores the extent of interdisciplinarity in search queries using the INSPEC and Compendex thesauri. The results show certain query formulation patterns associated with searching in an emerging and interdisciplinary area of nanotechnology such as: the use of multiword and compound query terms, extensive use of search terms beginning with the prefix "nano," hyphenated terms, spelling variations, a large number of query reformulations, and the use of acronyms. The results also indicate that 62% of the unique top terms resulting from mapping users' query terms to the INSPEC Classification codes represented two or more disciplines, specifically terms associated with the Classification code "A" representing "physics." The results have implications for information organization and representation, user interface design and federated searching in digital libraries and multi-subject databases.

Received 11 February 2010; Revised 22 May 2010; Accepted 22 May 2010

* The author would like to gratefully thank Thane Chambers for her assistance in data analysis and Randy Reichardt, Research Services Librarian (Engineering) at the University of Alberta, who provided me with the transaction log data for the NANOnetBASE e-book library.

1.0 Introduction

Nanoscale science and technology have seen rapid growth and expansion in new areas in recent years. Nanoscience and technology are characterized by nano researchers in an increasingly interdisciplinary domain, drawing upon such disciplines as chemistry, physics, materials science, and computer, electrical, mechanical, and biomedical engineering. A key challenge faced by nano researchers, as well as the information professionals providing related information services, is to efficiently identify information resources and to carry out inclusive and effective searches in a diverse and heterogeneous range of digital libraries, web-based databases, and search engines. It is essential to understand the problems and issues of interdisciplinary searching in a strategically and commercially significant area such as nanotechnology to support and enhance researchers' scholarly productivity and innovativeness, and to inform knowledge organization and representation practices in nanotechnology-related library and information services. Kutner (2000), in examining challenges

and issues in searching in interdisciplinary areas, argues that, although electronic bibliographic databases that provide sophisticated searching capabilities and multiple access points to the scholarly literature have been a boon to the interdisciplinary researcher, problems continue to exist in terms of lack of consistent interfaces and consistent controlled vocabularies across databases. While there exist many studies of general information seeking behaviour of interdisciplinary areas of humanities, the sciences, women's studies, and social sciences, very few studies have examined the nature of queries and query formulation and reformulation strategies of interdisciplinary searchers in nanoscience and technology.

2.0 Related work

This literature review focuses on the following areas: the interdisciplinary nature of the sciences; nanoscience and technology citation analysis and domain mapping; information search behaviour of interdisciplinary researchers; information search behaviour of scientists and engineers using electronic information resources; and representative transaction log analysis studies.

2.1 Interdisciplinarity

A number of studies have investigated the notion of interdisciplinarity in the context of science and technology. Qin et al (1997) investigated types and levels of collaboration in interdisciplinary research in the sciences; they concluded that biology and medical sciences are highly interdisciplinary. Pierce (1999) identified three types of interdisciplinary information transfer: Borrowing, Collaboration, and Boundary crossing. Using bibliometric techniques, Morillo et al (2003) measured interdisciplinarity in the sciences based on the Institute for Scientific Information (ISI) multi-assignation of journals in subject categories. They found that a high degree of interdisciplinarity was common in engineering, physics, and chemistry and concluded that new, emerging disciplines are highly interdisciplinary and show a predominance of "big interdisciplinarity." Haythornthwaite (2006) studied exchanges among three distributed interdisciplinary teams (one science and two social science teams) and found the highest indication of mutual exchange in reports of joint research in the science team. Leydesdorff (2007) has introduced a new indicator called "betweenness centrality" as a measure of interdisciplinarity of scientific journals.

2.2 Citation analysis and domain mapping

Most information science studies published relating to nanoscience and technology focus on citation analysis, scientometric, and bibliometric studies of journals and publications in the field of nanoscience and technology (Meyer and Persson 1998; Meyer 2000, 2001; Eto 2003; Morillo et al 2003; Schummer 2004a; Calero et al 2006; Leydesdorff 2007), mapping and visualization of the area of nanoscience and technology (Boyack et al 2002; Zitt and Bassecoulard 2006; Bassecoulard 2007; Milojević 2007), and identifying core journals in the area of nanotechnology (Leydesdorff 2007). A special issue of *Scientometrics* (Volume 70, Number 3, March 2007) focuses on various aspects of nanoscience and technology-related trends and issues. Suggestions for the development of an ontology for the area of nanoscience and technology have also been made by LIS researchers such as Tanaka (2005). Citation analysis studies have also been conducted to show the interdisciplinarity of other scientific areas such as medical and behavioural sciences (McCain 1989), environmental sciences (Steele and Stier 2000), biotechnology (McCain 1995), and chemical engineering (Peters et al 1995).

2.3 Information search behaviour of interdisciplinary researchers

Klein (1996) notes that meeting the interdisciplinary needs of today's library users begins with understanding the activities that create them and their place in the knowledge system. Drawing on the studies carried out by Mote (1962), as well as Packer and Soergel (1979), Bates (1996) concluded that studying researcher information seeking in interdisciplinary fields may tell us not only about the needs and problems of people in those fields—something we very much need to learn about—but also about what factors, in general, contribute to ease and difficulty in scholarly information seeking. She notes that interdisciplinary researchers constitute a significant and distinctive class of scholars, much deserving of research on their particular information needs and information-seeking behavior. Many other researchers have investigated the challenges and issues of searching in interdisciplinary areas (Kutner 2000; Palmer 1999; Palmer and Neumann 2002; Foster 2004; Weisgerber 1993; White 1996).

2.4 Information search behavior of scientists and engineers using electronic information resources

Several studies have focused on the ways in which new technologies such as email, electronic publishing, and the Internet have affected the information seeking behaviour of petroleum engineers (Joseph 2001), as well as scientists and engineers (Llull 1991); informal information sharing among scientists and engineers (Poland 1991); and information sources used by design, process, and manufacturing engineers (Kwasitsu 2003). Pinelli (1991), for instance, reviews the literature of information seeking practices of scientists and engineers to discuss some of the differences in the information seeking patterns of scientists and engineers. Allen (1977) comments that engineers are more likely to use specific forms of literature such as handbooks, standards, specifications, and technical reports. King et al (1994) provide a comprehensive review of engineers' information needs, seeking processes, and use. Leckie et al (1996) provide an analysis and interpretation of empirical studies on the information habits and practices of three groups: engineers, health care professionals, and lawyers to propose a general model of information seeking. Other researchers have investigated the information seeking patterns of engineers and research scientists in an industrial environment (Ellis and Haugan 1997); astronomers, chemists, mathematicians, and physicists (Brown 1999); environmental scientists (Murphy 2003); engineers (Hertzum and Pejtersen 2000); and interdisciplinary scholars in the areas of the arts, social sciences, and science and information media studies (Spanner 2001). Palmer (2005, 2006, and Palmer et al 2007) discusses different levels of research work and information use within an interdisciplinary brain research context. Chau et al (2006) developed a knowledge portal for nanoscience and technology information called NanoPort incorporating collection building, meta-searching, keyword suggestion, and various content analysis techniques such as document summarization, document clustering, and topic map visualization. User evaluation of NanoPort showed that NanoPort could provide results with higher precision than benchmarking search engines such as Google and NanoSpot (Qin et al 2004).

The aim of present research is to investigate the information search behaviour of nanoscience and technology researchers through analyzing transaction logs of a nano electronic book digital library called NANOnetBASE. In particular, it examines the extent of interdisciplinarity in the queries formulated by nano researchers.

