Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

Exploration of Interdisciplinarity in Nanotechnology Queries: The Use of Transaction Log Analysis and Thesauri

Ali Shiri
School of Library and Information Studies, University of Alberta, Edmonton, Alberta
ashiri@ualberta.ca

Abstract
Nanoscience and technology is characterized by nano researchers as 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 information professionals involved in organizing and providing the 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. This demand emphasizes the importance of thinking about and developing methodological models for investigating interdisciplinary knowledge organization practices. This 2008 study examined the extent of interdisciplinarity in user queries submitted to the NANOnetBASE digital library. Transaction logs of the digital library were analyzed to explore users' search behaviour patterns and to examine the extent to which user queries were interdisciplinary. The Inspect thesaurus and Classification codes were utilized as the disciplinary or interdisciplinary focus of the queries. The results 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 contribute to the development of more critical information organization and classification practices in such an increasingly interdisciplinary domain as nanoscience and technology.

Subject terms: digital libraries, interdisciplinarity, knowledge organization, information retrieval

Introduction and context

The interdisciplinary nature of nanoscience and technology poses challenges of efficient searching and retrieval of pertinent information from a wide range of electronic sources. These challenges affect both nano researchers and information professionals who provide library and information services to them. To create better information search services and systems and to properly organize and represent nanotechnology information, it is essential to explore and identify problems and issues associated with interdisciplinary searching in the context of nanotechnology. At present, there is very little research to inform the design and development of information services, knowledge organization systems, user interfaces and digital libraries that can accommodate interdisciplinary searching in nanotechnology. Bates (1996) argues that studying information seeking in interdisciplinary fields can inform us of the needs and problems of people in those fields and notes that interdisciplinary researchers constitute a significant and distinctive class of scholars, deserving of research on their information-seeking behavior.

The literature on the information search behaviour of nanoscience and technology information searchers is scarce. 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 Presson, 1998), mapping and visualization of the area of nanoscience and technology (Boyack et al., 2002), and identifying core journals in the area of nanotechnology (Leydesdorff, 2007). Citation analysis studies have also been conducted to show the interdisciplinarity of other scientific areas such as medical and behavioural sciences McCain, 1995), environmental sciences (Steele, 2000), and chemical engineering (Boyack et al., 2002). Bates (1996) argues that studying researcher information seeking in interdisciplinary fields may tell us not only about

ISSN: 2324-9773

Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

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, 2006; Foster, 2004; Weisgerber, 1993).

Transaction log analysis has been employed to study users of web search engines (Jansen and Pooch, 200; Wen et al., 2002), OPACs (Lau et al., 2006), library websites (Ghaphery, 2005), digital libraries (Jones et al., 2000; Zuccala and Thelwall, 2007), and electronic journals (Nicholas et al., 2005). Jansen (2006) presents transaction log analysis as a methodological framework along with strengths and limitations of this method in research.

The objective of this study was to examine the extent to which queries formulated by nanoscience and technology researchers are interdisciplinary. In particular, the aim was to investigate the disciplinary foci of user queries through the analysis of users transaction logs and their comparison with well-established thesauri covering the area of nanoscience and technology.

Methodological framework

Various methodological considerations have been taken into account to carry out this study. In particular, the analysis was based on Jansen's (2006) transaction data analysis method including term level, query level and session level analyses. Furthermore, two thesauri, namely Inspec and Compendex were utilized to carry out the mapping of users' query terms. This analysis was conducted to explore the potential value of these thesauri in supporting users' query formulation. To this end, the following mapping situations were taken into consideration: exact match, partial match and broader/narrower term match. For the analysis of data relating to the mapping of users' query terms to thesauri, the methodology used in Shiri and Revie (2004) was adopted where various types of mappings between users' query terms and thesaurus terms were established. In addition to the above thesauri, the Inspec Classification Codes were used to map the disciplines represented by the query terms. The Inspec terms that were found to be exact matches for searcher' 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.

Research questions

- To what extent are thesauri capable of providing a basis for identifying interdisciplinary queries in the area of nanoscience and technology?
- Can disciplinary and interdisciplinary queries be identified? What subject areas are covered by the queries?

