

An Informetric Analysis on Intellectual Structures with Multiple Features of Academic Library Research Papers*

복수 자질에 의한 지적 구조의 계량정보학적 분석연구:
국내 대학도서관 분야 연구논문을 대상으로

Sanghee Choi**

ABSTRACT

The purpose of this study is to identify topic areas of academic library research using two informetric methods: word clustering and Pathfinder network. For the data analysis, 139 articles published in major library and information science journals from 2005 to 2009 were collected from the Korean Science Citation Index database. The keywords that represent research topics were gathered from two sections: an abstract and titles in references. Results showed that reference titles usefully represent topics in detail, and combining abstracts and reference titles can produce an expanded topic map.

초 록

이 연구는 계량정보학적 기법을 적용하여 대학도서관의 연구분야를 파악하고자 하는 것이다. 적용된 계량정보학 기법은 용어클러스터링과 패스파인더 네트워크 알고리즘이다. 연구분야 분석을 위하여 2005년부터 2009년 동안 주요 학술지에 발표된 139건의 논문을 한국과학기술인용색인서비스(KSCI)에서 검색하여 데이터 컬렉션을 구축하였으며 연구분야를 분석하는 데 사용한 자질은 초록과 인용문헌의 제목에서 추출되었다. 분석 결과 인용문헌의 제목은 세분화된 연구분야를 표현하는 데 적절한 것으로 분석되었으며 초록과 인용문헌의 제목을 결합하여 적용하면 연관주제로 확장하여 주제구조를 효과적으로 표현해주는 것으로 나타났다.

Keywords: academic libraries, informetrics, word clustering, Pathfinder network, topic analysis
대학도서관, 계량정보학, 용어클러스터링, 패스파인더네트워크, 주제분석

* This work was supported by the research fund of the Catholic University of Daegu.

** Assistant Professor, Catholic University of Daegu. Department of Library Science(shchoi@cu.ac.kr)

▪ Received : 17 May 2011 ▪ Revised : 24 May 2011 ▪ Accepted : 8 June 2011

▪ Journal of the Korean Society for Information Management, 28(2): 65-78, 2011.

[DOI:10.3743/KOSIM.2011.28.2.065]

1. Introduction

Mapping a domain has attracted many researchers since various informetric methods were developed to successfully illustrate intellectual structures. It fascinated not only information scientist but also other scientists from various fields showing major subject areas, authors, and journals. Especially, to identify subject structure of a domain, it is crucial to use appropriate representatives to the subjects. Descriptors are the most popular features for subject analysis. However, sometimes they failed to effectively illustrate subject areas with the small number of specific terms. They are not effective to identify subject structure at the broad level, either. Therefore, to illustrate subject areas in a wide range, it is necessary to explore the performance of various factors such as abstract words or title words in references.

This study aims to explore the research areas of academic libraries by informetrically with multiple features such as abstracts and titles. To collect multiple features for the generation of intellectual structures, it used the research papers in three major library and information science journals; *Journal of the Korea Library and Information Science Society (JKLISS)*, *Journal of the Korean Society for Library and Information Science (JKSLIS)*, and *Journal of the Korean Society for Information Management (JKSIM)*. Keywords representing the topics of articles in these journals were extracted from two sections: abstracts and titles in references.

Academic libraries have been a core subject in library and information science research studies.

Leading from the front, these libraries embracing new concepts and technologies in this field. Academic library research has played an important role in introducing the latest practices and also in suggesting significant theoretical concepts. Thus, identifying topics in academic library research is an important way of predicting where the research areas of this field are going. Several studies have identified academic library research domains informetrically including digital libraries and archival studies, but few attempts to examine academic library research in its own right have been made.

Abstracts have been considered an appropriate source of keywords that accurately represent the subject of an article. However, because an abstract usually consists of 5-10 sentences, authors must make the most precise word choices to represent their articles' core concepts. The words used in abstracts are therefore considered to be competent subject descriptors.

Similarly, titles of articles that are listed in article references are usually assumed to be closely related to the subjects of these articles.

Authors usually cite various information sources to verify or affirm their arguments. As a result, information sources listed in references can also supply accurate subject representations for articles – at least to some extent. Compared to words in abstracts, title words in references are less specific; however, they are more diverse so as to closely represent the subject of each article. Title words in references can supplement the generalities of abstract words to represent more aspects of the article's subject.