2.5 Transaction log analysis studies

Transaction log analysis involves the following three steps: TLA involves the following three major stages, which are as follows: a) collection: the process of collecting the interaction data for a given period in a transaction log, b) preparation: the process of cleaning and preparing the transaction log data for analysis; and c) analysis: the process of analyzing the prepared data (Jansen 2006). Peters (1993) provides a review of transaction log analysis library and experimental IR systems used in the 1980s and early 1990s. Transaction log analysis has been employed to study users of the Internet in public libraries (Curry 2005); search engines (Jansen and Pooch 2001; Cothey 2002; Ke et al 2002; Wen et al 2002; Jansen et al 2005; Mat-Hassan and Levene 2005; Sanderson and Dumais 2007); comparative evaluation of web search engines (Jansen and Spink 2005); OPACs (Millsap and Ferl 1993; Blecic et al 1998; Lau and Goh 2006); library websites (Ghaphery 2005); digital libraries (Jones et al 2000; Mahoui and Cunnigham 2000; Marchionini 2000; Nicholas et al 2006a; Zuccala et al 2007); electronic journals (Eason et al 2000; Davis and Solla 2003; Jamali et al 2005; Nicholas et al 2005; Nicholas et al 2006b); electronic books (Hughes and Buchanan 2001; Connaway and Snyder 2005); and organizational interfaces (Marchionini 2002). Peters (1993) provides a useful discussion of the history and development of transaction log analysis. Troll Covey (2002) provides a very useful guide on the use of transaction log analysis for studying users of digital libraries, OPACs, and library websites. Jansen (2006) presents transaction log analysis as a methodological framework along with strengths and limitations of this method in research. In the present study, transaction logs of an electronic book digital library have been analyzed to investigate the search behaviour patterns of nanoscience and technology searchers, in particular the extent of interdisciplinarity in the queries formulated by nanoscience and technology searchers.

3.0 Methodology

The literature on the information search behaviour of nanoscience and technology information searchers is scarce. The overarching objective of this study was to investigate the information search behaviour of nano researchers interacting with a nanoscience and technology electronic book digital library. A more specific objective was to examine whether the data can provide a basis for investigating the interdisciplinary na-

ture of the queries formulated by nano researchers. As a result, the nature of this research is exploratory rather than hypothesis-driven.

3.1 Research questions

The research questions addressed in this study are listed below. The first three focus on search characteristics and strategies of nano researchers while the remaining two are concerned with interdisciplinarity in the searchers' queries.

– What are the search behaviour characteristics of nano searchers, including session length, query characteristics, and the use of query operators?
– What query formulation and reformulation strategies did users employ?
– What search strategies were adopted by the users (e.g., basic vs. advanced search features, failed and erroneous queries, and spelling issues)?
– Can disciplinary and interdisciplinary queries be identified? What subject areas are covered by the queries formulated by nano researchers?
– Are thesauri capable of providing a basis for identifying interdisciplinary queries in the area of nanoscience and technology?

Table 1 shows the variables used in this study along with their descriptions and the coding method.

The data analysis was based on Jansen's (2006) transaction data analysis method, which includes term level, query level, and session level analyses. The INSPEC thesaurus and the INPSEC Classification Codes were used to map the disciplines represented by the query terms. The INPSEC terms that were found to be exact matches for searchers' query terms were traced to their top terms using their broader terms. All of the top terms were then compared with the INSPEC Classification codes. The classification codes were used due to their discipline-based categorization. A detailed description of the dataset and data analysis methods is provided in the following sections.

3.2 The NANOnetBASE e-book digital library

NANOnetBASE is an e-book database for nanotechnology and nanoscience researchers consisting of nanoscience and technology references in an online collection. This full-text database contains 45 titles that can be accessed in a variety of ways. Researchers can browse titles from subject lists (e.g., Biomedical Engineering, Computer Engineering, Electrical Engineering, Electrical Engineering Communications, Electronics, General, Industrial Engineering & Manufacturing, Laser & Optical Engineering, Mechanical Engineering, and Nanoscience/Nanotechnology) or search within all titles using keywords.

3.3 Transaction log dataset

The data used in this study consisted of transaction logs from the NANOnetBASE digital library between July 2004 and October 2006. The data was acquired from a large Canadian university with strong nanoscience and technology research profile where more than 140 nano researchers are currently active. They include faculty members as well as graduate students and post-doc researchers. It should be noted that the actual users were not involved in this study. Rather, their interactions with the digital library, cap

Variable	Definition	Coding
Session	Session length and the type of interactions users had with the system (e.g., time spent on formulating and submitting a query and time spent on viewing the results). A transaction log may contain many sessions in which users go to digital library for purposes other than searching.	Manual coding: identifying and recording sessions using IP addresses
Query terms	Number of terms per query	Manual count
Search strategies	Construction of queries using Boolean operators; queries with acronyms	Identified and counted queries with operators and acronyms
Query reformulation	Any change to the initial query terms	Manual comparison of initial vs. reformulated terms
Subject areas of queries	Discipline-based analysis of queries	Comparison of query terms with two thesauri, INSPEC and Compendex

Table 1. Definition of variables studies

NANOnetBASE

Activity Type	Activity Detail	User Identifier	Date & Time
Book Viewed	Nano Science and Technology: Novel Structures and Phenomena	129.128.217.125	7/30/2004 4:17:50 PM
Advanced Search	Wright, Patrick	129.128.198.21	9/17/2004 11:09:14 AM
Book Viewed	Handbook of Nanoscience, Engineering, & Technology	129.128.217.4	9/17/2004 1:34:31 PM
Book Viewed	Handbook of Nanoscience, Engineering, & Technology	129.128.216.76	9/22/2004 9:20:42 PM
Book Viewed	Nanoelectromechanics in Engineering and Biology	129.128.216.76	9/27/2004 5:46:32 PM
Advanced Search	vacuum	129.128.216.76	9/28/2004 1:47:06 AM
Quick Search	vacuum	129.128.216.76	9/28/2004 1:47:06 AM
Search Results Viewed		129.128.216.76	9/28/2004 1:47:06 AM
Book Viewed	Nanoelectromechanics in Engineering and Biology	129.128.216.76	9/28/2004 4:33:10 PM

Table 2. Snippet from a sample transaction log

tured in the NANOnetBASE server, were used as research data in this study. In total, 1921 transactions were analyzed. Table 2 provides a sample of transaction logs that were analyzed in the study. The table has the following four columns:

- Date and time of activity.
- User identifier (in the form of IP address).
- Activity detail (query or book viewed).
- Activity type: book viewed, advanced search, quick search, search results viewed, or managed account information.

3.4 Data analysis

Jansen (2006) suggests three common levels of analysis for examining transaction logs (i.e., term, query, and session), so we based our data analysis on these. Before the transactions were analyzed, the transaction logs were converted into a spreadsheet to facilitate data manipulation.

Term level analysis: the number, nature, and frequency of user terms.

Query level analysis: includes query characteristics as follows:
- Query formulation and reformulation (and the reasons for query reformulations),
- Query construction strategies (use of Boolean operators, quick vs. advanced search strategies etc.),
- Query classification (based on disciplines).

Session level analysis: includes session length and the type of interactions users had with the system (e.g., time spent on formulating and submitting a query and time spent on viewing the results). This session definition is similar to the definition of a unique visitor that is used by commercial search engines and organizations to measure Web site traffic (Jansen 2006).

4.0 Results

The results and findings presented here are based on the order of the research questions.

In order to provide a brief overview of the general results, Table 3 shows the number of sessions, queries, reformulated queries, and queries using advanced search features.

Variable	Number
Sessions	552
Queries	393
Queries using acronyms	40
Reformulated queries	155
Queries using advanced search features	22

Table 3. Summary of session and query-related data

4.1 Session length

Jansen et al (2005) noted that a session is the entire series of queries submitted by a user during one interaction with the Web search engine. This definition was taken into consideration while analyzing the dataset in this study. Session length information for the dataset

was only available for the years 2005 and 2006. Table 4 shows the average time spent on each search step by an individual user; the overall search session length ranged from 30 seconds to more than 6 hours. This result should be contextualized by the fact that this is in a digital library that is an electronic book digital library with more than 44 titles. Some users spent more time on consulting the content than querying. The time spent on each query averaged 2 minutes and fifty-one seconds, but users spent more time looking at particular books and also viewing the results of their searches.