The System: NanoNetBase E-book Digital Library

NANOnetBASE is an e-book digital library for nanotechnology and nanoscience researchers. The full-text database consists of 45 titles that can be accessed in a variety of ways. The library provides a variety of search tools, such as Boolean operators, truncation, wildcard, stemming, fuzzy searches, field searching, phonic search, synonym search, and variable term weighting.

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 research university with strong nanoscience and technology research profile where more than 140 nano researchers are currently

Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

active. They include faculty members as well as graduate students and post-doc researchers. In total, 1921 transactions were analyzed.

Analysis methods

Jansen's (2006) transaction data analysis method was utilized to conduct term level, query level and session level analyses. Part of this study investigated the potential utility of thesauri in supporting users' search behaviour in nanotechnology (Shiri and Chambers, 2008). In this study, the Inspec thesaurus and Classification Codes were used to map the disciplines represented by the query terms. The Inspec terms that were found to be exact matches for users' 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 and a select number of documents in the INSPEC database to evaluate their extent of interdisciplinarity.

Results

One of the key questions of this research was to examine whether disciplinary or interdisciplinary queries can be identified in the search behaviour of the nano researchers. This is a complex and multidimensional question as interdisciplinarity can be interpreted from a variety of perspectives and using various methods. Previous research on the interdisciplinary nature of nanoscience and technology has largely focused on citation analysis and domain mapping. Here the intention is to explore if the data allows us to address this issue from a search behaviour perspective.

Query Terms

The average number of terms used per query was 2.11 and the largest number of terms used was 8. Most of the users were looking for subjects and their queries looked for subjects in a direct away. For example, if someone was looking for information about how 'textiles' are used in nanotechnology, the query would simply be *textile.* A few of the larger queries tended to be searches for book titles; which usually are not restricted to one or two words.

Table 1. Number of terms in queries

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

ISSN: 2324-9773

Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

Total	393	

The data in Table 1 shows that around 60% of the queries consisted of two or more terms. This finding is particularly interesting when it is compared with previous studies that found users tend to use very short queries. This can be explained by the fact that the digital library is very specific in its coverage and a closer examination of the queries shows that nano researchers tended to start with very specific, but long muti-term queries and reformulated their queries if they were considered to be too restrictive in terms of retrieval. In order to assess the extent to which users' terms mapped to the thesaurus, all terms entered by users were analyzed and their mapping situations are reported here. A brief description of different match types is provided below with a range of examples shown in Table 2.

Table 2. Examples of different types of term match in INSPEC and COMPENDEX

Match type	**User term**	**INSPEC thesaurus term**	**COMPENDEX thesaurus term**
Exact match	Vacuum	Vacuum (elementary particles)	Vacuum
	Carbon nanotube	Carbon nanotubes	Carbon nanotubes
	Sputtering	Sputtering	Sputtering
Partial match	Microchannel heat sink	Microchannel flow	Heat sinks microchannels
Narrower term match	Magnetic film	Magnetic thin film	Magnetic thin film
	Tunnelling	---	Electron tunnelling
	Photonics	Photonic switching systems Photonic band gap	---
Broader term match	DNA scaffolding	DNA	DNA
	vapour deposition stress	vapour deposition	vapour deposition

The Compendex thesaurus had exact matches for 49% of users' search terms compared to 37% exact matches offered by INSPEC. This is particularly interesting as details and narrower terms of exact match terms can be further explored by users to find out more about the knowledge map of that particular subject area and can also choose a wider variety of terms offered by these thesauri. A closer look at all the match types indicates that in general the Compendex thesaurus provides a better coverage for nanoscience and technology user terms. This is not to say that the Compendex thesaurus should be used instead of the INSPEC thesaurus. In fact, our observations of the nature of terms suggested by both thesauri show that there are terms whose exact or partial matches can only be found in INSPEC and not in Compendex.

Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

These observations suggest that both thesauri should be used in order to support users' query formulation and expansion behaviours.

Mapping Top Terms to the Inspec Classification Codes

The second type of analysis of the queries was conducted to investigate whether any element of interdisciplinarity can be observed in users' queries. The analytical method for this part was as follows. The Inspec thesaurus terms that were found to be exact matches for searcher' 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. Table 3 shows the classification structure and its treatment of various domains.

