Citation Flow of the *ASIST Proceedings* Using Pathfinder Network Analysis

패스파인더 네트워크 분석에 의한 *ASIST Proceedings* 인용흐름 연구

Heejung Kim*

ABSTRACT

In this study, pathfinder network analysis has been carried out to identify subject domains of documents which cited articles in the *ASIST Proceedings*. This represents how articles in the *ASIST Proceedings*are flowed and used in what subjects areas. For this analysis, 240 documents were selected through a search of the Scopus database. The complete linkage clustering method was used to draw out 16 clusters from 240 documents. Through MDS and pathfinder network analysis, knowledge networks of clusters have been produced. As a result, articles in the *ASIST Proceedings* relating to knowledge management, bibliometrics, information retrieval and digital libraries have been cited actively by other publications. The most frequent citation flow type of *ASIST proceedings* was citation from proceedings(ASIST) to reviews(ARIST), via journals, and the most popular subject areas related to documents were bibliometrics.

초 록

본 연구에서는 ASIST 프로시딩을 인용한 저널들을 중심으로 패스파인더 네트워크 분석을 수행함으로써, ASIST 프로시딩의 지식이 어떠한 주제영역을 중심으로 네트워크 구조를 형성하고 있는지를 확인하는 데에 그 목적이 있다. 이를 위하여 Scopus 데이터베이스에서 검색한 240개의 문헌을 대상으로 완전연결 클러스터링 기법을 통하여 16개 클러스터를 도출하였으며, MDS 및 패스파인더 네트워크 분석을 통하여 지식 네트워크를 매핑하였다. 지금까지 대부분의 경우 학술지를 대상으로 수행되어 온 네트워크 분석을 프로시딩을 대상으로 분석을 시도하였으며, 분석 결과 ASIST 프로시딩은 정보추구행태 및 탐색과 인터페이스, 계량서지학 및 지식관리 주제영역의 논문이 타 문헌에 활발하게 소비되고 있음을 확인할 수 있었다.

Keywords: citation analysis, pathfinder network analysis, bibliometrics
인용분석, 패스파인더 네트워크 분석, 계량서지학

* Director, International Vaccine Institute(heejung@ivi.int)

▪ Received : 21 May 2008 ▪ Revised : 29 May 2008 ▪ Accepted : 9 June 2008
▪ Journal of the Korean Society for Information Management, 25(2): 157-166, 2008.
〔DOI:10.3743/KOSIM. 2008.25.2.157〕

1. Introduction

1.1 Background and Purpose of Research

As with the dynamic flow of information, which transforms people's thinking and life, academic knowledge presented in publications also actively flows and evolves. For example, in the beginning, ideas and opinions formed in the proceedings. As time passed, some articles published in proceedings developed and were published in journals.

Until now, many researches attempting to identify the intellectual structure of some subject field mainly focused on journals. Different from prior researches, proceedings were selected to start the network analysis in this study, which would show prior characteristics to journals or reviews.

The purpose of this research is to identify the intellectual structure of *ASIST proceedings,* and citation flow among proceedings(*ASIST*), journals and reviews(*ARIST*), which will show the dynamic flows and evolution of information and knowledge.

1.2 Related Research

Lee(2006) suggested Parallel Nearest Neighbor Clustering and details of the methodology are described in the next chapter.

Lee, Moon, and Kim(2007) investigated intellectual structure of records management and archival Science in Korea using document clustering, a widely used method of text mining, and document similarity network analysis. 145 articles written on the subject of the field were selected from five major representative journals published from 2001 to 2006.

H. Kim and J.Y. Lee(2008a) analyzed the intellectual structure of records management and archives using pathfinder networks based on document cluster similarities. The data used in this study were 432 articles selected from six major journals in the field of records management and archives published from 2001 to 2004.

H. Kim and J.Y. Lee(2008b) provided archiving research trends from the perspective of the field of library and information science using profiling analysis method. The LISA database has been selected as the representative database in the library and information science field and articles have been searched via the keyword 'archiv*'.

2. Research Design

2.1 Data Collection

In this study, *ASIST proceedings* was selected as a target proceeding, because of the popularity and close relationship among journals(including *JASIST*) and reviews(*ARIST*) in library and information field.