Activity type	Time spent
Average Session	1:51:00
Average Query	0:02:51
Average Book Viewed	0:08:04
Average Search Results Viewed	0:05:57

Table 4. Search session length

It is important to note the difference between the average session of 1 hour 51 minutes and the individual activities that occurred during the sessions; that is, most users performed several activities during their interaction with the system. Users who chose to limit their session to one single action normally viewed a single book. Of the 187 sessions, 161 of these involved viewing a book. This is not surprising since it reflects the nature of the digital library resources itself.

4.2 Query terms

The average number of terms used per query was 2.11; the maximum terms used was 8. Most users were looking for subjects, and their queries reflected looking for subjects in a direct away. For example, if someone were looking for information about "textiles" and nanotechnology, the query would be simply "textile." Queries with more terms tended to be searches for book titles, which are usually longer than one or two words.

Number of terms used in queries	Number of Queries	Percentage of Queries
1	156	40%
2	137	35%
3	53	14%
4	19	5%
5	9	2%
6	13	3%
7	3	0.76%
8	3	0.76%
Total	393	

Table 5. Number of terms in queries

Table 5 shows that approximately 60% of queries consisted of two or more terms. This finding is particularly interesting when compared with previous studies that found users tend to use very short queries. This can be explained by the very specific coverage of this digital library and, based on closer examination of queries, that nano researchers tended to start with very specific, but long, muti-term queries and then reformulated their queries if they were considered to be too restrictive in terms of retrieval. A hypothesis could be formulated to test whether interdisciplinary searchers tend to formulate longer queries than disciplinary searchers.

4.3 Search strategies

One advantage to using transaction logs is the ability to see the entire search process. After performing a query, the user receives search results from the NANOnetBASE with the following information: number of times the search string appears in the document and a list of documents (usually chapter titles) indicating book title, authors, and file size. Unfortunately, there is little context for the user to determine whether retrieved results are actually relevant without opening the document itself. No abstract or brief annotation is provided for any of the documents found. Figure 1 shows the NANOnetBASE search results screen.

The search process begins with the query. During the time period of this study, most users started with a quick search easily carried out from the homepage. It is interesting to note that, in 2005, the number of Advanced Searches and Quick Searches performed were almost equal (i.e., 94 Advanced Searches and 90 Quick Searches), but Quick Searches were far more popular in 2004 (almost twice as many) and 2006 (almost four times as many). It is also interesting that most users did not move from Quick Search to Advanced Search, but preferred to reformulate their search within the search type they started with. This may suggest that the advanced search functionality was not viewed as desirable, relevant, or intuitive to the users.

NANOnetBASE has many search functions including single character, stemming, fuzzy search, phonic search, synonym search, numeric range, and variable term weighting. These search functions, however, are not well-used by users of this database. Only 22 queries out of 393 used any advanced search functions and the majority of these just used Boolean. Table 6 provides a summary of various search strategies adopted by searchers along with the frequency of those strategies in the dataset.

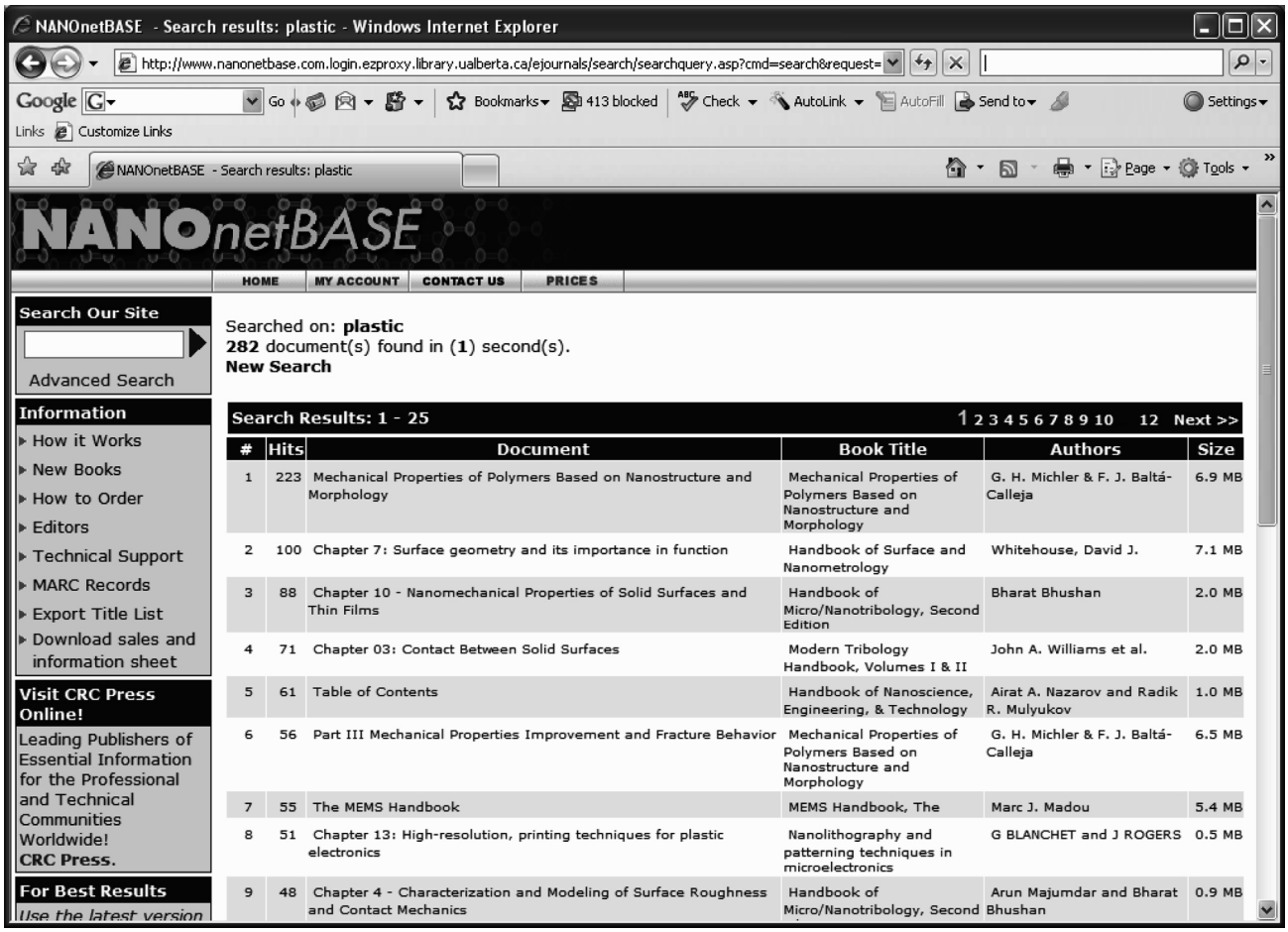

Figure 1. The NANOnetBASE search results screen

Search features	No.
Queries using advanced search functions	22
Queries using more than one advanced search function	4
Queries using 2 advanced search function	4
Queries using Boolean operators	20
Queries using AND	19
Queries using OR	1
Queries using phrase searching	4
Queries using parentheses	1
Queries using truncation	1

Table 6. Search strategies

Although the '?' character can be used to match any single character, no queries in this dataset used it. Furthermore, there were no uses of stemming, fuzzy search, phonic search, synonym search, numeric range, or variable term weighting. Among the variety of reasons users chose not to use these functions could be a lack of understanding how they work, inaccessibility of descriptions or instructions, reliance on Quick Search where those functions can be used but are not promoted in any obvious way, the simple and easy access to free text searching, or users' frustration when advanced search features retrieve no results. Interestingly, many consecutive searches revealed users searching for single and plural versions of the same word, so the use of wildcard characters or truncation has potential time-saving benefits.

