

\*

## Descriptor Profiling for Research Domain Analysis

(Panjun Kim)\*\*,

(Jae Yun Lee)\*\*\*

가  
 가  
 가  
 가

### ABSTRACT

This study aims to explore a new technique making complementary linkage between controlled vocabularies and uncontrolled vocabularies for analyzing a research domain. Co-word analysis can be largely divided into two based on the types of vocabulary used: controlled and uncontrolled. In the case of using controlled vocabulary, data sparseness and indexer effect are inherent drawbacks. On the other case, word selection by the author's perspective and word ambiguity. To complement each other, we suggest a descriptor profiling that represents descriptors(controlled vocabulary) as the co-occurrence with words from the text(uncontrolled vocabulary). Applying the profiling to the domain of information science implies that this method can complement each other by reducing the inherent shortcoming of the controlled and uncontrolled vocabulary.

descriptor, author, profiling, domain analysis, intellectual structure, mapping, co - word analysis, controlled vocabulary, uncontrolled vocabulary

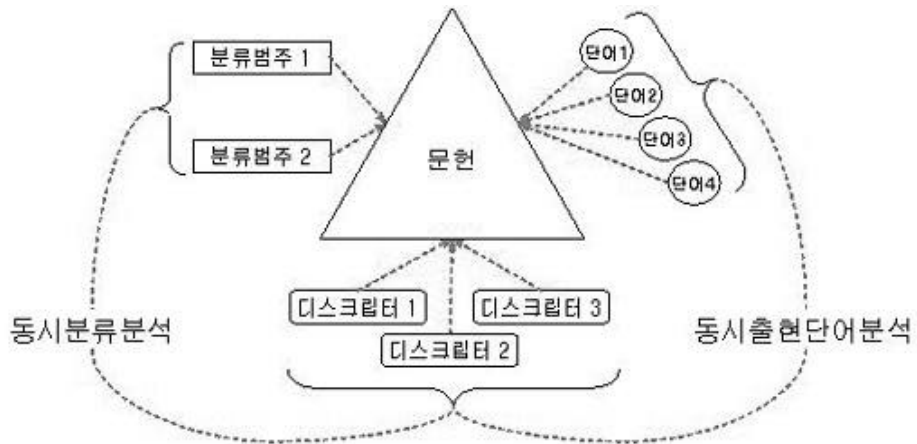
---

\* 14 (2007 8 17 , )  
 \*\* (dpblueseas@hanmail.net), 1  
 \*\*\* (memexlee@kgu.ac.kr),  
 :2007 11 30  
 :2007 12 11

1.

1980 (co - word analysis) (Callan et al. 1983; Callan et al. 1986). (co - classification analysis; Todorov 1989), (co - heading analysis; Todorov & Winterhager 1990), (co - descriptor analysis; McCain 1995), (co - term network analysis; Jacobs 2002) , < 1>

(He, 1999). (database tomography; Kostoff 1993; Kostoff et al. 2005), (tech mining; Porter & Cunningham 2005)



가  
 가 . (Bhattacharya and Basu 1998; Widhalm et al. 2001; Watt, Porter, and Minsk 2004), 가

(Courtial et al. 1984)

(full-text)

가  
 (Voutilainen 1993; Ding et al. 2001; Buter and Noyons 2002; Porter 2005; Kostoff et al. 2005; Janssens et al. 2006).

가  
 가 .

가 .

2.