To determine the characteristics of the subject networks of articles citing the *ASIST Proceedings,* 240 documents were selected from the Scopus database(through the function 'REFSRCTITLE').

2.2 Methodology

The processes of the analysis were as follows: First, Cosine similarity among the 240 documents were produced, and LTF*IDF weights were applied. For the next step, the second Pearson correlation coefficient was produced. For the creation of clusters, the complete linkage clustering method was applied. Considering size balances(similar size) among clusters, 16 clusters were produced.

For the next step, the similarities among these 16 clusters were identified to produce multidimensional scaling. After that, MDS(Multidimensional Scaling) and the PFNET method were applied. The PNNC algorithm was also applied, and 4 clusters produced.

The PNNC method, which Lee(2006) suggested, refers to Parallel Nearest Neighbor Clustering. Descriptors and journals in PFNet were categorized systematically using the Parallel Nearest Neighbor Clustering(PNNC) algorithm(Lee 2006). The PNNC algorithm is a hierarchical clustering method that connects a node with its nearest neighbor.

> It is similar to single link agglomerative clustering algorithm but differs in merging sequence. The PNNC algorithm works by:
>
> ① Assuming each node as a cluster with size one.
>
> ② Finding the nearest neighbor cluster(s) of each cluster.
>
> ③ Connecting every cluster with its nearest neighbor(s). (If there are more than one nearest neighbor exists, connecting them all)
>
> ④ If more than one cluster remains, step ② and ③ are repeated(H. Kim and J. Y. Lee 2008a).

3. Analysis and Results

240 articles cited the *ASIST Proceedings*. The articles from ASIST Proceedings which were cited six or more times are listed in <Table 1> below:

Among 240 articles, sources of articles(publication titles), which cited ASIST proceedings more than 3 times, are listed in <Table 2> below:

16 clusters were selected to analyze subject networks and the relations between the 240 articles. Considering similarities among clusters and within clusters, clusters and the top four documents in each group can be suggested as follows:

For the next step, 2 dimension mapping was carried out through the application of MDS with the similarity among the 16 clusters. Especially, four clusters were produced through the PNNC method. The results are presented in <Figure 1> below:

As shown in <Figure 1>, there are big four clusters are located.

The cluster on the far left, PNNC 2 is focused on bibliometrics. It is composed of citation analysis, web link analysis, and biblioinformatics.