To effectively present research topics within academic libraries, the performance of each element (abstract words and reference title words) were examined. The two elements were also combined in the performance analysis as the third variation.

The analyses of extracted keywords were performed by two informetric methods: word clustering and Pathfinder network algorithm. Abstracts and references used in this study were gathered from the Korean Science Citation Index (KSCI) database.

2. Previous studies

Research trends have been an attractive subject for many scholars who study such patterns within scientific communities. Domain analysis, subject mapping, and topic mapping are typical keywords for this research area. Most research trend analyses are derived with informetric methods.

According to a review of informetric studies about research trend analysis in the 2008 *Annual Review of Information Science Technology*, the analytical approach to research specialties consists of four categories: sociological, bibliographical, communicative, and cognitive (Morris & Van Der Veer Martens 2008).

Among these approaches, the bibliographical approach (which is generally called bibliometrics or informetrics) is the most popular. Bibliometric methods have been chosen to analyze the research trends of many disciplines. For example, citation

analysis is the most common technique used to identify the most up-to-date research topics (Kessler 1963; Small 1973; White & McCain 1998; Jarnerving 2001; Åström 2007). The preferred objects for informetric analysis are academic journal articles, which have structuralized locations for subject-related terms. Because of this advantage, informetric analyses are often performed in the disciplines around which academic journal publication has flourished, such as the scientific and technological research fields.

Informetric techniques have been particularly widely adapted for analysis of the intellectual structures within life science and medical science areas (Seglen & Aksnes 2000; Ananiadou & McNaught 2006). The patent area (Fattori, Peddrizzi, & Turra 2003; Tesng, Lin, & Lin 2007) and The areas of Earth science (Kostoff, Eberhartm, & and Toothman 1998; Kostoff, et al. 2001) are also popular areas in which to analyze intellectual structures. In the social sciences, the major areas for domain analysis include informatics and file maintenance (Persson 1994; White & McCain 1998; Kim & Lee 2007; Kim & Lee 2009) and economy and management (McCain 1991; Sullivan 2001; Miller 2004).

Several researchers have also focused on library and information science and archival studies for informetric analysis. To identify research trends in Korea's library and information science field, Seo (1997) located the major topics by examining articles in the *Journal of the Korean Society for Information Management (JKSIM)* and the *Journal of the American Society of Information Science (JASIS)*. She

compared the research trends of these two journals by analyzing citations of their articles.

Oh and Lee (2005) conducted a similar study in which they analyzed citations in 20 years of articles published in *JKSIM*. Other major informetric studies on sub-subject in library and information science include Kim's (2005) comparison of subject structures in archival research from two countries, Korea and the U.S. She also investigated cocitation patterns of major authors in archival journals. However, Lee (2008) pointed out the limitation of co-author analysis, which is due to the time lag of citing previous publications, and suggested an informetric analysis method instead (using author coupling to bibliometrically present research trends). Lee, Kim, and Kim (2010) analyzed current research trends of digital libraries with the text mining method. These studies have suggested some efficient informetric methods, such as document clustering and cluster-based networks.

3. Data collection and methodology

Data for analysis were obtained from the Korean Science Citation Index database. The articles on academic libraries were searched with one Korean query word ('대학도서관') and two English query words ('academic library' and 'university library') in the titles, keywords, and abstract sections of three journals: the *Journal of Korean Library and Information Science Society* (JKLISS), the *Journal of the Korean*

Society for Library and Information Science (JKSLIS), and the *Journal of the Korean Society for Information Management* (JKSIM). These journals are major publishers of studies in which the trends of academic library research have emerged.

The span of the search was limited to 2005-2009. KSCI project began and its service launched in 2002, so its data coverage started at 2002. However, the bibliographic information from the three journals started to be accumulated from 2005; therefore, the data collected for this study had to have been published between 2005 and 2009.

139 articles were collected by the process explained above: 65 from JKLISS; 43 from JKSLIS, and 31 from JKSIM. Bibliographic factors such as abstracts and references were accumulated in three data collections. The first collection was an abstract word file in both Korean and English. The second collection was a file of title words in references. The third collection combined the previous two. Last, the three files were independently analyzed by two informetric methods: word clustering and Pathfinder network algorithm.

Of the possible clustering methods, the Ward clustering algorithm was chosen to map the subject areas in academic libraries' research. The term weight used for clustering was term frequency and Cosine similarity was calculated among word vectors from each file. The designated number of clusters was determined by the total number of words to clusters; clusters were generated by as many as 10% of the total words. In clustering abstract words, the total number of words was 102, so the designated cluster

number is 10. The total number of reference title clusters is 18, because the total number of reference title words that formed clusters was 182. The combined collection generated 30 clusters because words to cluster comprised 324 terms.