(full-text)

가  
 가

가 (Callan et al. 1983; Todrov 1990; Tijssen 1992; McCain 1995; Spasser 1997; Noyons and van Raan 1998).

profiling)

(descriptor

(van Raan 1987)

( )

가 , 가

< 2>

가 , 가

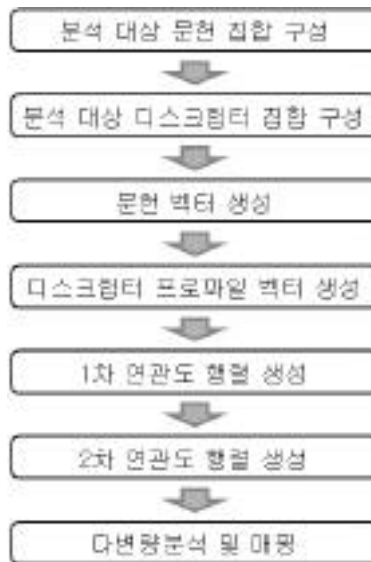
: ( )

가

가 가 < 3>

3.

1 :



< 2>

1

:1

2

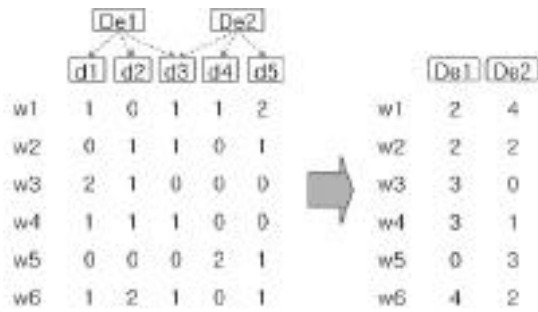
2

:1

2

2

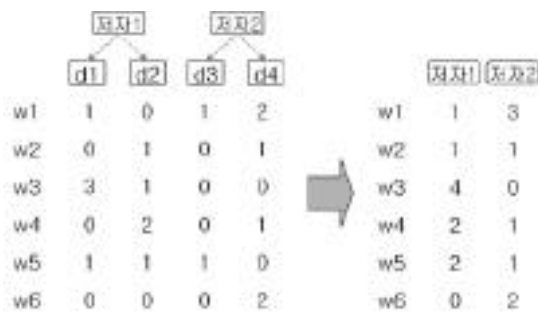
가



De : 디스크릴터, d : 문헌, w : 단어,

↘ : 문헌에 디스크릴터 할당

< 3>



d : 문헌, w : 단어,

↘ : 저자가 발표한 문헌

< 4>

가  
 1995  
 ( 2005) 2006 12  
 가  
 가  
 < 4> 4.1 가  
 가 Journal Citation Reports  
 , 가 2001 2006 6  
 가 5 Journal  
 가 of the American Society for Information  
 Science & Technology, Information  
 Processing & Management, Journal of  
 Documentation 3  
 4.  
 LISA  
 1995

< 1 >

searching	648
online information retrieval	628
world wide web	291
evaluation	148
citation analysis	133
information seeking behaviour	130
computerized information storage and retrieval	111
research	109
automatic text analysis	108
bibliometrics	104

&lt; 2&gt;

5

( )

	citation analysis	information seeking behaviour	computerized information storage and retrieval	automatic text analysis	bibliometrics
2 - dimension	0	0	0	0	5.28
3 - dimension	4.47	0	0	0	0
aamulehti	0	0	0	0	0
aat	0	0	0	0	0
ab	0	0	0	0	5.69
abandon	0	0	0	0	0
abbott	0	0	0	0	0
abbrevi	0	0	13.95	11.46	0
abc	0	0	4.95	0	0
abel	0	0	0	0	0
abi	0	0	0	0	0
abid	0	0	0	0	4.69
abil	0	6.39	5.43	4.77	0
abl	1.77	2.65	2.38	3.54	1.77
aboard	0	0	0	0	0
abort	5.69	0	0	0	0
yule	0	0	4.28	7.24	0
yule - simon	0	0	0	0	8.93
z39	0	0	0	0	0
zadeh	0	0	0	0	0
zagreb	0	0	0	0	0
zatocod	0	0	0	0	0
zeal	0	0	0	0	0
zealand	0	0	0	0	0
zero	4.95	0	0	0	0
zero - order	0	0	0	0	0
zhang	0	0	0	0	0
zipf	9.98	2.63	0	2.63	24.75
zipf - distribut	0	0	0	0	0
zipfian	4.95	0	0	0	4.95
zipf - mandelbrot	0	0	0	0	7.28
zipf - type	5.69	0	0	0	0
zone	0	0	0	0	4.69
zoom	0	0	0	0	0
zprise	0	0	0	0	0
z - score	0	0	0	6.28	0