〈Table 1〉 Articles of ASIST Proceedings Cited More than 6 times

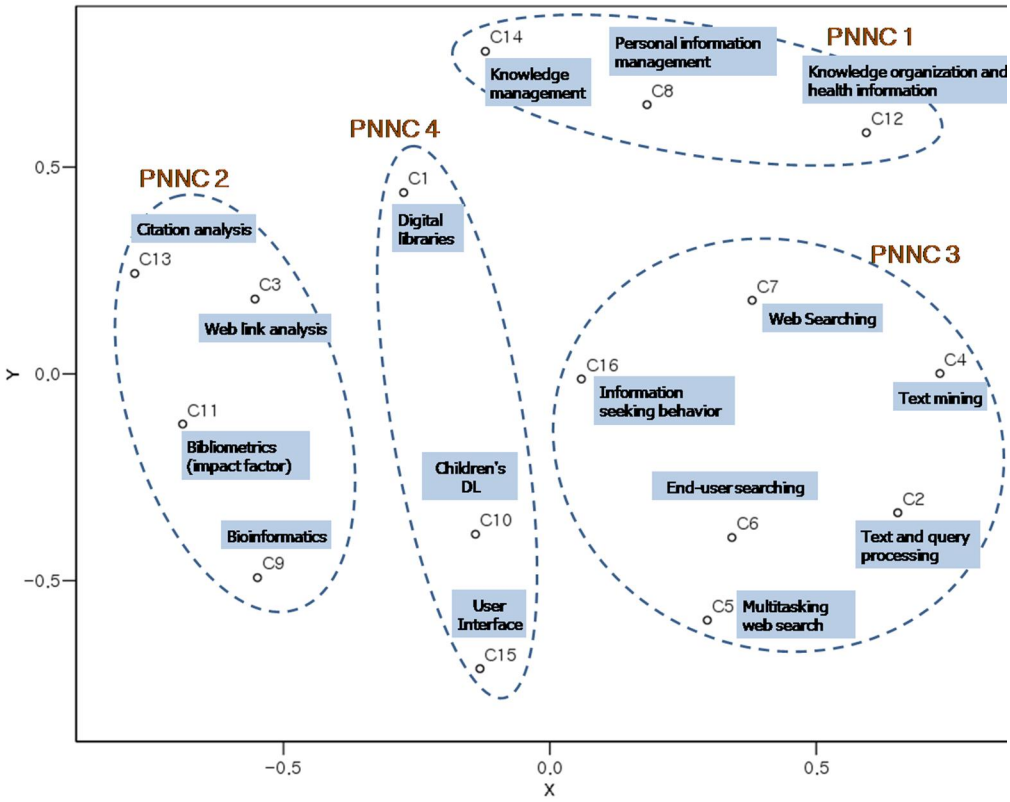
Cited documents	Freq.
Jones, W.P., Dumais, S.T., Bruce, H., Once found, what next? A study of 'keeping' behaviors in the personal use of web information(2002) Proceedings of ASIST 2002.	16
Larson, R.R., Bibliometrics of the world wide web: An exploratory analysis of the intellectual structure of cyberspace(1996) Proceedings of ASIS 1996.	12
Fry, J., Talja, S., The cultural shaping of scholarly communication: Explaining e-journal use within and across academic fields(2004) Proceedings of ASIST 2004.	10
Miwa, M., User situations and multiple levels of users goals in information problem solving processes of Ask ERIC users(2001) Proceedings of ASIST 2001.	10
Ozmutlu, S., Ozmutlu, H.C., Spink, A., Multitasking Web search: Implications for design(2003) Proceedings of ASIST 2003.	10
Rieh, S.Y., Xie, H., Patterns and sequences of multiple query reformulations in web searching: A preliminary study(2001) Proceedings of ASIST 2001.	8
Bilal, D., Draw and tell: Children as designers of Web interfaces(2003) Proceedings of ASIST 2003.	8
Tang, R., Thelwall, M., Exploring the pattern of links between Chinese university Web sites(2002) Proceedings of ASIST 2002.	8
Fast, J.V., Campbell, D.G., "i still like Google": University student perceptions of searching OPACs and the Web(2004) Proceedings of ASIST 2004.	7
Large, A., Beheshti, J., Moukad, H., Information seeking on the web: Navigational skills of grade-six primary school students(1999) Proceedings of ASIS 1999.	7
Large, A., Beheshti, J., Nesset, V., Bowler, L., Children as designers of Web portals(2003) Proceedings of ASIST 2003.	6

〈Table 2〉 Sources of Articles(publication titles)(which cited ASIST proceedings more than 3 times)

Titles	Freq.
Journal of the American Society for Information Science and Technology	50
Information Processing and Management	13
Proceedings of the ASIST Annual Meeting	13
Annual Review of Information Science and Technology	12
Library and Information Science Research	9
Journal of Information Science	7
Information Research	7
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	7
Proceedings of the ACM IEEE Joint Conference on Digital Libraries	6
Library Trends	5
Online Information Review	5
Scientometrics	5
Journal of Documentation	4
Conference on Human Factors in Computing Systems Proceedings	4
Proceedings of the ACM International Conference on Digital Libraries	4
Journal of Digital Information	3
Communications of the ACM	3

〈Table 3〉 Clusters and Top Four Documents in Each Group According to the Similarities (parts of the article titles are suggested)