Along with word clustering, a Pathfinder network (PFnet) algorithm was used to profile the research subjects. The PFnet algorithm was suggested by Schvaneveldt (1990) to visualize the intellectual structures of domains. The research subject maps of academic libraries were generated by PFnet at three levels: abstract, reference title, and combined. The top three words (chosen by high frequency) were extracted from the abstract, reference title, and combined files. Using the selected words, a profile similarity matrix was made between each word's vectors. From the similarities between terms, main topics of research papers were extrapolated; next, a PFnet algorithm mapped a similarity matrix in two-dimensional spaces. For this study, the PROXSCAL from SPSS's Multidimensional Scaling program was used to produce the maps.

4. Data analysis

4.1 Data analysis by word clustering

4.1.1 Topics derived by abstract words

Ten clusters were generated by terms found in abstracts. To denote the topic of each cluster, the three terms with the highest frequency were selected. These terms best represented the theme of the cluster they belonged to.

Few clusters related to library resources. Because Cluster 3 contains major topic words such as 'book', 'collection', and 'selection', its contents are defined as 'collection development' research. Cluster 7 is about 'digital resource management', based on analysis of terms it contains.

Users and evaluations of academic libraries were also identified as major topics. Cluster 1 represents research topics related to 'users and satisfaction' and Cluster 2 represents 'information resource and users.' Similar topics are 'user education' (Cluster 9) and 'information services' (Cluster 5).

Terms related to evaluation models and indicators

<Table 1> The results of clustering abstract words

Cluster no.	Top three words that represent the topic of the cluster	Cluster no.	Top three words that represent the topic of the cluster
C1	이용자, 이용, 만족도	C6	운영, Korea, operation
C2	use, public, 공공	C7	자원, 관리, 전자
C3	도서, 선정, collection	C8	평가, model, 지표
C4	inform, user, resource	C9	교육, education, 프로그램
C5	service, 파악, 가능	C10	사서, 주제, librarian

* All English terms are stemmed by Porter stemming algorithm (Porter 1980) throughout the clustering process.

were gathered in Cluster 8. Research topics related to ‘library operation’, ‘national libraries’, and ‘library policy’ were identified in Cluster 6. ‘Subject librarian’ (Cluster 10) was verified as a research topic by abstract terms.

4.1.2 Topics derived by title words in references

By clustering reference title terms, 18 research topics were found. However, few were similar to those defined by abstract words such as ‘evaluation and standards’ (Cluster 9) and ‘user and use’ (Cluster 4). Moreover, analysis of reference title words divided most of the topics identified by abstract words into narrower categories. For example, ‘digital resource management’ (Cluster 7 based on abstract words) became ‘future resource management’ (Cluster 5), ‘access to science resource’ (Cluster 14), and ‘model of web catalog’ (Cluster 16).

Topics related to ‘library operation, national libraries, and library policy’ (Cluster 6 using abstract

words) also became more detailed. Clustering reference title words yielded 5 related topics: ‘national public services’ (Cluster 2), ‘library environmental changes’ (Cluster 10), ‘national strategy for digital library’ (Cluster 11), ‘nation, state, and university’ (Cluster 12), which embraces library operation topics by library type; and ‘online strategy and planning’ (Cluster 17).

Information services yielded two sub-subjects: ‘reference service and digital resource’ (Cluster 15) and ‘information communication and use’ (Cluster 1), both derived from clustering reference title words.

Evaluation clusters discovered by abstract word clustering also yielded two clusters: ‘evaluation and standards’ (Cluster 9) and ‘quality and satisfaction’ (Cluster 18).

‘Satisfaction’ and ‘user’ affiliated with the same cluster via abstract words, but ‘satisfaction’ divided into a new cluster that included the more specific concept of ‘quality.’

Two sub-topics on education also appeared, (e.g.