2006, 3, 2,245, 40%  
 'searching', 'online information retrieval', 'world wide web', 'evaluation', 'research', 22  
 (DE) : 가 (TI) (AB) 가 ( : 6,433 ) 4.2  
 (2006) (AU) IDF 가 6,433 log TF x 2,245  
 2,908, 3  
 12 18 가 Visual Foxpro  
 109, 1.5, 24, 85 < 2>  
 < 3> 1 ( ) ( )

Descriptor	ID	DE1	DE2	DE3	DE4	DE5	DE6	DE7	DE8	DE9	DE10
citation analysis	DE1	1.000	0.173	0.176	0.157	0.462	0.319	0.198	0.220	0.183	0.217
information seeking behaviour	DE2	0.173	1.000	0.257	0.141	0.179	0.310	0.248	0.216	0.307	0.316
computerized information storage and retrieval	DE3	0.176	0.257	1.000	0.363	0.178	0.394	0.280	0.154	0.226	0.174
automatic text analysis	DE4	0.157	0.141	0.363	1.000	0.176	0.230	0.211	0.110	0.171	0.137
bibliometrics	DE5	0.462	0.179	0.178	0.176	1.000	0.300	0.172	0.204	0.188	0.193
performance measures	DE6	0.319	0.310	0.394	0.230	0.300	1.000	0.480	0.151	0.235	0.191
search engines	DE7	0.198	0.248	0.280	0.211	0.172	0.480	1.000	0.110	0.277	0.141
information science	DE8	0.220	0.216	0.154	0.110	0.204	0.151	0.110	1.000	0.176	0.256
internet	DE9	0.183	0.307	0.226	0.171	0.188	0.235	0.277	0.176	1.000	0.246
information communication	DE10	0.217	0.316	0.174	0.137	0.193	0.191	0.141	0.256	0.246	1.000



SPSS 1 ( ) < 3>  
 1 ( )  
 < 3> SPSS  
 < 4> 2 ( 1 ) ( )

Descriptor	ID	DE1	DE2	DE3	DE4	DE5	DE6	DE7	DE8	DE9	DE10
citation analysis	DE1	1.000	0.672	0.670	0.623	0.895	0.786	0.697	0.726	0.712	0.718
information seeking behaviour	DE2	0.672	1.000	0.751	0.627	0.684	0.809	0.776	0.729	0.840	0.818
computerized information storage and retrieval	DE3	0.670	0.751	1.000	0.865	0.688	0.883	0.838	0.654	0.779	0.691
automatic text analysis	DE4	0.623	0.627	0.865	1.000	0.646	0.773	0.745	0.579	0.697	0.614
bibliometrics	DE5	0.895	0.684	0.688	0.646	1.000	0.785	0.696	0.714	0.719	0.709
performance measures	DE6	0.786	0.809	0.883	0.773	0.785	1.000	0.920	0.697	0.816	0.737
search engines	DE7	0.697	0.776	0.838	0.745	0.696	0.920	1.000	0.634	0.810	0.686
information science	DE8	0.726	0.729	0.654	0.579	0.714	0.697	0.634	1.000	0.719	0.769
internet	DE9	0.712	0.840	0.779	0.697	0.719	0.816	0.810	0.719	1.000	0.791
information communication	DE10	0.718	0.818	0.691	0.614	0.709	0.737	0.686	0.769	0.791	1.000

< 5> ( )