C1	Digital libraries	C9	Bioinformatics
D41	Digital libraries	D156	Developing a protocol for bioinformatics analysis
D226	Designing, implementing, and evaluating	D150	Where do molecular biology graduate students find
D135	Scholarly work and the shaping of digital access	D166	Bioinformatics
D120	The INVENT framework: Examining the role	D104	Diffusion pattern of the use of genomic databases
C2	Text and query processing	C10	Children's DL
D87	Automated judgment of document qualities	D51	The evolution of the international children's digital
D215	Can Document-genre Metadata Improve Information	D15	Supporting elementary-age children's searching
D115	Is Google enough? Comparison of an internet search	D131	Children's conceptual structures of science categories
D125	What type of page is this? Genre as web descriptor	D139	Children's information seeking and the design
C3	Web link analysis	C11	Bibliometrics(impact factor)
D212	Why do web sites from different academic subjects	D169	On the scholarly use of electronic journals
D207	Methods for reporting on the targets of links	D98	The history and meaning of the journal impact factor
D210	Disciplinary and linguistic considerations for academic	D34	Evaluating E-contents beyond impact factor
D144	National and international university department	D147	The "impact factor" revisited
C4	Text mining	C12	Knowledge organization and health information
D195	Using MEDLINE as a knowledge source	D88	Semantic representation of consumer questions
D72	A large scale, corpus-based approach	D107	Exploring and developing consumer health
D12	The impact of number of query words on image	D17	How consumers search for health information
D170	A query analytic model for image retrieval	D44	Semantics and knowledge organization
C5	Multitasking Web search	C13	Citation analysis
D108	Multitasking during Web search sessions	D42	Citation analysis
D176	Information task switching and multitasking	D209	Bibliographic and Web Citations
D193	Multitasking Web search on Alta Vista	D143	Web citation data for impact assessment
D213	Multitasking web searching and implications	D18	How is science cited on the web?
C6	End-user searching	C14	Knowledge management
D196	Elicitation purposes and tasks during mediated	D165	Managing social capital
D240	Study of mediated successive searching	D202	Explaining knowledge sharing in organizations
D185	Using the information see ker to elicit construct	D71	On the nature of knowledge: Rethinking popular
D171	From information retrieval to information interaction	D142	Knowledge processes. A strategic foundation
C7	Web searching	C15	User Interface
D112	Saving and using encountered information	D174	Designing web portals in intergenerational teams
D201	Sharing encountered information: Digital libraries	D100	Bonded design: A novel approach to intergenerational
D186	On the Web at home: Information seeking and Web	D97	Search histories for user support in user interfaces
D214	Investigating web searching behavior in home	D62	Just the facts ma'am? A contextual approach
C8	Personal information management	C16	Information seeking behavior
D43	Personal information management	D159	Internetworking of factors affecting successive search
D152	Personal, anticipated information need	D157	Search moves and tactics for image retrieval
D154	Personal, anticipated information need	D160	Validation of a model of information seeking
D204	Stuff goes into the computer and doesn't come out	D190	The effects of domain knowledge on search tactic



〈Figure 1〉 PROXSCAL, Complete Linkage, 16 clusters and PNNC 4 categories

The cluster in the center, PNNC 4 shows digital libraries subjects.

The cluster on the lower right, PNNC 3 is strongly related to information retrieval.

The cluster on the upper right, PNNC 1 shows a strong relationship to knowledge management.