4.4 Search for acronyms

About 10% of the queries made use of some form of acronym. This suggests that, as in many scientific fields, acronyms are a commonly-used and intrinsic language form within the literature of nanoscience and technology. Table 7 shows the extent of the use of acronyms in the searches.

The search function of NANOnetBASE does not handle acronyms intelligently, so users must know and use their full forms as well. If an acronym is entered, only a small number of results are retrieved because the system does not offer simultaneous full-form and ac-

Queries containing acronyms	% of 393 queries
Queries using acronyms	9.67
Queries using 1 acronym by itself	4.33
Queries using acronyms and full form	4.58
Queries using 2 acronyms	0.76

Table 7. Queries using acronyms

ronym searches to make the document selection process easier. To ensure that all relevant documents are retrieved, users must combine the acronym with its full-form equivalent using the Boolean operator OR.

4.5 Query reformulation

A key question of this research was to examine how users reformulated their queries and the reasons for doing so. Query reformulation can be interpreted in a variety of ways. For instance, adding terms to the initial search terms is considered to be one of the ways in which a query can be expanded whereas dropping any terms from the initial query terms would be considered query refinement or modification. Through the analysis of users' terms and the various ways in which they reformulated their initial queries, we identified a number of query reformulation strategies discussed in this section. Table 8 shows basic statistics about the number of total queries and the reformulated queries in the analyzed dataset. One important observation is that approximately 40% of queries were reformulated.

Query details	No. of queries
Number of sessions	552
Total number of queries	393
Total number of reformulated queries	155

Table 8. Number of Initial and reformulated queries

Table 9 shows types of reformulation strategies and the number of reformulations carried out by users. One interesting pattern was the use of phrases in searching, and many reformulated queries were, in fact, the result of phrase searching. The "major shift" category refers to a reformulation strategy where the user tries a new query related to the initial query but not by any hierarchical or semantic relationships. "Expansion" and "major shifts" queries constitute 37% of the 155 total reformulated queries. Also, 37% of reformulated queries were hierarchically or synonymously related to the initial queries.

Reasons for query reformulation	No of instances
Query changed (spelling errors)	9
Query expansion (additional terms added)	32
Major shifts	38
Query narrowed (narrower term used)	21
Query broadened (broader term used)	26
Query restructured (query terms moved around)	6
Synonymous terms used	23
Total	155

Table 9. Type and number of reformulation strategies

Figure 2 provides a graphical representation of the reasons for query reformulation and the number of instances for each reformulation reason.

One interesting observation regarding the nature of the queries submitted by nano searchers was that, of 132 terms starting with the prefix "nano," a few were very general (e.g., nanotechnogy or nanoscience) but most represented very specific query terms; for example, "nano-onion," "nonoparticle," "nanofabrication," "nanotubes," and "nanocones." Due to a lack of consistency in the ways these terms appear in different documents (e.g., sometimes hyphenated, sometimes not), some searchers were not successful at retrieving any documents using these terms if they happened to try the wrong form. Figure 2 refers to a shift to a totally different topic. Table 10 below shows examples of query reformulations.

4.6 Interdisciplinarity in queries

Another key question of this research was to examine whether disciplinary or interdisciplinary queries can be identified in the search behaviour of nano researchers. This is a complex and multidimensional question because interdisciplinarity can be interpreted from a variety of perspectives and using various methods (Palmer 1999; Morillo et al 2003). Previous research on the interdisciplinary nature of nanoscience and technology has largely focused on citation analysis and domain mapping (Meyer and Persson 1998; Meyer 2010; Zitt 2006). Here the intention is to explore whether the data allows us to address this issue from a search behaviour perspective. Some specific questions in this regard might be:

– Are there any queries that represent more than one discipline?

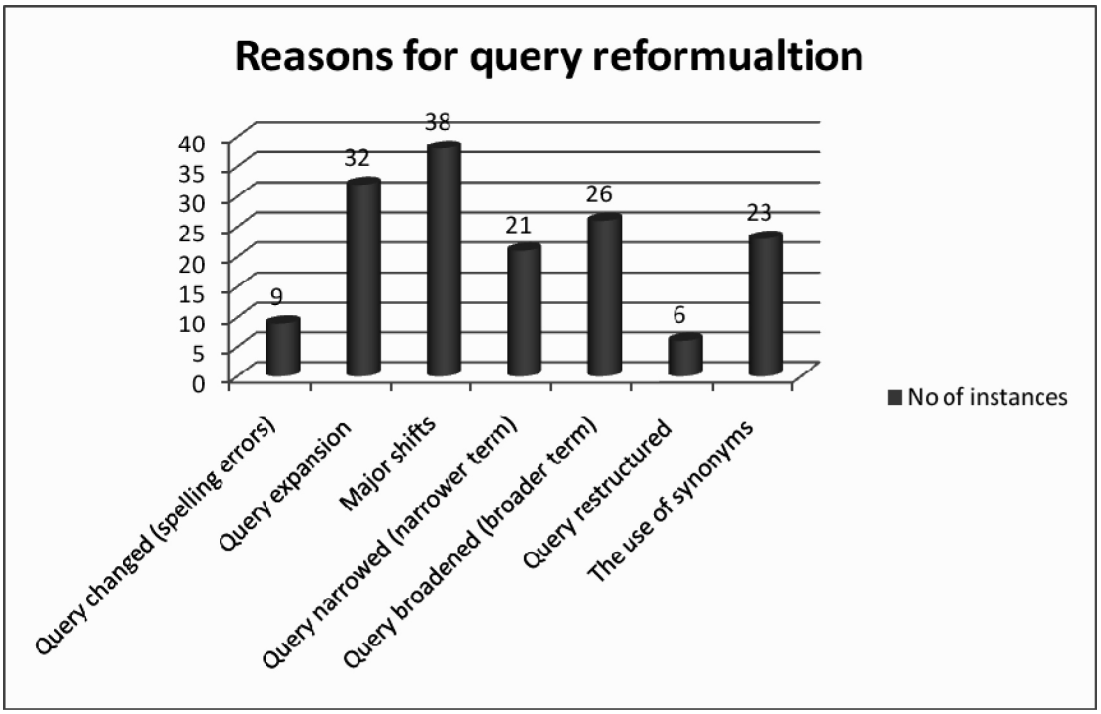

Figure 2. Query reformulation strategies

Initial query	Spelling error	Query expansion	Query modification (narrower)	Query modification (broader)	Query restructured	synonyms	Major shift
Protiein	Protein						
nanotube		Carbon nantubes					
DNA			DNA scaffolding				
Microchannel heat sink				Heat sink			
DNA wire					Wiring DNA		
Nanoparticle toxicology						Nanoparticle toxicity	
Intel							Transistor size

Table 10. Examples of query reformulation strategies

– Are there any specific characteristics associated with interdisciplinary queries?
– Are interdisciplinary queries more complex and sophisticated than disciplinary queries?

Because we did not have access to the users themselves or their research and educational backgrounds, it was not possible to base any evaluation on their perception of interdisciplinarity. It would be instructive to compare users' perception of interdisciplinarity with their actual search behaviour, their research topics, and the results of the query analysis. However, in order to analyze the data to support our analysis of the extent of interdisciplinarity in the queries, it was decided to conduct some analyses of user terms and compare them with thesauri in two commercial databases, namely INSPEC and Compendex, and then compare them to the INSPEC classification codes that represent disciplines.

4.7 Mapping user terms to INSPEC and Compendex thesauri

The first type of analysis was to map users' terms with the terms in the two thesauri attached to the INSPEC and Compendex databases. This has been reported in another paper (Shiri & Chambers 2008) showing that approximately 50% of terms entered by users matched those of the Inspec thesaurus and that around 62% of users' search terms were matched those of the Compendex thesaurus.