Descriptor	ID	DE1	DE2	DE3	DE4	DE5	DE6	DE7	DE8	DE9	DE10
citation analysis	DE1	133	0	0	0	7	7	2	2	1	4
information seeking behaviour	DE2	0	130	4	0	0	6	3	1	4	6
computerized information storage and retrieval	DE3	0	4	111	2	0	14	2	0	0	0
automatic text analysis	DE4	0	0	2	108	2	0	1	0	0	0
bibliometrics	DE5	7	0	0	2	104	5	1	1	2	1
performance measures	DE6	7	6	14	0	5	78	18	0	0	0
search engines	DE7	2	3	2	1	1	18	75	0	5	0
information science	DE8	2	1	0	0	1	0	0	72	0	3
internet	DE9	1	4	0	0	2	0	5	0	72	2
information communication	DE10	4	6	0	0	1	0	0	3	2	66

< 4> 2 analysis (DE1)

< 5>

< 6>

< 7>

< 6> 'user behaviour' 가

< 5> 'citation analysis (DE1)가  
' bibliometrics (DE5) ' performance measures' (DE6) 7  
, 2 'information seeking behaviour' 2 , 12

< 4> ' performance measures' (DE6) ' bibliometrics (DE5)가' citation < 7> ' information science'

< 6> ' user behaviour '

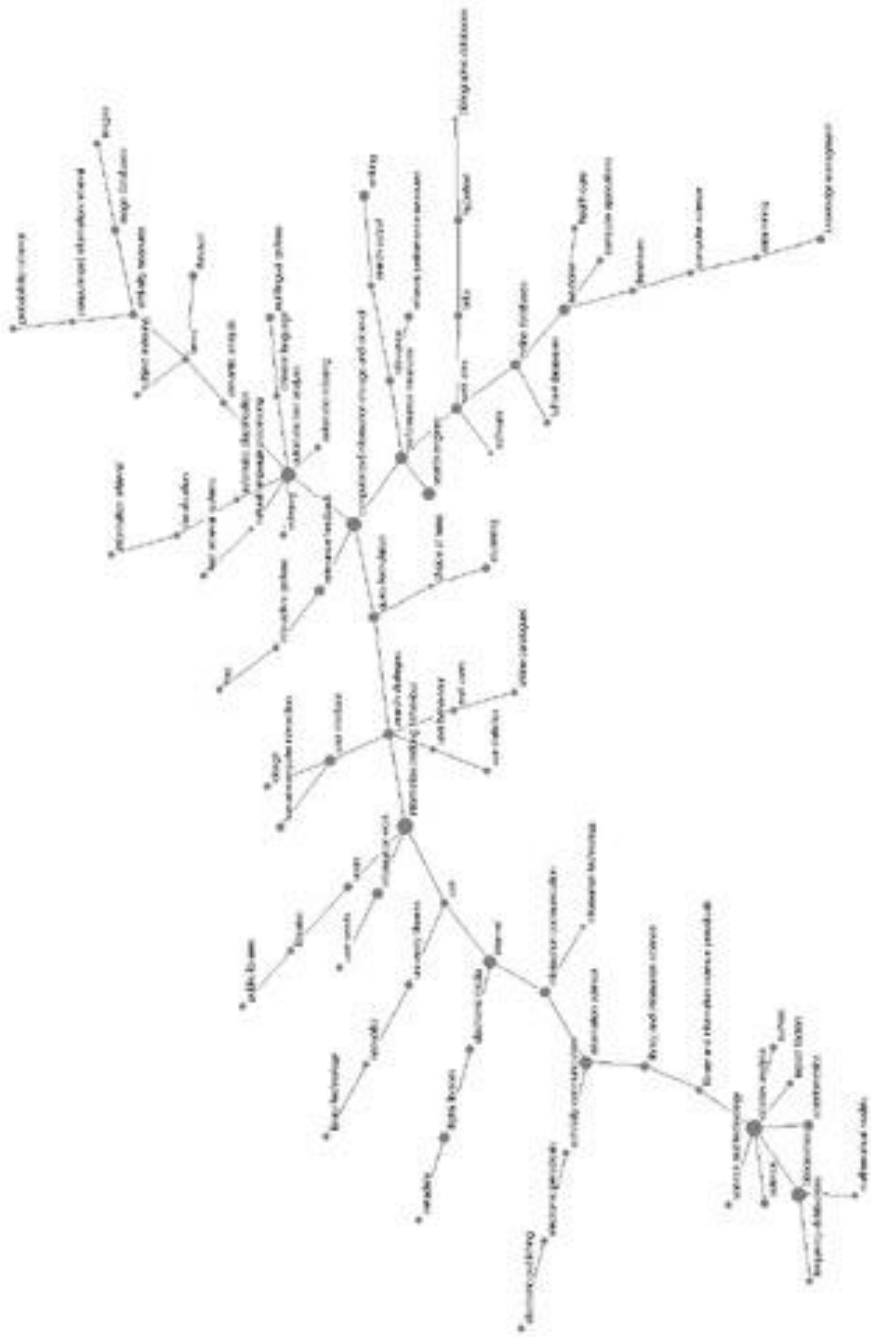
1	use statistics	use statistics
2	information seeking behaviour	end users
3	end users	performance measures

