Survey on Convergence Study Fields Focusing on Papers and Patents

Tae-Seok Lee*, Su-Mi Shin**
*,**KISTI, Korea,
E-mail : tsyi@kisti.re.kr*, sumi@kisti.re.kr**

1. Introduction

KISTI (Korea Institute of Science and Technology Information) is operating NDSL, a website that provides scientific and technological information services. NDSL is the largest website in Korea that constructs and provides information on papers and patents. That is why a great number of researchers are using the website to search for information. This paper studies how to analyze and measure the search log, connectivity, and association in the field of research. [3].

2. Methods of association analysis

After identifying the session of the researcher who logged in in NDSL, it was assumed that the researcher searched the data for the same purpose in one session of visit. The frequency of simultaneous appearances of papers and patents was investigated using Science and Technology Standard Classification and IPC large classification, respectively [4]. Apriori was selected as the analysis technique, and R was used as the statistical analysis tool [5].

[Table 1] Standard fields of science and technology in NDSL

A. Mathematics,	B. Physics,	C. Chemistry,
D. Bioscience,	E. Earth science,	F. Mechanics,
G. Material,	H. Chemical process,	I. Electrics and electronics,
J. Information,	K. Telecommunication,	L. Agriculture, forestry and fisheries,
M. Health and medical,	N. Environment,	O. Energy resource,
P. Atomic energy,	Q. Construction and transportation,	
R. Air, space, astronomy and oceanography,		
S. Technical innovation/science and technology policy,		V. Business and economics,
W. Liberal art,	X. Sociology,	Y. Art,
Z. Others		

3. Analysis on the association between research fields of papers

The result of extracting showed that the fields of papers used in the search session, 113,291 cases of observation data were generated. As use of NDSL increased, information on paper using session also increased. As it was found that each session used an average of 2.3 sessions simultaneously, the number increased from 2012 by 2.26%. This shows that researchers are using data from more diverse fields. In detail, papers in health and medical (M), agriculture, and forestry and fishery (L) fields were actively used.

In the result of association analysis, 245 rules excluding duplicated rules were generated. While the average support rate decreased as the number of subject sessions increased from the previous year, the average confidence level slightly increased. With an increase of 0.1, the average lift was 2.455.

To explore the linking rules, data with support rates of 0.001 or higher were filtered from cases with 100 or more “uses” from a total of 110,000 cases. Among the rules that had confidence levels of 0.6 or higher, 10 rules were identified in order of level of lift. When the fields used at the same time were arranged based on the order of level of lift, the top 3 rules of being used together are “Mechanics, Business and economics, Sociology, and Information”, followed by “Chemical process, Agriculture, forestry and fisheries, Health and medical, and Bioscience” and “Chemistry, Material, Chemical process”. Key words such as “Robot, Automobile” appeared in the first rule, along with “Microbe, Antioxidant, Cosmetics, Kimchi, Coffee, Food, Senior Citizens, Obesity, Diabetes, Ginseng”, showing that papers on chemistry, health and medical, and bioscience were used together. The order of confidence was “Chemistry, Health and medical, Agriculture, forestry and fisheries” (0.82), “Chemical process, health and medical, and Agriculture, forestry and fisheries” (0.81), and “Chemistry and Agriculture, forestry and fisheries”(0.75), just like in 2012. The result showed that more studies related to health and medical and food are being made.

[Table 2] Association rule of fields used in papers in 2013

Number	Association result	support	Confidence	Lift
1	{F,V,X}=>{J}	0.005879	0.619643	4.864431
2	{H,L,M}=>{D}	0.009394	0.625847	4.707166
3	{C,G}=>{H}	0.005506	0.65261	4.70637
4	{C,M,Z}=>{D}	0.005049	0.600806	4.518833
5	{D,Z}=>{L}	0.026191	0.829622	3.575894

4. Association analysis of patent IPC

In the extraction of fields of patents used in search sessions, 67,567 cases of observation data were generated. Use of patents increased in 2013, and in detail, use of A (Daily essentials), B (Treatment and operation, Transportation), and G (Physics) fields largely increased. In the result of association analysis, 245 association rules were generated excluding duplicated ones. A comparison of aspects of the association rules with the result in 2012 showed that the average support decreased as the number of subject sessions increased, while the average confidence slightly increased and average lift was 2.455, increased by 0.1.

To explore association rules, the rules were examined according to their levels of improvement for rules with confidence levels of 0.36 or higher. When the rules with support levels of 0.005 or higher were selected, the fields in the rules found to be used combined in 337 or more cases. In the result, the rules with the highest confidence levels were “Chemical and metallurgy, Mechanic engineering, and Processing and operation”, “Daily essentials, Processing and operation, and Chemistry and metallurgy”, “Chemistry and metallurgy, Physics, and Processing and operation”, and “Chemical and metallurgy, Physics, and Electrics” in order of use. The order was “Chemical and metallurgy, Mechanic engineering, Processing and operation”, “Processing and operation, Electrics, Mechanic engineering”, “Processing and operation, Physics, Electrics”, and “Chemical and metallurgy, Physics, Processing and operation” when ranked according to the level of improvement, which is similar to that of the previous year. The fields of patent not used in much of the research were “Fabric and paper” and “Fixed structure”. Some of the key patents researched were Nano structure fabrication, Waste water phosphate, Biodegradation, Lithium secondary battery, and lamination film. The main key words were Antibiotic material, Lysine, Gene cloning, Micro bubble, Bubble spray, Automobile, Screw, and Nut.

[Table 3] Rules of association between fields of patent used in research in 2013

Number	Association rule	support	Confidence	Lift
1	{C,F}=>{B}	0.006900	0.582524	3.633267
2	{B,H}=>{F}	0.005674	0.226762	3.063832
3	{B,G}=>{F}	0.005060	0.224872	3.038299
4	{B,G}=>{H}	0.008587	0.381601	2.636442
5	{C,G}=>{B}	0.007590	0.413937	2.581772

5. Conclusion

Association analysis was used to find in what main fields the convergence studies are carried out from the association of fields used in papers by researchers and patents. In the result, the fields showing high associations were found. The result of this study can be used to figure out the trend of using papers and patents in different disciplines following the new stream of research and to search recommendation and customized service. In the future, further studies are required on models to predict the stream of research and use of information [1][2].

6. References

[1] Hassan, A., Jones, R. and Klinkner, K.L., “Beyond DCG: user behavior as a predictor of a successful search.” Proc. WSDM, pp. 221-230. 2010.
[2] Ryen W. White, Paul N. Bennett, Susan T. Dumais, “Predicting Short-Term Interests Using Activity-Based Search Context”, Proc. ACM, pp.1009-1018, 2010.
[3] Soo-Sang Lee, Cheng-Guang Wei, “A Study on the Search Behavior of Digital Library Users”, Journal of Korean Library and Information Science Society, vol. 10, no. 4, pp.139-158, 2009.
[4] Tae-Seok Lee, Do-Heon Jeong, Young-Su Moon, Minsoo Park, Mi-hwan Hyun, “An Analytic Study on the Categorization of Query through Automatic Term Classification”, JIPS, vol. 19-D, no. 2, pp.133-138, 2012.
[5] H.Toivonen, “Sampling large databases for association rules”, In the VLDB Journal, pp 134-145, 1996.