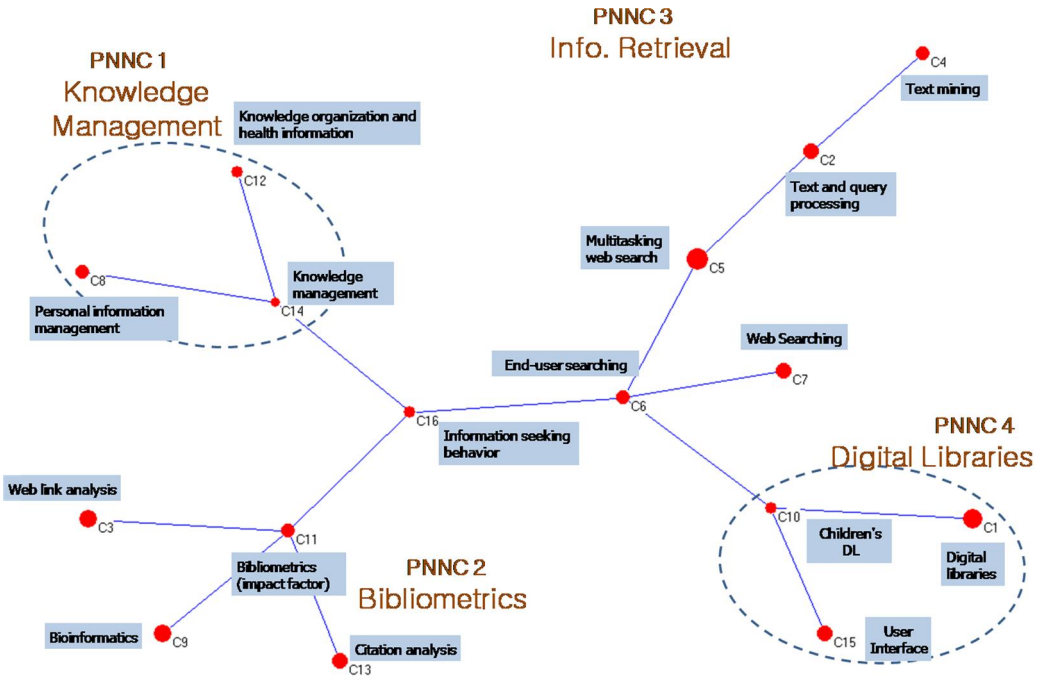
The results depicted in <Figure 1> show the size and clear distinction among 4 subject categories.

However, the relationship among these 4 categories cannot be found in this MDS map.

Therefore, more detailed interrelationship and the flow of subject through pathfinder networks are shown in <Figure 2>.

<Figure 2> shows not only the subject networks and the flow and order of subjects within each of the 4 categories, but also the interrelationship between the four clusters.

Looking at PNNC 3, labeled as information retrieval, C 16 and C 6, which are located in the center of the map, are the main hub nodes, and the pivot nodes, which also play a role in connecting the two different groups. The subject of C 16 is 'information seeking behavior,' and that of C 6 is 'end user searching.' Multitasking web search



〈Figure 2〉 Complete Linkage, 16 clusters, PFNet, PNNC 4 categories

is connected and evolves to text and query processes, and this is connected to text mining again. The subject 'web searching' is no longer extended.

By following the branches, the flow of changing subjects can be recognized.

PNNC 4, labeled digital libraries, shows three detailed subjects flows: 1) children's digital libraries, 2) user interface, and 3) digital libraries.

PNNC 1 and PNNC 2 have clear subject scope, knowledge management, and bibliometrics.

To figure out citation flow of ASIST proceedings, three citation types are suggested and matched to PNNC categories as shown in <Table 4>.

Type A shows the citation flow from proceedings(ASIST) to reviews(ARIST) via proceedings. Subjects of Type A are focused on PNNC 1 and PNNC 3, knowledge management and information retrieval.

Type B shows the citation flow from proceedings(ASIST) to reviews(ARIST) via journals. Subject of Type B is focused on PNNC 2, bibliometrics.

Type C shows the citation flow from proceedings(ASIST) to reviews(ARIST) directly.

Subject of Type C is focused on PNNC 2, bibliometrics, too.

Cross-table can be suggested in different way shown in <Table 5> as follows.

〈Table 4〉 Citation Flow and PNNC Categories

Type	Citation Flow	Titles	Doc. No.	Cluster No.	PNNC1 KM	PNNC2 Biblio-metrics	PNNC3 IR	PNNC4 DL
A	Proceedings (ASIST) → Proceedings → Reviews (ARIST)	Proceedings of the ASIST	D204	C8	V			
		Proceedings of SIGCHI on HF in CS	D177	C8	V			
		Proceedings of SIGCHI on HF in CS	D112	C7			V	
		Proceedings of ACM/IEEECS JCDL	D201	C7			V	
		Proceedings of SIGCHI on HF in CS	D200	C8	V			
B	Proceedings (ASIST) → Journals → Reviews (ARIST)	JASIST	D235	C10				V
			D153	C1				V
			D237	C16			V	
			D186	C7			V	
			D238	C5			V	
			D219	C3		V		
			D222	C3		V		
			D209	C13		V		
			D233	C6			V	
		IPM	D207	C3		V		
		Scientometrics	D210	C3		V		
			D194	C13		V		
		JIS	D227	C3		V		
			D229	C3		V		
		Online Information Review	D226	C1				V
			D208	C11		V		
			D225	C3		V		
		Library & Information Science Research	D173	C2			V	
			D239	C14	V			
		Information Research	D152	C8	V			
			D191	C9		V		
		Communications of the ACM	D91	C12	V			
C	Proceedings (ASIST) → Reviews	ARIST	D205	C5			V	
		ARIST	D166	C9		V		
		ARIST	D110	C9		V		
		ARIST	D165	C14	V			
		ARIST	D111	C9		V		
		ARIST	D161	C10				V
		ARIST	D163	C3		V		