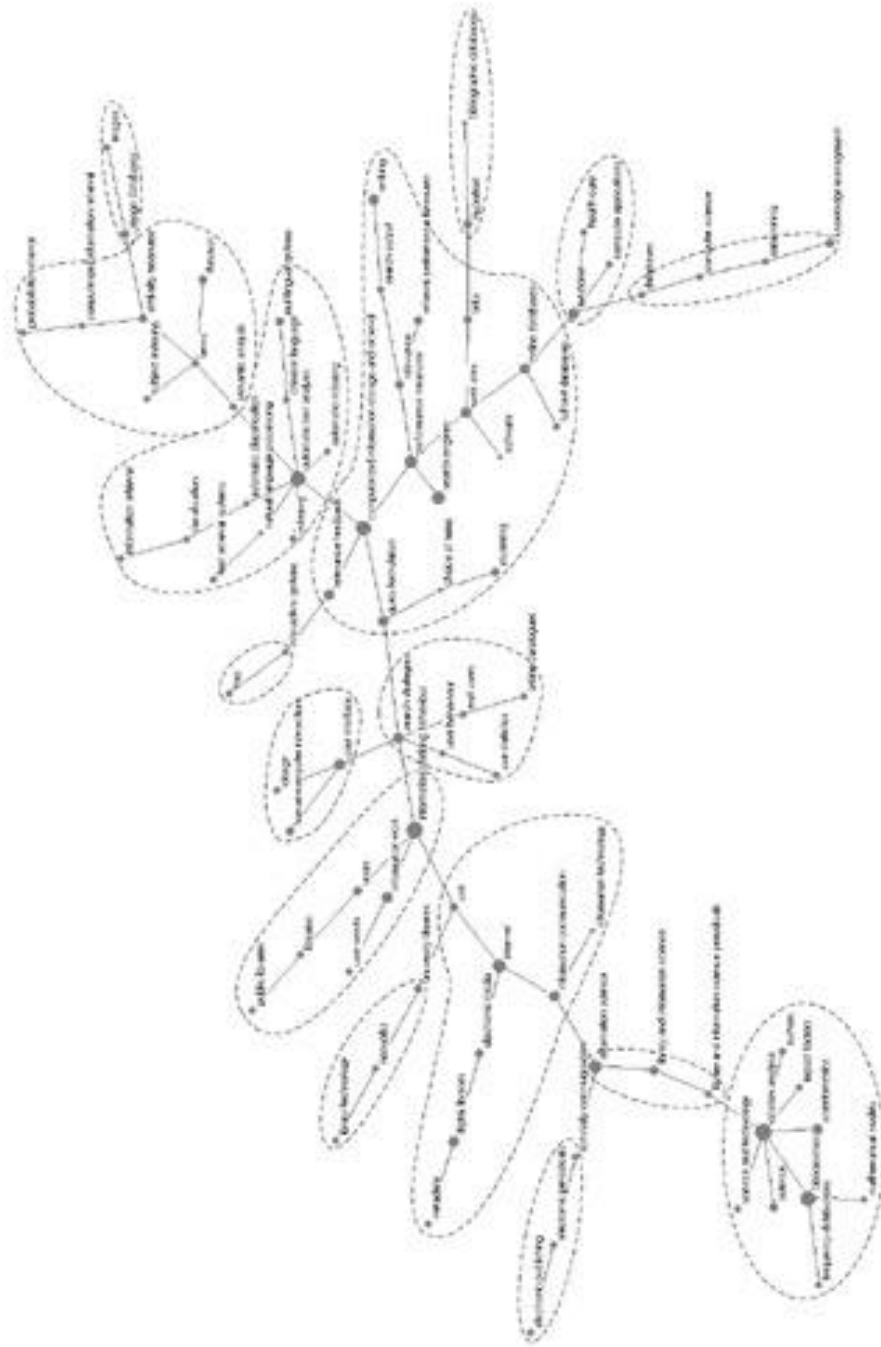
< 7> ' information science '