4.8 Mapping top terms to the INSPEC Classification codes

The second type of analysis investigated whether any element of interdisciplinarity could be observed in users' queries. To analyze this, the INPSEC thesaurus terms that were exact matches for searchers' query terms were traced to their top terms using broader term indications in the thesaurus. All the top terms were then compared with the INSPEC Classification codes, which represent discipline-based categorizations. Appendix A shows the classification structure and its treatment of various domains.

Each INSPEC top term was searched in the subject field of the INSPEC database to retrieve records for each term with their subject fields containing the top term. Following the search, details of up to 10 documents were examined to ensure that the top term was, in fact, among the first three terms assigned to each document in the database. Then, classification codes for those records were examined and recorded to analyze the disciplines and sub-disciplines they represented. The rationale for this method is that it takes into account the document collection as context for top terms; in other words, top terms associated with classification codes and the document collection formed the basis for the analysis of interdisciplinary queries.

Top terms and classification codes were entered into an MS Excel spreadsheet and analyzed (Appendix A lists the INSPEC top terms and the INSPEC Classification Codes utilized). In total, 64 unique top terms were used for searching NANOnetBASE and were compared with the INSPEC Classification Codes. Table 11 summarizes the analysis of top terms with regards to their disciplinary or interdisciplinary origins.

Terms representing disciplines	No. of top terms	%
Top terms representing one discipline	24	37.5
Top terms representing two disciplines	22	34.3
Top terms representing three or more disciplines	18	28.2
Total	64	

Table 11. Percentage of top terms representing disciplines

The top terms mapped to the INSPEC Classification Codes represent the following disciplines: A - Physics, B - Electrical Engineering & Electronics, C - Computers & Control, and E - Manufacturing & Production Engineering.

One important observation about queries representing two or three disciplines is that most had the INSPEC codes A and B for Physics and Electrical Engineering & Electronics, respectively. As Table 11 shows, approximately 62% of the top terms represent two or more disciplines. Our findings also suggest that some query terms represented a greater degree of interdisciplinarity. In other words, there may be query terms that occur in more than one discipline or are discussed, researched, and searched from a variety of perspectives. A closer examination of the top terms and some of the associated classification codes provides another interesting aspect of query interdisciplinarity. Among the terms with a higher degree of interdisciplinarity, there were such terms as "coatings," "fluidics," "control equipment," "interface phenomena," "surface treatment," and "vacuum techniques."

Another interesting observation relates to code A8 in appendix "A" or "Cross-disciplinary physics & related areas of science & technology." There are a number of top terms that have been assigned this classification code thus demonstrating another dimension of interdisciplinarity. That is, terms such as "electrochemical devices," "transistors," "disperse systems," and "electric field effects" appear in documents that show more interdisciplinarity than reflected in the codes assigned to them. Another interesting observation is that 52 out of the 64 top terms had the classification code "A" (i.e., physics) which suggests physics plays a key role in nanoscience and technology and that a majority of nanoscience and technology search queries are, in fact, associated with this discipline.

5.0 Discussion

Interdisciplinary searching in the area of nanotechnology is an interesting topic as it sheds light on the ways in which knowledge organization systems, digital libraries, portals, and other web services can be developed. The following discussion brings together and discusses our results with findings reported previously. Session length indicates the level of users' engagement and also their interaction with the system. Jones et al (2000) suggest that over half (54.34%) of user sessions they studied had a duration of five or fewer minutes, while two thirds (66.43%) had a duration of ten minutes or less. Marchionini (2002) reported that many researchers use 30 minutes as a sensible idle time interval for a session delimiter..Spink and Jansen (2004) noted that the mean session time associated with an identifier was more than two hours for some search engine datasets but most sessions lasted less than 15 minutes. Wolfram (2008) found that users were more likely to engage in extensive searching using the OPAC and specialized search system as compared to search engines and bibliographic databases. In the present study, session length was found to be in line with those from Wolfram (2008) and Spink & Jansen (2004); the average session lasted for an hour and fifty-one minutes. This finding should be viewed in light of users interacting with an electronic book digital library so a significant amount of time was spent viewing individual electronic books. It can be argued that, in the context of electronic book libraries, users generally spend more time compared to OPAC searches, search engine use, and bibliographic databases consultation.

One search pattern related to search modes found in this study was that approximately 40% of searches that were carried out in 2004 to 2006 used the advanced search mode. This may be partly explained by the NANOnetBASE interface not being flexible enough to support users' retrieval of partially relevant results in their initial searches. Furthermore, approximately 85% of searches were subject searches reflecting the importance of providing subject-based search support for users of e-book digital libraries. All of the query terms were checked against the electronic book titles that were available in the NANOnetBASE digital library. This observation is particularly important as users could easily browse the list of titles available in the library, hence the focus on subject-based searches.

Number of terms per query and query length represent search behaviour characteristics of interest to researchers. Jansen et al (2005) reported 2.35 average number of terms per query for Altavista search engine users. Wolfram (2008) found that query length varies in different web-based information retrieval environments; users tend to formulate shorter queries in bibliographic databanks or specialized search services (1.81 and 1.78 terms per query, respectively) and longer queries when using OPACs or search engines (3.66 and 2.62 terms per query respectively). In the present study, the average number of terms per query was 2.11. An interesting observation in this study was that approximately 24% of queries had three or more terms. This can be explained by users carrying out searches in an electronic book collection specific to the area of nanoscience and technology, where users could formulate very specific but sometimes long and hyphenated queries. Some queries formulated by nano researchers are multifaceted and are more sophisticated compared to queries submitted to general search engines.

In total, 22 queries made use of Boolean operators, but it must be noted that the interface does not explicitly provide search options using Boolean operators, thus partly explaining the limited use. Gerhard, Su, and Rubens (1998) note that despite keyword, Boolean, and other sophisticated search techniques making it easier to access interdisciplinary and newly emerging bodies of literature, poorly constructed keyword searching still results in the retrieval of much irrelevant material. In this study, query reformulations and consecutive searches were related to the poor performance of keyword searching. A close examination of users' queries showed that approximately 10% of their queries included some form of acronym. Because nanotechnology queries may represent multifaceted and interdisciplinary search terms, the use of acronyms should be considered one search strategies that digital libraries in this field must support. In the present study, users showed variability in their use of search terms, including American and British spellings. The matching algorithm and the interface should be designed to support users by providing cross references for variant spelling forms and acronyms, particularly the disambiguation of homographs such as 'PVD' to distinguish between Physical Vapour Deposition and Peripheral Vascular Disease.

One key search strategy that can be used to infer users' search patterns and behaviours is query reformulation. The analyzed data showed that approximately 40% (155 of 393) of the queries were subject to reformulations. Some reasons for query reformulations were spelling errors, hyphenated and combined words, narrowing the search, broadening the search, restructuring the query, and use of phrase searching

and synonyms. The reformulation reasons noted earlier indicate that nanotechnology information searchers require terminological support to carry out successful reformulations and to explore the conceptual structure of the collection. The use of multiword search terms and the order of the terms pose a challenge to nanotechnology searchers in terms of formulating queries. It is of particular importance that the interface and the search system be flexible enough to guide the user throughout the search process. Interactive term suggestion facilities based on the document collection can alleviate this problem.

Braun et al (1997) analyzed the early growth of nanoscience research during 1986-1995 and counted the occurrence of nano-prefix terms in the title of journal papers; they found 4,152 such papers during 1985-1995, which indicated a fast growth rate. Schummer (2004b) measured the growth of "nano-title-papers" in terms of annual growth rate and doubling times in different bibliographic databases and various disciplines. In the present study, approximately 132 of the search terms used had the prefix "nano," thus stressing the importance of incorporating ways to handle these terms by the search system.