〈Table 5〉 Cross-Table between Publications and PNNC

Type	Title	PNNC1			PNNC2				PNNC3						PNNC4			Total
		C8	C12	C14	C3	C9	C11	C13	C2	C4	C5	C6	C7	C16	C1	C10	C15	
A	proceedings	3											2					5
B	JASIST				2			1			1	1	1	1	1	1		22
	IPM				1													
	Scientometrics				1			1										
	JIS				2													
	OIR				1		1								1			
	LISR			1					1									
	IR	1				1												
	Commu. ACM		1															
C	ARIST			1	1	3					1					1		7
Total		7			15				8						4			34

As shown in <Table 5>, the most frequent citation flow type of ASIST proceeding was type B, which represents citation from proceedings(ASIST) to reviews(ARIST), via journals. The most popular subject areas related to documents were PNNC 2, bibliometrics.

4. Conclusion

As a result of performing clustering and pathfinder network analysis of documents which cited the *ASIST Proceedings,* four PNNC were identified: PNNC1 knowledge management, PNNC2 bibliometrics, PNNC3 information retrieval, and PNNC4 digital libraries.

Three citation types were suggested to show the knowledge flow of ASIST proceedings: Type A, from proceedings(ASIST) to reviews(ARIST) via proceedings, Type B, from proceedings(ASIST) to reviews(ARIST) via journals, and Type C, from proceedings(ASIST) to reviews (ARIST) directly.

Related subjects were different according to types: Type A was close to PNNC 1 and PNNC 3, knowledge management and information retrieval. On the other hand, Type B and Type C both were close to PNNC 2, bibliometrics.

Overall, the most frequent citation flow type of ASIST proceeding was type B, citation from proceedings(ASIST) to reviews(ARIST), via journals, and the most popular subject areas related to documents were PNNC 2, bibliometrics.

In summary, the *ASIST Proceedings* have been cited actively, especially in relation to knowledge management, information retrieval, digitallibraries, and bibliometrics.

Further research can be suggested with the expansion of the scopes of proceedings and data.

References

Kim, H., Lee, J. Y. 2008a. "Exploring the Emerging Intellectual Structure of Archival Studies using Text mining: 2001-2004." *Journals of Information Science,* 34(3): 356-369.

Kim, H., Lee, J. Y. 2008b. "Archiving Research Trends in LIS Domain using Profiling Analysis." *Scientometrics,* xx(x), accepted for publication.

Kim, P., Lee, J. Y. 2007. "Mapping Knowledge Domains with Descriptor Profiling," *Proceedings of the 14th Conference of Korean Society for Information Management,* 14: 131-138.

Lee, J. Y. 2006. "A Novel Clustering Method for Examining and Analyzing the Intellectual Structure of a Scholarly Field." *Journal of the Korean Society for Information Management,* 23(4): 215-231.

Lee, J. Y., Moon, J. and Kim, H.. 2007. "Examining the Intellectual Structure of Records Management & Archival Science in Korea with Text Mining." *Journal of the Korean Society for Library and Information Science,* 41(1): 345-372.

Porter, M. 1980. "An Algorithm for Suffix Stripping." *Program,* 14(3): 130-137.

Salton, G., Mcgill, M. J. 1983. *Introduction to Modern Information Retrieval*, McGrawHill Book Co., New York.

Schvaneveldt, R. W. 1990. *Pathfinder Associative Networks: Studies in Knowledge Organization*, Ablex, Norwood, N. J.

Scopus homepage. <http://www.scopus.com>.