Automatic Subject Classification of Korean Journals

Seon-Heui Choi

Department of Domestic Information
Korea Institute of Science and Technology Information, Daejeon, South Korea

Byung-kyu Kim

Department of Domestic Information
Korea Institute of Science and Technology Information, Daejeon, South Korea

ABSTRACT

Subject classification of journals is important because it can be utilized for the improvement of scholarly information services and analysis by research area. The classification by experts in a subject area wastes a lot of time and expense. On the other hand, the simple classification with basic information, such as the journal title has limitations. To solve this problem, this paper suggests the automatic classification of Korean journals using the SCI journals information cited by Korean journals, and an analysis of the classification result. In particular, this study adopted the WoS subject categories for classification to support the base for comparison between the Korean citation database and the global citation database (KSCI vs. SCI).

Key words: Korean Journal, Subject Classification, KSCI, SCI, WoS Subject Category.

1. INTRODUCTION

Subject classification of journals is necessary because it can be utilized for the improvement of scholarly information services and analysis of intellectual structure by research area. The classification by experts in a subject area wastes a lot of time and expenses. On the other hand, the simple classification with basic information, such as the journal title, has limitations. To solve this problem, this paper suggests the automatic classification of Korean journals using the Science Citation Index (SCI) journals information cited by Korean journals, and analysis the classification result. For this, this study adopts the Web of Science (WoS) subject categories for classification to support the base for comparison between the Korean citation database and the global citation database (KSCI vs. SCI). To show the result effectively, we maps WoS subject categories to 14 subjects by Leydedorff [5].

2. RELATED RESEARCH

Thomson Reuter's (before ISI) WoS is the first citation index database for citation analysis and the most well-known in the world. Thomson Reuter's WoS covers a variety of

disciplines: Science, Social science and Arts & Humanities, and supports citation indices such as SCIE, SSCI and A&HCI (classified by subject respectively). Specifically, all of the journals (8,100) of WoS are classified by subject categories of WoS and this is one of reasons for much research related citation analysis. Dewey Decimal Classification (DDC) is used to classify journals widely in libraries [4]. Mahdi and Joorabchi suggested the method automatic subject classification of scientific literature using citation metadata based on DDC [6]. We adopted the subject categories of WoS for subject classification of Korean journals instead of DDC in this study.

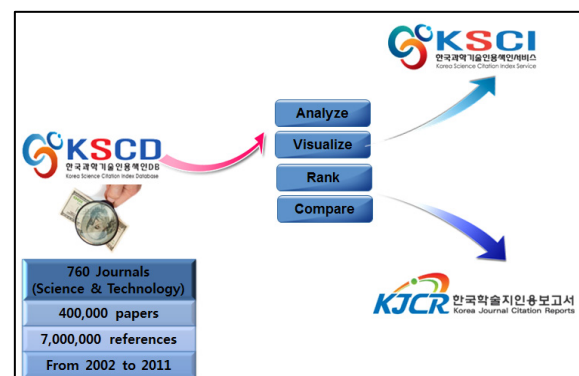## 3. AUTOMATIC SUBJECT CLASSIFICATION OF KOREAN JOURNALS

Fig. 1. KSCD, KSCI and KJCR

We utilized KSCD (Korea Science Citation Database), which supports the base for KSCI of KISTI (Korea Institute of Science and Technology Information). KSCI covers all of the important Korean Journals and supports the same level of service as WoS and JCR (Journal Citation Reports). KSCD covers the period from 2002 to 2013, and there are 760 journal titles, but it has changed with the criteria of journal selection annually. Figure 1 shows the relation among KSCD, KSCI, KJCR (Korea Journal citation Reports).

The previous studies for citation analysis of Korean researchers, using KSCD, revealed some significant findings, as described below. First, journals are the most cited document type among scholarly works (e.g books, research reports, thesis, proceeding etc.). Second, the citation rate of global journals to domestic journals is 8:2. Third, Korean researchers prefer journal papers in the global citation index service ,such as SCI or SCOPUS, over non-indexed journal papers for citing references [1]-[3]. Subject classification information for both SCI and SCOPUS journals can be obtained very easily from their own service sites.

The result of related research provided the motivation and background for this study. In particular, this paper utilized SCI journal, which is the most widely used in the world. WoS subject categories for SCI journals consist of 180 categories, and 1.6 subject categories are allocated on average per journal in JCR 2010.
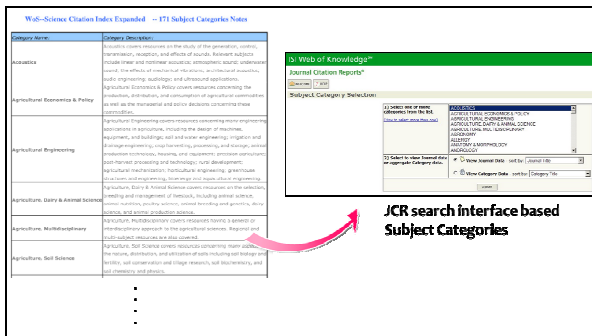