McCain (1989) studied descriptor versus citation-based retrieval in the medical behavioural sciences literature and found the overlap between the results retrieved using descriptors and the results retrieved through citation-based searches was low. The use of INSPEC thesaurus terms along with INSPEC Classification Codes provided an analytical framework for this type of categorization in this study. Sixty-two percent of the 64 unique top terms resulting from mapping users' query terms to the INSPEC thesaurus represented two or more disciplines and these queries were associated with the disciplines of Physics and Electrical Engineering & Electronics. One hypothesis based on this finding is that queries related to Physics and Electrical Engineering & Electronics represent a considerable number of nanoscience and technology search queries. This finding concerning the interdisciplinarity of queries is particularly important, indicating that some queries are, in fact, more interdisciplinary than others.

6.0 Conclusion

This study was based on an analysis of transaction logs from an electronic book digital library in the area of nanoscience and technology. The results show that there are certain patterns associated with searching in an emerging and interdisciplinary area of nanotechnology. Examples of these patterns are the use of multiword and compound query terms, hyphenated terms, spelling variations, query reformulation, and the use of acronyms. One hypothesis that can be formulated based on this study is that interdisciplinary searchers tend to carry out more query reformulation and modification compared with disciplinary searchers. This hypothesis should be tested by a user-centred study, in which user contextual factors, such as task, intent, education level, and computer and information literacy could be captured. This study also demonstrated that knowledge structures such as thesauri attached to multidisciplinary databases and the INSPEC Classification Codes can contribute to the identification of interdisciplinary queries. In particular, the INSPEC and Compendex thesauri can provide interactive term suggestion facilities to support users in exploratory browsing and searching of nanotechnology information collections by providing a structure for consecutive or reformulated searches. In addition, the knowledge structures in thesauri contextualize terms and how they are used in different disciplines and subject areas. Other findings suggest that the use of acronyms and full form of terms should be accommodated to allow users to disambiguate their search terms; the user interface/system should provide opportunities to use either hyphenated or non-hyphenated words without penalizing the user by zero hits. The user terms that will be matched with the terms in the thesaurus, either exactly or partially, could be shown to the user in addition to the discipline or disciplines in which these terms are used. This way, the thesauri could be used as an interdisciplinary map of several different domains.

The NANOnetBASE user interface does not provide query reformulation or modification facilities: interdisciplinary digital library services should offer interactive query reformulation support such as term suggestions or a document collection map as well as ranking algorithms that account for the importance of query terms in various disciplinary or interdisciplinary documents. This area also calls for revisiting information organization and representation strategies based on the notion of disciplinarity. For instance, subject terms can be categorized, contextualized, and presented based on their relevance and importance in different disciplines. From a searcher's standpoint, search terms can be used to retrieve documents in a collection based on disciplinary or interdisciplinary ranking or focus. If we can prove that some queries are more interdisciplinary than others, then it should

be possible to design algorithms that match terms to documents based on disciplinary or interdisciplinary focus. The interface could then categorize and display the retrieved results based on the degree of interdisciplinarity or the dominant disciplines that have a higher degree of association with the terms searched. Furthermore, the user interface could be designed such that terms and their relationships as well as their associated disciplines be shown to the user for further exploration and examination. This type of design calls for a more enriched use of the knowledge structures in thesauri within search user interfaces, by drawing upon both thesaurus term relationships and classification categories. For instance, thesaurus terms and their broad level facets could be shown to the user along with the documents retrieved to allow users to form disciplinary and interdisciplinary views of terms and documents. Faceted browsing and exploratory search interfaces that make use of taxonomies, thesaurus terms or high level categories and facets, are becoming particularly popular. For instance, Flamenco Search Interface Project at the University of Berkeley School of Information (Hearst 2006) provides interesting and innovative ways of incorporating facets into search user interfaces to support users' navigation and exploration of information.

This study was limited to transaction log analysis and did not involve real users. Further research should examine the interdisciplinary search behaviour of real users and their perceptions and impressions of their research topics, search terms, and vocabulary problems. Vocabulary problems are considered one of the key issues in interdisciplinary searching and can be analyzed from two different perspectives. One perspective would identify the difficulties that nanoscience and technology searchers experience as they search for a particular term, while the other would examine the ways in which digital libraries and information retrieval systems in this area represent information. Further research should investigate the value and the use of co-occurrence analysis of terms in interdisciplinary areas such as nanotechnology to provide support for interdisciplinary searchers. The analytical framework developed in this study to classify queries based on their subject disciplines can be used to study other interdisciplinary domains; a particular strength of this framework is the combined use of real search terms entered by interdisciplinary searchers with thesaurus terms and classification codes to provide a solid basis for the study of query interdisciplinarity.

References

Allen, Thomas. J. 1977. *Managing the flow of technology: technology transfer and the dissemination of technological information within the R&D organization.* Cambridge, MA: MIT Press.

Bassecoulard, Elise 2007. Mapping nanosciences by citation flows: a preliminary analysis. *Scientometrics* 70: 859-80.

Bates, Marcia. J. 1996. Learning about the information seeking of interdisciplinary scholars and students. *Library trends* 45n2: 155-64.

Blecic, Deborah., Bangalore, Nirmala S., Dorsch, Josephine L., Henderson, Cynthia L., Koenig, Melissa H., and Weller, Ann C. 1998. Using transaction log analysis to improve OPAC retrieval results. *College and research libraries* 59: 39-50.

Boyack, Kevin W., Wylie, Brian N., and Davidson, George S. 2002. Domain visualization using VxInsight for science and technology management. J*ournal of the American Society for Information Science and Technology* 53: 764-74.

Braun, Tibor., Schubert, A., and Zsindely, Sandor. 1997. Nanoscience and nanotechnology on the balance. *Scientometrics* 38: 321-25.

Brown, Ceceila M. 1999. Information-seeking behaviour of scientists in the electronic information age: astronomers, chemists, mathematicians, and physicists. *Journal of the American Society for Information Science* 50: 929-43.

Calero, Clara, Buter, Renald, Valoes, Ceclia. C., and Noyons, Ed. 2006. How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics* 66: 365-76.

Chau, Michael, Huang, Zan., Qin, Jialun, Zhou, Yilu, and Chen, Hsinchun. 2006. Building a scientific knowledge web portal: The NanoPort experience. *Decision support systems* 42: 1216-38.

Connaway, Lynn S., 2005, February. Transaction log analysis of electronic (e-book) usage. *Against the grain* 85-89. Available http://www.oclc.org/research/publications/archive/2005/connaway-snyder-atg.pdf

Cothey, Vivian. 2002. A longitudinal study of World Wide Web users' information-searching behavior. *Journal of American Society for Information Science and Technology* 53: 67-78.

Curry, Ann. 2005. Reaching beyond the reference books in public libraries. *International journal of web based communities (IJWBC)* 1: 450-63.

Davis, Philip. M., and Solla, Leah. 2003. An IP-Level analysis of usage statistics for electronic journals in

chemistry: making inferences about user behaviour. *Journal of the American Society for Information Science and Technology* 54: 1062-8.

Eason, Ken, Richardson, Sue, and Yu, Liangzhi. 2000. Patterns of use of electronic journals. *Journal of documentation* 56: 477-504.

EI thesaurus description 2008. Available http://www.ei.org/databases/other.html.

Ellis, David, and Haugan, Merete. 1997. Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of documentation* 53: 384-403.

Eto, Hajime. 2003. Interdisciplinary information input and output of a nano-technology project. *Scientometrics* 58: 5-33.

Foster, Allan. 2004 A nonlinear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology* 55: 228-37.

Gerhard, Kristin H., Su, Mila C., & Rubens, Charlotte C. 1998. An empirical examination of subject headings for women's studies core materials. *College and research libraries* 59: 130-38.

Ghaphery, Jimmy. 2005. Too quick? Log analysis of quick links from an academic library website. *OCLC systems & services* 21n3: 148-55.

Haythornthwaite, Caroline. 2006. Learning and knowledge networks in interdisciplinary collaborations. *Journal of the American Society for Information Science and Technology* 57: 1079-92.