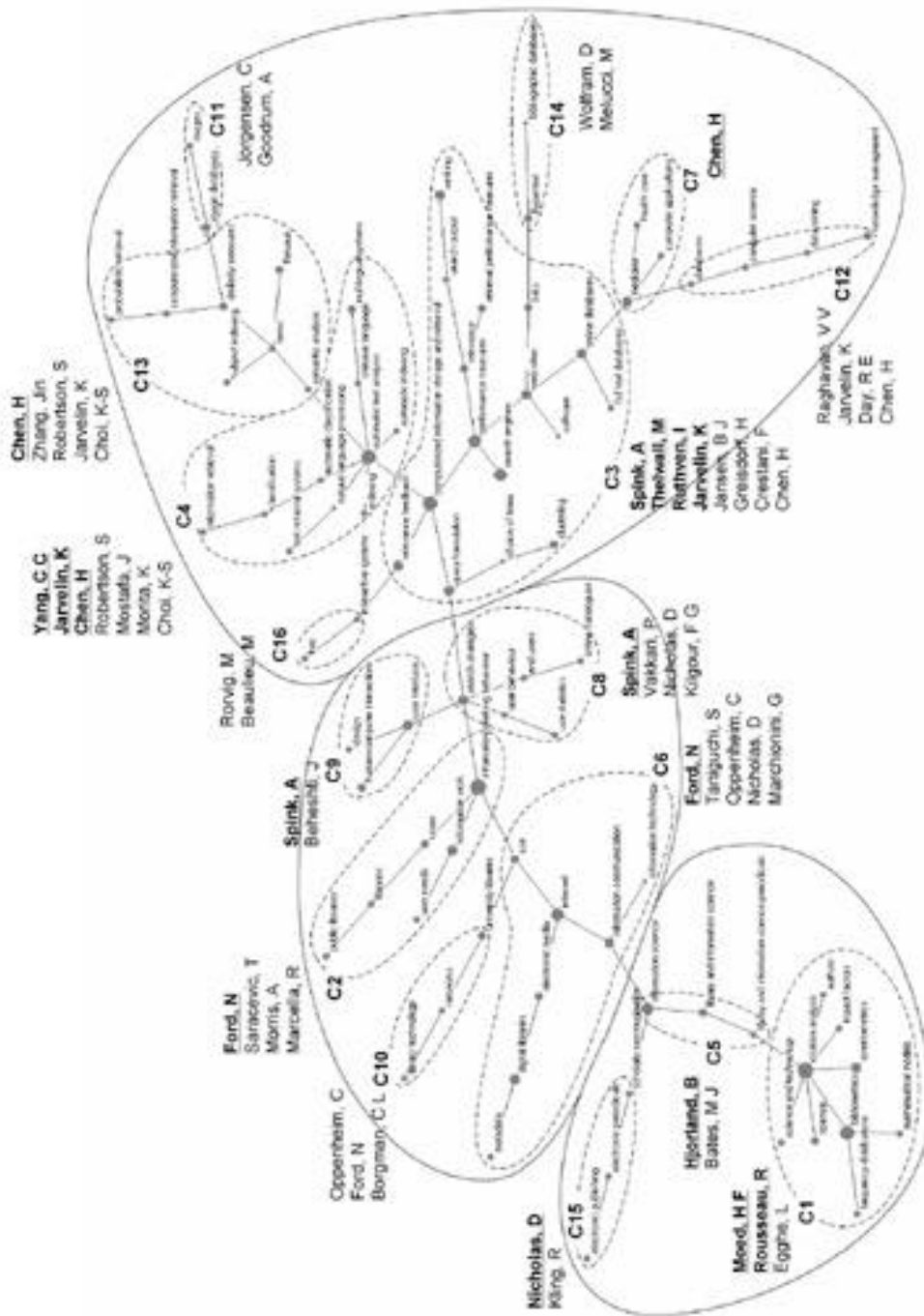
1	library and information science	science and technology
2	library and information science periodicals	scholarly communication
3	information communication	library and information science periodicals