<Table 2> The results of clustering reference title words

Cluster no.	Top three words that represent the topic of the cluster	Cluster no.	Top three words that represent the topic of the cluster
C1	communication, us, inform	C10	experience, change, model
C2	nation, public, service	C11	국립, 전략, 디지털
C3	공공도서관, 공공, 한국	C12	국가, state, 대학교
C4	user, 관리, 이용	C13	trend, 정책, education
C5	resource, future, management	C14	access, 학술, science
C6	문헌, 모형, 교육	C15	참고, reference, 전자
C7	faculties, college, librarian	C16	목록, web, 모델
C8	환경, 활성화, 운영	C17	online, strategy, plan
C9	evaluation, 기준, 평가	C18	품질, 만족, 체계

‘evaluation case’). These topics associated with the topic ‘education program’ via abstract words analysis. ‘Education program’ became more specific, as ‘education trend and policy’ and ‘document and education model.’

Wide-range analysis of reference titles almost doubled the number of topics via an increase of nearly 80% in the number of words to cluster. However, the latter increase seemed neither to expand the topic areas identified by abstract words in a wide range nor to introduce totally new concepts. Although some clusters such as ‘experience and models for change’ (Cluster 10) and ‘public library’ (Cluster 3) may be new topic areas identified by analysis of reference title words, most of the topics yielded by analysis of abstract words became more segmented rather than more widely expanded.

4.1.3 Topics derived by combined words in abstracts and titles in references

In this study, abstract words and reference title words were combined in order to investigate whether combining two elements can map topic areas better than adapting each element independently.

Clustering abstract and reference title words identified several new topics identified. From Cluster 4, ‘Seoul, Korea, and document’ had never been discovered by previous analyses. ‘Collaboration and region’ is also a new topic cluster, as is Cluster 16 (‘knowledge trends and statistics’). ‘Yearbook and organization’ (Cluster 20), ‘homepage, websites’ (Cluster 23), and ‘market and value’ (Cluster 29) made their first appearances in the third group analysis (clustering abstract and reference title words). Combining the first two elements yielded new re

<Table 3> The results of clustering abstract and reference title words

Cluster no.	Top three words that represent the topic of the cluster	Cluster no.	Top three words that represent the topic of the cluster
C1	technology, use, information	C16	trend, statistics, 지식
C2	모델, public, service	C17	cooperation, scholar, collection
C3	user, 이용자, 이용	C18	센터, 확보, 지원
C4	서울, 문헌, 한국	C19	Operation, issue, 국립
C5	전문, 사서, 교육	C20	연감, unit, organization
C6	standard, 기준, 평가	C21	direct, book, electronic
C7	모형, 환경, 운영	C22	경험, 효과, process
C8	profession, college, librarian	C23	홈페이지, 사이트, web
C9	resource, system, 관리	C24	선정, 장서, 도서
C10	commun, model, educ	C25	봉사, 참고, reference
C11	change, future, management	C26	측정, 만족, 만족도
C12	협력, 지역, 활성화	C27	efficiency, product, data
C13	plan, 전략, 대학교	C28	검색, catalog, 목록
C14	문화, 공공도서관, 공공	C29	market, value, 가치
C15	science, access, 학술	C30	perform, differ, 지표

search topics and also segmented some of the research topic areas identified in the previous analyses.

Clustering reference title words yielded narrower topic areas related to the topics in both the abstract words and reference title words analyses. ‘Special education and librarian’ (Cluster 5) and ‘education and communication model’ (Cluster 10) were revealed as sub-topics of ‘education program.’ ‘Efficiency of process’ (Cluster 22) and ‘measure of satisfaction’ (Cluster 26) were shown to be segmented from ‘evaluation.’

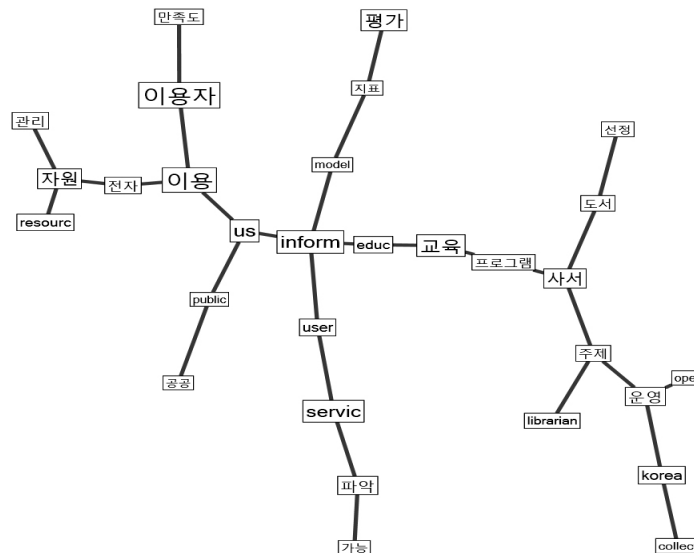
The analysis of clustering reference title words showed several topic areas to be very similar. ‘User and use’ (Cluster 3), ‘evaluation and standards’ (Cluster 6), ‘access to science resources’ (Cluster 15), and ‘reference service’ (Cluster 25) were almost identical to some of the topic areas derived from reference title words.