Hearst, Marti. 2006. Design recommendations for hierarchical faceted search interfaces. In *Association for Computing Machinery Special Interest Group on Information Retrieval (SIGIR) Workshop on Faceted Search*, pp. 26-30.

Hertzum, Morten, and Pejtersen, Annelise M. 2000. The information-seeking practices of engineers: searching for documents as well as for people. In*formation processing & management* 36: 761-78.

Hughes, Carol A., and Buchanan, Nancy L. 2001. Use of electronic monographs in the humanities and social sciences. *Library hi tech* 19n4: 368-375.

INSPEC User Guide (n.d.). Available http://www2.theiet.org/publish/support/inspec/document/UserG/UserGuide.pdf

Jamali, Hamid R., Nicholas, David, and Huntington, Paul. 2005. The use and users of scholarly e-journals: A review of log analysis studies. *Aslib Proceedings* 57: 554-71.

Jansen, Bernard J. 2006. Search log analysis: What it is, what's been done, how to do it. *Library & information science research* 28: 407-32.

Jansen, Bernard J., and Pooch, Udo. 2001. Web user studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology* 52: 235–46.

Jansen, Bernard J., and Spink, Amanda. 2005. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information processing & management* 42: 248–63.

Jansen, Bernard J., Spink, Amanda, and Pedersen, Jan. 2005. A temporal comparison of AltaVista Web searching. *Journal of the American Society for Information Science and Technology* 56: 559-70.

Jones, Steve, Cunningham, Sally Jo, McNab, Rodger, and Boddie, Stefan. 2000. A transaction log analysis of a digital library. *International journal on digital libraries* 3: 152-69.

Joseph, Lura E. 2001. Information-seeking and communication behavior of petroleum geologists. *Science & technology libraries* 21n3/4: 46-62.

Ke, Hao-Ren, Kwakkelaar, Rolf, Tai, Yu-Min, and Chen, Li-Chun. 2002. Exploring behaviour of e-journal users in science and technology: Transaction log analysis of Elsever's ScienceDirect OnSite in Taiwan. *Library & information science research* 24: 265-91.

King, Donald W., Caste, J., and Jones, H. 1994. *Communication by engineers: a literature review of engineers' information need, seeking processes, and use.* Washington, DC: Council on Library Resources.

Klein, Julie T. 1996. Interdisciplinary needs: The current context. *Library trends* 45n2: 134–54.

Kutner, Laurie A. 2000. Library instruction in an interdisciplinary environmental studies program: Challenges, opportunities, and reflections. *Issues in science and technology librarianship* 28. Available http://www.library.ucsb.edu/istl/00-fall/ .

Kwasitsu, Lishi. 2003. Information-seeking behaviour of design, process, and manufacturing engineers. *Library & information science research* 25: 459-76.

Lau, Eng P., and Goh, Dion H. 2006. In search of query patterns: A case study of a university OPAC. *Information processing & management* 42: 1316-29.

Leckie, Gloria J., Pettigrew, Karen E., and Sylvain, Christian. 1996. Modelling the information-seeking of professionals: a general model derived from research on engineers, health care professionals, and lawyers. *Library quarterly* 66: 161-93.

Leydesdorff, Loet. 2007. "Betweenness centrality" as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology* 58: 1303-19.

Llull, Harry. 1991. Meeting the academic and research information needs of scientists and engineers in the university environment. *Science and technology libraries* 11n3: 83-90.

Mahoui, Malika, and Cunningham, Sally. J. 2000. A comparative transaction log analysis of two computing collections. *Lecture notes in computer science* 1293: 418-23.

Marchionini, Gary. 2000. Evaluating digital libraries: A longitudinal and multifaceted view. *Library trends* 49n2: 304-33.

Marchionini, Gary. 2002. Co-evolution of user and organizational interfaces: A longitudinal case study of WWW dissemination of national statistics. *Journal of the American Society for Information Science and Technology* 53: 1192-1209.

Mat-Hassan, Mazlita, and Levene, Mark. 2005. Associating search and navigation behavior through log analysis. Journa*l of the American Society for Information Science and Technology* 56: 913-34.

McCain, Katherine. W. 1995. Biotechnology in context: A database-filtering approach to identifying core and productive non-core journals supporting multidisciplinary R&D. *Journal of the American Society for Information Technology* 46: 306-17.

McCain, Katherine W. 1989. Descriptor and citation retrieval in the medical behavioral sciences literature: Retrieval overlaps and novelty distribution. *Journal of the American Society for Information Science* 40: 110-14.

Meyer, Martin S. 2000. Patent citations in a novel field of technology: What can they tell about interactions between emerging communities of science and technology? *Scientometrics* 48: 151-78.

Meyer, Martin. S. 2001. Patent citation analysis in a novel field of technology: An exploration of nanoscience and nano-technology. *Scientometrics* 51: 163-83.

Meyer, Martin S., and Persson, O. 1998. Nanotechnology - interdisciplinarity, patterns of collaboration and differences in application. *Scientometrics* 42: 195-205.

Millsap, Larry, and Ferl, Terry. 1993. Search patterns of remote users: An analysis of OPAC transaction logs. *Information technology and libraries* 11: 321-43.

Milojević, Stasa. 2007. Big science, nano science? Mixed method approach to mapping the evolution and structure of nanotechnology. *Information research* 12n4 poster colisp09. Available http://Information R.net/ir/12-4/colis/colisp09.html

Morillo, Fernanda, Bordons, Maria, and Gomez, Isabel. 2003. Interdisciplinarity in science: a tentative typology of disciplines and research areas. *Journal of the American Society for Information Science & Technology* 54: 1237-49.

Mote, L. J. B. 1962. Reasons for the variations in the information needs of scientists. J*ournal of documentation* 18: 169-75.

Murphy, Janet. 2003. Information seeking habits of environmental scientists: A study of interdisciplinary scientists at the environmental protection agency in Research Triangle Park, North Carolina. *Issues in science and technology librarianship* 38. Available www.istl.org/03-summer/refereed.html.

Nicholas, David, Huntington, Paul, Jamali, Hamid R., and Tenopir, Carol. 2006a. What deep log analysis tells us about the impact of big deal, Case study OhioLink. *Journal of documentation* 62: 482-508.

Nicholas, David, Huntington, Paul, Jamali, Hamid R., and Tenopir, Carol. 2006b. Finding information in (very large) digital libraries: A deep log approach to determining differences in use according to method of access. *Journal of academic librarianship* 32: 119-26.

Nicholas, David., Huntington, Paul, and Watkinson, Anthony. 2005. Scholarly journal usage: The results of deep log analysis. *Journal of documentation* 61: 246-80.

Packer, Katherine H., and Soergel, Dagobert. 1979. The importance of SDI for current awareness in fields with severe scatter of information. *Journal of the American Society for Information Science* 30: 125-35.

Palmer, Carole L. 2006. Weak information work and 'doable' problems in interdisciplinary science. In Grove, Andrew, ed., *Information realitites: shaping the digital future for all--Proceedings of the 69th ASIS&T Annual Meeting, 3-8 November 2006, Austin, Texas*. Available: http://onlinelibrary.wiley.com/doi/10.1002/meet.14504301108/abstract (accessed 7 January 2011).

Palmer, Carole, L. & Neumann, Laura J. 2002. The information work of interdisciplinary humanities scholars: Exploration and translation. *The library quarterly* 72: 85-114.

Palmer, Carole L. 1999. Structures and strategies of interdisciplinary science. *Journal of the American Society for Information Science* 50: 242–53.

Palmer, Carole L. 2005. Scholarly work and the shaping of digital access. *Journal of the American Society for Information Science and Technology* 56: 1140-53.

Palmer, Carole L., Cragin, Melissa H., and Hogan, Timothy P. 2007. Weak information work in scien-

tific discovery. *Information processing & management* 43: 808-20.