가 (Clustering based Network; CBNNet)  
 가 ( 2007)  
 ‘library and information science’  
 가 1 Pajek(de Nooy et al.  
 2005)  
 17  
 SPSS  
 가  
 Ward  
 4.3  
 가  
 Ward  
 2 16 3  
 < 5> . < 6>  
 Visual Foxpro < 5> Ward  
 16  
 < 8> 2

	( )
Chen, H	C4(automatic text analysis), C7(medicine), C13(data mining)
Ford, N	C2(information seeking behavior), C6(Internet, digital libraries)
Jarvelin, K	C3(computerized information storage and retrieval), C4(automatic text analysis)
Moed, H F	C1(bibliometrics, citation analysis)
Nicholas, D	C15(electronic publishing)
Rousseau, R	C1(bibliometrics, citation analysis)
Ruthven, I	C3(computerized information storage and retrieval)
Spink, A	C3(computerized information storage and retrieval), C8(search strategies), C9(user interface)
Thelwall, M	C3(computerized information storage and retrieval)
Yang, C C	C4(automatic text analysis)







, < 7>  
3  
가 12 1  
3 C1 H. F. Mode  
R. Rousseau가 2  
16 < 6> L. Egghe  
(bibliometrics, Price  
citation analysis) Garfield가  
(information science), (Internet), 가  
(information seeking 가  
behaviour), (search strategies),  
(computerized information  
storage & retrieval)  
(automatic text  
analysis) (databases) 2  
< 7> 3 H. Chen 10  
, , < 8> 10  
93  
< 7> 16 가 5.  
가  
3 12 6  
93  
1

, ( , , CBNet  
 , ) ( , 16 3  
 , , )  
 가 .