4.2 Data analysis by Pathfinder network

To identify the academic library research topics, Pathfinder network maps of research topics were generated by PFnet algorithm. These maps were also produced according to three features; abstract words, reference title words, and a combination of abstract and reference title words. One of the purposes in this study is to explore the nature of abstract keywords and title keywords of references, so that all terms used for the analysis were controlled at the minimum level. The lack of controls over the keywords will be discussed in the conclusion.

4.2.1 By abstract words

The analysis result of abstract words shows ‘이용자’ (user), ‘이용’ (use), ‘정보’ (information), and ‘교육’ (education) are important areas in academic library research. According to Figure 1, relationships



〈Figure 1〉 Research topic map by abstract words using PFnet

and connections.

Unlike Figure 2, Figure 3 shows big nodes that are similar to those in Figure 1, such as ‘이용자’ (user), ‘이용’ (use), ‘정보’ (information), ‘교육’ (education), and ‘서비스’ (services).

It also presents similar relationship between major concepts. ‘이용자’ (user) and ‘이용’ (use) show almost the same connection with ‘정보’ (information) that appears in the map of abstract words. Similarly, ‘사서’ (librarian) and ‘교육’ (education) repeat their close relationship in Figure 3, whereas they are close in Figure 1 and apart in Figure 2. The keyword ‘평가’ (evaluation) is also connected to ‘이용자’ (users), as in Figure 1. However, in Figure 3, ‘이용자’ (users) links ‘평가’ (evaluation) to ‘정보’ (information). By contrast, ‘정보’ (information) is the link that connects ‘평가’ (evaluation) and ‘이용자’ (users) in the map based upon abstract words.

Despite these differences and the increased numbers of concepts, the research topic map by combined words has similar structure and connection to the map by abstract words. Therefore, the map shown in Figure 3 can be interpreted as an expanded research topic map of abstract words.

5. Conclusion

This paper has shown how words in abstracts and reference titles can be analyzed by informetric methods to present the nature of academic library research topics. Overall, words in abstracts and reference titles prove to be able to extrapolate research topics at

different levels.

When word clustering is applied, abstract words show the broad concept of research topics while reference title words present more segmented topic areas. Even though the difference in the number of words might be a cause of this result, reference words are useful not only to segment topics identified by abstract words but also to discover new topic areas.

Three topic maps were generated by applying a PFnet algorithm to words found within abstracts and reference titles. These maps demonstrate some significant findings. First, both abstract words and reference title words are useful in presenting the relationships between research topics. In the map of abstract words, major concepts within academic library research are clearly visualized. In the map of reference title words, the major concepts are less clear; however, it does show more diverse aspects of academic library research. Second, the correlation between the map of combined words and the map of abstract words noteworthy because the map of combined words inherits the relationships identified by the map of abstract words, but not those identified by the map of reference title words. Even though the number of reference title words is almost twice the number of abstract words, the map based upon combined words can be interpreted as the expanded version of the map based upon abstract words.

In the topic maps, several key subjects of academic researches were identified. Users, information resources, service, librarian and evaluation were major subject terms for academic library researches. Even though the map of reference title words didn't show

the connection among users, evaluation and information, two maps, abstract keywords map and combined keywords map, illustrated the close relationships of them.

This work, however, clearly has some limitations. For example, there was no authority control for synonyms and cross-linguistic terms. Abstracts of the analyzed articles are provided in both Korean and English, and Korean and English papers appear in

their references. Because the lack of authority control for this problem might hinder better results, future work should include authority control. Furthermore, this study adapts only abstracts and reference titles, in order to clearly compare the characteristics of each element. It is therefore recommended that future research explore more diverse elements for the mapping of research topics.