Table 3. Outline of the Inspec Classification

A – Physics	***B - Electrical Engineering & Electronics***
A0 General	B0 General topics, engineering mathematics & materials science
A1 The physics of elementary particles & fields	B1 Circuit theory & circuits
A2 Nuclear physics	B2 Components, electron devices & materials
A3 Atomic & molecular physics	B3 Magnetic & superconducting materials & devices
A4 Fundamental areas of phenomenology	B4 Optical materials & applications, electro-optics & optoelectronics
A5 Fluids, plasmas & electric discharges	B5 Electromagnetic fields
A6 Condensed matter: structure, thermal & mechanical properties	B6 Communications
A7 Condensed matter: electronic structure, electrical, magnetic, & optical properties	B7 Instrumentation & special applications
A8 Cross-disciplinary physics & related areas of science & technology	B8 Power systems & applications

ISSN: 2324-9773

doi: 10.7152/acro.v20i1.12883

A9 Geophysics, astronomy Astrophysics	
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	***D - Information Technology*** D1 General & management aspects D2 Applications D3 General systems & equipment D4 Office automation – communications D5 Office automation – computing
E- Manufacturing & Production Engineering E0 General topics in manufacturing & production engineering E1 Manufacturing & production E3 Industrial sectors	

Each of the Inspec top terms was searched in the subject field of the Inspec database. This process retrieved a number of records for each term with their subject fields containing the top term. Following the search, details of a maximum of ten documents were examined to ensure that the top term is in fact among the first three terms assigned to each document in the database. Then, classification codes of those records were examined and recorded to analyze the disciplines and sub-disciplines that those classification codes represented. The rationale for this method lies in the fact that it takes into account the document collection as context for those top terms. In other words, top terms in association with classification codes and the document collection formed the basis for the analysis of interdisciplinary queries.

ISSN: 2324-9773

Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

Top terms and classification codes were all entered into an Excel spreadsheet and were analyzed. In total, there were 64 unique top terms that were used for searching NANOnetBASE and were compared with the Inspec classification codes. Table 4 summarizes the analysis of top terms with regards to their disciplinary or interdisciplinary origins as is reflected in the Inspec classification codes.

Table 4. Percentage of top terms representing disciplines

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	

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 the queries representing two or three disciplines is that a majority of them had the Inspec codes A and B, namely Physics and Electrical Engineering & Electronics. As Table 4 depicts, around 62% of the top terms represent two or more disciplines. The finding also suggests that some of the query terms represented a greater degree of interdisciplinarity. In other words, there may be query terms that occur in a 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'. Also, an interesting observation is related to the code A8 in table 3 which represents '*Cross-disciplinary physics & related areas of science & technology'*. There are a number of top terms that have been assigned this classification code which further demonstrates another dimension of interdisciplinarity. For instance, terms such as ' electrochemical devices', ' transistors', 'disperse systems' and 'electric field effects'. All of these terms appear in the documents that show more interdisciplinarity as reflected in the codes assigned to them. Another interesting observation regarding the disciplines represented by the query terms is that 52 out of the 64 top terms had the classification code 'A', which represents the area of physics. This finding is particularly important as it suggests that physics play a key role in nanoscience and technology and that a majority of nanoscience and technology search queries are, in fact, associated with this discipline.

Discussion and conclusion

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. This study also demonstrated that thesauri, attached to multidisciplinary databases, have the potential to contribute to the query formulation and reformulation carried out by nano researchers and to

the identification of interdisciplinary queries. In particular, the Inspec and Compendex thesauri can be used to provide interactive term suggestion facilities and to support users in exploratory browsing and searching of nanotechnology information collections. For instance, providing a browsing structure for consecutive or reformulated searches will assist users in their search process. In addition, the knowledge structures in thesauri provide a context for a given term and how it occurs in different disciplines and subject areas. The interface to an interdisciplinary digital library such as the NANOnetBase should offer users interactive query reformulation support, such as term suggestions or a map of the document collection as well as ranking algorithms that take into account 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, for example, search terms entered by nano researchers 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 will then be able to 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. This study was limited to transaction log analysis and did not involve real users. Research should examine the interdisciplinary search behaviour of real users and their perceptions and impressions of their research topics, search terms and vocabulary problems.

References

Bates, M. J. (1996). Learning about the information seeking of interdisciplinary scholars and students. *Library Trends, 45*(2), 155-164.