Peters, H. P. F., Braam, R. R., and van Raan, A. F. J. 1995. Cognitive resemblance and citation relations in chemical engineering publications. *Journal of the American Society for Information Science* 46: 9-22.

Peters, Thomas. 1993. The history and development of transaction log analysis. *Library hi tech* 42: 41-66.

Pierce, Sydney J. 1999. Boundary crossing in research literatures as a means of interdisciplinary information transfer. *Journal of American Society for Information Science* 50: 271-78.

Pinelli, Thomas E. 1991. The information-seeking habits and practices of engineers. *Science & technology libraries* 11: 5-25.

Poland, Jean M. 1991. Informal communication among scientists and engineers. *Science & technology libraries* 11: 61-73.

Qin, Jian, Lancaster, Fredrick W. and Allen, Bryce. 1997. Types and Levels of Collaboration in Interdisciplinary Research in the Sciences. *Journal of the American Society for Information Science* 48n: 893-916.

Qin, Jialun, Zhou, Yilu, and Chau, Michael. 2004. Building domain-specific web collections for scientific digital libraries: a meta-search enhanced focused crawling method. *Joint Conference on Digital Libraries (JCDL) 2004*: 135-41. Available http://www.informatik.uni-trier.de/~ley/db/conf/jcdl/jcdl2004.html#QinZC04

Sanderson, Mark and Dumais, Susan. 2007. Examining repetition in user search behavior. In Amati, Giambattista, Carpineto, Claudio and Romano, Giovanni, eds., *Advances in information retrieval, 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007, Proceedings. Lecture Notes in Computer Science* 4425, Springer 2007, pp. 597-604.

Schummer, Joachim. 2004a. Interdisciplinary issues of nanoscale research. In Baird, Davis, Nordmann, Alfred and Schummer, Joachim eds., *Discovering the nanoscale.* Amsterdam: IOS Press, pp. 9-20.

Schummer, Joachim. 2004b. Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics* 59: 425-65.

Shiri, Ali, and Chambers, Thane. 2008. Information retrieval from digital libraries: Assessing the potential utility of thesauri in supporting users' search behaviour in an interdisciplinary domain. In Arsenault, Clément and Tennis, Joseph T., eds., *Culture and identity in knowledge organization: Proceedings of the 10th international conference for knowledge organization, Montreal, Quebec August 5-8, 2008*. Advances in knowledge organization 11. Würzburg: Ergon-Verlag, pp. 184-89.

Spanner, Don. 2001. Border crossings: Understanding the cultural and informational dilemmas of interdisciplinary scholars. *The journal of academic librarianship* 27: 352-60.

Spink, Amanda, and Jansen, Bernard J. 2004. *Web search: public searching of the Web.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Steele, Thomas W., and Stier Jeffrey C. 2000. The impact of interdisciplinary research in the environmental sciences: A forestry case study. *Journal of the American Society for Information Science* 51: 376-484.

Tanaka, Michiko. 2005. Toward a proposed ontology for nanoscale research. In Vaughan, Liwen, ed., *Data, information, and knowledge in a networked world: Proceedings of the Canadian Association for Information Science annual conference June 2-4 2005*. Available: http://www.cais-acsi.ca/search.asp?year=2005 (accessed 7 January 2011).

Troll Covey, Denise. 2002. Usage and usability assessment: library practices and concerns. Council on Library and Information Resources, Washington DC. Available www.clir.org/pubs/reports/pub105/pub105.pdf.

Weisgerber, David W. 1993. Interdisciplinary searching - problems and suggested remedies - A report from the ICSTI group on interdisciplinary searching. *Journal of documentation* 49: 231-54.

Wen, Ji -Rong Nie, Jian. -Yun, and Zhang, Hong. 2002. Query clustering using user logs. *ACM transaction on information systems* 20: 59-81.

White, Howard. 1996. Literature retrieval for interdisciplinary syntheses. *Library Trends* 45n2: 239-64.

Wolfram, Dietmar. 2008. Search characteristics in different types of Web-based IR environments: Are they the same? *Information processing & management* 44: 1279-92.

Zitt, Michael, and Bassecoulard, Elise. 2006. Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information processing & management* 42: 1513-31.

Zuccala, Alesia, Thelwall, Mike., Oppenheim, Charles, and Dhiensa, Rajveen. 2007. Web intelligence analyses of digital libraries: A case study of the National electronic Library for Health (NeLH). *Journal of documentation* 63: 558-89.

Appendix A.
Outline of the INSPEC Classification Codes

<table>
<tr>
<td>A – Physics
A0 General
A1 The physics of elementary particles & fields
A2 Nuclear physics
A3 Atomic & molecular physics
A4 Fundamental areas of phenomenology
A5 Fluids, plasmas & electric discharges
A6 Condensed matter: structure, thermal & mechanical properties
A7 Condensed matter: electronic structure, electrical, magnetic, & optical properties
A8 Cross-disciplinary physics & related areas of science & technology
A9 Geophysics, astronomy Astrophysics</td>
<td>B - Electrical Engineering & Electronics
B0 General topics, engineering mathematics & materials science
B1 Circuit theory & circuits
B2 Components, electron devices & materials
B3 Magnetic & superconducting materials & devices
B4 Optical materials & applications, electro-optics & optoelectronics
B5 Electromagnetic fields
B6 Communications
B7 Instrumentation & special applications
B8 Power systems & applications</td>
</tr>
<tr>
<td>C- Computers & Control
C0 General & management topics
C1 Systems & control theory
C3 Control technology
C4 Numerical analysis & theoretical computer topics
C5 Computer hardware
C6 Computer software
C7 Computer applications</td>
<td>D - Information Technology
D1 General & management aspects
D2 Applications
D3 General systems & equipment
D4 Office automation – communications
D5 Office automation – computing</td>
</tr>
<tr>
<td colspan="2">E- Manufacturing & Production Engineering
E0 General topics in manufacturing & production engineering
E1 Manufacturing & production
E3 Industrial sectors</td>
</tr>
</table>

Appendix B.
INSPEC top terms mapped to the INSPEC Classification Codes

INSPEC Top terms	INSPEC classification codes
chemical reactions	A3, A8, B, C
coatings	A8160H, A6150J, A6116D, B0550
collision processes	A5, A3
computation theory	A,B, C
computer software	A,B, C, E
conductors (electric)	B3, A, C
control equipment	B, C, E
convertors	B0, B2, B7, C
direct energy conversion	B0, B8, C, E
disperse systems	A6, A8
electric field effects	A6, A7, A8
electric properties	A6, A7, A8
electrochemical devices	A6, A8, B
electron tubes	A4, B, E
elementary particle interactions	A1
elementary particle theory	A0,A1
elementary particles	A1
environmental factors	B0
films	A1
fluidics	A, B, E
fluids	A, B, E
heat transfer	A, B
imaging	C, B
inorganic compounds	A, B, E
instrumentation	A,C
insulation	A,B
interface phenomena	A, B, E
magnetic field effects	A,B, C
magnetism	A
materials	A6
mathematics	A, B, C
measurement	A, A8, B, E
mechanics	A,B, C, E
medicine	A8, B
metals	A8, A6
micromechanical devices	A, B, A8
molecular electronics	B
molecules	A8
networks (circuits)	B, C
optics	A, B
organic compounds	A8

INSPEC Top terms	INSPEC classification codes
production	E, A
pumps	A, B
quantum interference devices	A, B
radio applications	B, A
semiconductor devices	A, B
spectra	A
surface phenomena	A, B, A8
surface treatment	A,A8, B, E
switchgear	B
switching	B
telecommunication	B, C
thermodynamics	A
transistors	B, A8
transmission lines	B
transport processes	A
tunneling	A
vacuum (elementary particles)	A
vacuum apparatus	A, B
vacuum microelectronics	A, B
vacuum techniques	A, B, E
vapour deposition	A, B
waves	A
wiring	E, A,B