References

- Ananiadou, Sophia and McNaught, John. eds. 2006. *Text Mining for Biology and Biomedicine*. Norwood: Artech House Publishers.
- Åström, Fredrik. 2007. "Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990-2004." *Journal of the American Society for Information Science and Technology*, 58(7): 947-957.
- Fattori, M., G. Padrazzi, and R. Turra. 2003. "Text Mining applied to patent mapping: a practical business cases." *World Patent Information*, 25(4): 335-342.
- Janervig, B. 2001. "A comparison of two bibliometric methods for the mapping of the research front." *Scientometrics*, 65(2): 245-263.
- Kessler, M. M. 1963. "Bibliographic coupling between scientific papers." *American Documentation*, 14(1): 10-25.
- Kim, Hee-Jung. 2005. "A Study on comparison of intellectual structure in records management and archives using author cocitation." *Journal of the Korean Society for Library and Information Science*, 39(3): 207-224.
- Kim, Heejung and Jae Yun Lee. 2009. "Archiving research trends in LIS domain using profiling analysis." *Scientometrics*, 80(1): 75-90.
- Kim, Panjun and Jae Yun Lee. 2007. "Descriptor profiling for research domain analysis." *Journal of the Korean Society for Information Management*, 24(4): 285-303.
- Korean Citation Science Index.
<<http://ksci.kisti.re.kr>>.
- Kostoff, Ronald N., Antonio J. del Rio, James A. Humenik, and Anna Maria Ramirez. 2001. "Citation mining: integrating text mining and bibliometrics for research user profiling." *Journal of the American Society for Information Science and Technology*, 52(13): 1148-1156.

- Kostoff, Ronald N., H. J. Eberhart, and D. R. Toothman. 1998. "Database tomography for technical intelligence: A roadmap of the near-earth space science and technology literature." *Information Processing & Management*, 34(1): 69-85.
- Kostoff, Ronald N., and Joel A. Block. 2005. "Factor matrix text filtering and clustering." *Journal of the American Society for Information Science and Technology*, 56(9): 946-968.
- Lee, Jae Yun. 2008. "Bibliographic author coupling analysis: A new methodological approach for identifying research trends." *Journal of the Korean Society for Information Management*, 25(1): 173-190.
- Lee, Jae Yun, Heejung Kim and Pan Jun Kim. 2010. "Domain analysis with text mining: Analysis of digital library research trends using profiling methods." *Journal of Information Science*, 36(2): 144-161.
- McCain, Katherin W. 1995. "R&D themes in information science: A preliminary co-descriptor analysis." *Proceedings of the 5th Biennial Conference of the International Society for Scientometrics and Informetrics*. Pine Forest, IL. June 7-10. 275-282.
- McCain, Katherin W. 1991. "Mapping economics through the journal literature: An experiment in journal cocitation analysis." *Journal of the American Society for Information Science*, 42(4): 290-296.
- Miller, T.W. 2004. *Data and Text Mining: A Business Applications Approach*. Upper Saddle River, NJ: Prentice Hall.
- Morris, Steven A. and Van der Veer Martens, Betsy. 2008. "Mapping research specialties." *Annual Review of Information Science and Technology*. Edited by Blaise Cronin. Medford NJ: Information Today.
- Oh, Se-Hoon and Too-Young Lee. 2005. "Research trends of information science in Korea." *Journal of the Korean Society for Information Management*, 22(3): 167-289.
- Persson, O. 1994. "The intellectual base and research fronts of *JASIS* 1986-1990." *Journal of the American Society for Information Science and Technology*, 45(1): 31-38.
- Seglen, P. O., and D. W. Aksnes. 2000. "Scientific productivity and group size: A bibliometric analysis of Norwegian microbiological research." *Scientometrics*, 49(1): 125-143.
- Seo, Eun-Gyung. 1997. "An analytical study on research patterns in information science." *Journal of the Korean Society for Information Management*, 14(1): 269-291.
- Small, Henry. 1973. "Co-citation in the scientific literature: A new measure of the relationship between two documents." *Journal of the American Society for Information Science*, 24(4): 265-269.
- Sullivan, D. 2001. *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*. Chichester: John Wiley & Sons.
- Tseng, Yuen-Hsein, Chi-Jen Lin, and Yu-I Lin. 2007. "Text mining techniques for patent analysis." *Information Processing & Management*, 43(5):

1216-1247.

Todorov, R. 1992. "Displaying content of scientific journals: A co-heading analysis." *Scientometrics*, 23(2): 319-334.

White, Howard D. and Katherine W. McCain. 1998.

"Visualizing a discipline: An author co-citation analysis of information science, 1972-1995." *Journal of the American Society for Information Science*, 49(4): 327-355.