,  
 .  
 ,  
 .  
 10  
 1995 2006  
 12 3  
 2,245 가 LISA  
 가 DB

가

Ward



- . 2006. .  
 .  
 23(3): 69-90.  
 . 2005. .  
 . 2007. .  
 . 14  
 147 -  
 154.
- Bhattacharya, Sujit, and Prajit K. Basu. 1998. "Mapping a research area at the micro level using co - word analysis." *Scientometrics* 43(3): 359-372.
- Buter, R.K. and E.C.M. Noyons. 2002. "Using bibliometric maps to visualize term distribution in scientific papers." In *IEEE Proceedings of the 6th International Conference on Information Visualisation*, London, July 2002, pp. 697-705.
- Callon, M., J. Law, and A. Rip. 1986. *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World* London: The Macmillan Press Ltd.
- Callon, M., Jean -Pierre Courtial, W. A. Turner, and S. Bauin. 1983. "From translations to problematic networks: an introduction to co - word analysis." *Social Science Information* 22(2): 191-235.
- Courtial, J. -P., M. Callon, and M. Sigogneau. 1984. "Is indexing trustworthy? Classification of articles through co - word analysis." *Journal of Information Science* 9: 47 -56.
- de Nooy, W., A. Mrvar, and V. Batagelj. 2005. *Exploratory Social Network Analysis with Pajek*. New York: Cambridge University Press.
- Ding, Y., Gobinda G. Chowdhury, and Schubert Foo. 2001. "Bibliometric cartography of information retrieval research by using co - word analysis." *Information Processing & Management*, 37(6): 817-842.
- He, Qin. 1999. "Knowledge discovery through co - word analysis." *Library Trends*, 48(1): 133-159.
- Jacobs, N. 2002. "Co - term network

- analysis as a means of describing the information landscapes of knowledge communities across sectors." *Journal of Documentation*, 58(5): 548-562.
- Jassens, F., J. Leta, W. Gläzel, and B. De Moor. 2006. "Towards mapping library and information science." *Information Processing & Management*, 42(6): 1614-1642.
- Kostoff, R. N. 1993. "Database tomography for technical intelligence." *Competitive Intelligence Review*, 4(1): 38-43.
- Kostoff, R. N., R. Tshiteya, K. M. Pfeil, J. Humenik, and G. Karipis. 2005. "Power source roadmaps using bibliometrics and database tomography." *Energy*, 30(5): 709-730.
- McCain, Katherine W. 1995. "R&D themes in information science: A preliminary co-descriptor analysis." In M. Koenig and A. Bookstein (eds.), *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*, Rosary College, Pine Forest, IL, June 7-10, 1995, (Medford, N.J.: Learned Information, 1995), pp. 275-282.
- Noyons, E.C.M., and A.F.J. van Raan. 1998. "Monitoring scientific developments from a dynamic perspective: self-organized structuring to map neural network research." *JASIS* 49(1): 68-81.
- Porter, Alan L., and Scott W. Cunningham. 2005. *Tech Mining: Exploiting New Technologies for Competitive Advantage*, Hoboken, NJ: John Wiley & Sons, Inc.
- Spasser, M. A. 1997. "Mapping the terrain of pharmacy: co-classification analysis of the international pharmaceutical abstracts database." *Scientometrics* 39(1): 77-97.
- Tijssen, R.J.W. 1992. "A Quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research." *Research Policy*, 21: 27-44.
- Todorov, R. 1989. "Co-classification analysis for science mapping: an example from superconductivity." In A. F. J. van Raan, A. J. Nederhof, and H. F. Moed. (eds.) *Science and Technology Indicators* Leiden: DSOW Pr.,

- University of Leiden, pp. 263-270
- Todrov, R., and M. Winterhager. 1990. "Mapping Australian geophysics: a co-heading analysis." *Scientometrics* 19(1-2): 35-56.
- Voutihainen, A. 1993. "NPto0: a detector of English noun phrases." In Proceedings of the workshop on very large corpora Columbus Ohio, Ohio University, June 1993.
- Watts, R. J., A. Porter, and B. Minsk. 2004. "Automated text mining comparison of Japanese and USA multi-robot research." Ed. by A. Zanasi, N. Ebcken & C. Brebbia. *Data Mining*. WIT Press, pp. 61-74.
- Widhalm, Clemens, Ute Gigler, Alexander Kopcsa, and Edgar Schiebel. 2001. "Co-occurrence and knowledge mapping to identify hot topics and key players in the field of mobility and transport." In Proceedings of the 8th Conference of the International Society for Scientometrics and Informetrics Sydney, July 2001, pp. 751-758.