Braun, T.; Schubert, A. & Zsindely, S. (1997). Nanoscience and nanotechnology on the balance. *Scientometrics*, 38(2), 321-325.

Foster, A. (2004) A Nonlinear Model of Information-Seeking Behavior. *Journal of the American Society for Information Science and Technology, 55(3),* 228-237.

Ghaphery, J. (2005) Too quick? Log analysis of quick links from an academic library website. *OCLC Systems & Services, 21*(3), 148-155.

Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research, 28*(3), 407-432.

Jansen, B. J., & Pooch, U. (2001). Web user studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology, 52*(3), 235−246.

Jansen, B. J., & Spink, A. (2005). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management, 42*(1), 248−263.

Jones, S., Cunningham, S., & McNab, R. (2000). A transaction log analysis of a digital library. *International Journal on Digital Libraries, 3*, 152-69.

Kutner, L. A. (2000). Library instruction in an interdisciplinary environmental studies program: Challenges, opportunities, and reflections. *Issues in Science and Technology Librarianship*, (28). Retrieved July 30, 2008, http://www.library.ucsb.edu/istl/00-fall/ .

Lau, E. P., & Goh, D. H. (2006). In search of query patterns: A case study of a university OPAC. *Information Processing and Management, 42*(5), 1316-1329.

Leydesdorff, L. (2007). "Betweenness centrality" as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology, 58*(9), 1303-1319.

Marchionini, G. (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.

McCain, K. 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(4), 306-317.

Meyer, M., & Persson, O. (1998). Nanotechnology - interdisciplinarity, patterns of collaboration and differences in application. *Scientometrics, 42*(2), 195-205.
Boyack, K. W., Wylie, B. N., & Davidson G. S. (2002). Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology, 53*(9), 764-774.

Nicholas, D., Huntington, P., & Watkinson, A. (2005) Scholarly journal usage: The results of deep log analysis. *Journal of Documentation*, *61*(2), 246-280.

Palmer, C. (2006). Weak information work and 'doable' problems in interdisciplinary science. Proceedings of the 69th ASIS&T Annual Meeting, 3-8 November 2006, Austin, Texas.

Peters, H. P. F., Braam, R. R., & van Raan, A. F. J. (1995). Cognitive resemblance and citation relations in chemical engineering publications. *Journal of the American Society for Information Science, 46*(1), 9-22.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics, 59*(3), 425-465.

Shiri, A.; Chambers, T.(2008) Information Retrieval from Digital Libraries: Assessing the Potential Utility of Thesauri in Supporting Users' Search Behaviour in an Interdisciplinary Domain. In: *Proceedings of the 10th International Conference of the International Society for Knowledge Organization (ISKO)*. Montreal, 5-8 August 2008.

Spink, A. & Jansen, B.J. (2004). *Web search: Public searching of the Web*, Dordrecht, the Netherlands: Kluwer Academic Publishers.

Shiri, A. (2009). Exploration of interdisciplinarity in nanotechnology queries: The use of transaction log analysis and thesauri. 20th Annual ASIS SIG/CR Workshop: Bridging Worlds, Connecting People: Classification Transcending Boundaries
doi: 10.7152/acro.v20i1.12883

Gerhard, K. H., Su, M. C., & Rubens, C. C. (1998). An empirical examination of subject headings for women's studies core materials. *College and Research Libraries,59*(2), 130-138.

Steele, T. W., & Stier J. C. (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. *Journal of the American Society for Information Science, 51*(5), 376-484.

Weisgerber, D. W. (1993). Interdisciplinary searching - problems and suggested remedies - A report from the ICSTI group on interdisciplinary searching. *Journal of Documentation, 49*(3), 231-254.

Wen, J. -R, Nie, J. -Y, & Zhang, H. (2002). Query clustering using user logs. *ACM Transaction on Information Systems, 20*(1), 59-81.

Wolfram, D. (2008). Search characteristics in different types of Web-based IR environments: Are they the same? *Information Processing and Management*, 44(3), 1279-1292.

Zuccala, A., Thelwall, M., Oppenheim, C. & Dhiensa, R. (2007). Web intelligence analyses of digital libraries: A case study of the National electronic Library for Health (NeLH). *Journal of Documentation. 63(4), 558-589.*

ISSN: 2324-9773