Fig. 2. WoS Subject Categories

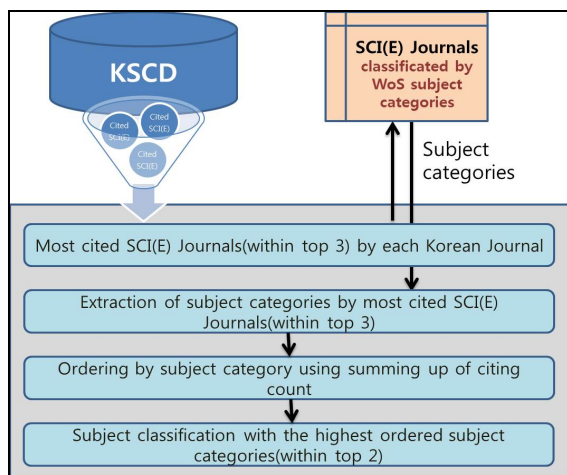Fig. 3 shows the method and the process for automatic subject classification.

Fig. 3. Process of Automatic Subject Classification

The detail on Fig. 3 is as below.

1. Investigation of WoS subject categories for SCI(E) Journals.
2. Investigation of most cited SCI(E) journals (within top 3) with cited count of each KSCI journal
3. Searching Wos Subject categories of SCI(E) journals from step 2 in step 1 result list.
4. Summing up cited count by subject category found in step 3.
5. Classification of subject of KSCI journals based on step 4. (allocation of two subject categories in descending order by summing up cited count)

KSCI's 743 journals were classified by subject categories once they were submitted, as figure3 shows, and 3 journals, which do not cite any SCI(E) journals were excepted. Verification for the automatic classification of subject categories was accomplished using 90 Korean SCI(E) journals. The matching percentage was 87.8% through comparing the answer (already classified subject categories by WoS) and the test (automatic subject classification). Specifically, the percentage when checking the answer among subjects (the average number of subjects categories of Korean SCI(E) journals is 1.37) was included in two test subjects. Through the verification test, we confirmed the precision of the research result. Table1 shows the result about 20 journals among the classified KSCI 743 journals.

Table 1. Classified KSCI Journals (20 among 743)

Index	Journal (in KSCD)	WoS Subject Category #1	WoS Subject Category #2
1	Structural Engineering and Mechanics	ACOUSTICS	ENGINEERING, MECHANICAL
2	Asian-Australasian Journal of Animal Sciences	AGRICULTURE, DAIRY & ANIMAL SCIENCE	FOOD SCIENCE & TECHNOLOGY
3	Allergy, Asthma & Immunology Research	ALLERGY	IMMUNOLOGY
4	International Journal of Control, Automation, and Systems	AUTOMATION & CONTROL SYSTEMS	ENGINEERING, ELECTRICAL & ELECTRONIC
5	Experimental and Molecular Medicine	BIOCHEMISTRY & MOLECULAR BIOLOGY	MULTIDISCIPLINARY SCIENCES
6	BMB Reports	BIOCHEMISTRY & MOLECULAR BIOLOGY	MULTIDISCIPLINARY SCIENCES
7	Journal of Microbiology and Biotechnology	BIOTECHNOLOGY & APPLIED MICROBIOLOGY	MICROBIOLOGY
8	Yonsei Medical Journal	CARDIAC & CARDIOVASCULAR SYSTEMS	HEMATOLOGY
9	BioChip Journal	CHEMISTRY, ANALYTICAL	BIOCHEMICAL RESEARCH METHODS
10	Journal of Medicinal Food	CHEMISTRY, APPLIED	FOOD SCIENCE & TECHNOLOGY
11	Journal of the Korean Society for Applied Biological Chemistry	CHEMISTRY, APPLIED	FOOD SCIENCE & TECHNOLOGY
12	Journal of Ginseng Research	Chemistry, Medicinal	Plant Sciences
13	Bulletin of the Korean Chemical Society	CHEMISTRY, ORGANIC	CHEMISTRY, MULTIDISCIPLINARY
14	Journal of Industrial and Engineering Chemistry	CHEMISTRY, PHYSICAL	ENVIRONMENTAL SCIENCES
15	Journal of clinical neurology	CLINICAL NEUROLOGY	PERIPHERAL VASCULAR DISEASE
16	Journal of Korean Neurosurgical Society	CLINICAL NEUROLOGY	SURGERY
17	KSII Transactions on Internet and Information Systems (TIIS)	COMPUTER SCIENCE, INFORMATION SYSTEMS	ENGINEERING, ELECTRICAL & ELECTRONIC
18	Computer and concrete	CONSTRUCTION & BUILDING TECHNOLOGY	MATERIALS SCIENCE, MULTIDISCIPLINARY
19	Tissue Egineering and Regenerative Medicine	ENGINEERING, BIOMEDICAL	MATERIALS SCIENCE, BIOMATERIALS
20	The Korean Journal of Chemical Engineering	ENGINEERING, CHEMICAL	CHEMISTRY, PHYSICAL

4. COMPARISON OF SCIENTIFIC OUTPUT BY SUBJECTS (KSCI VS. SCIE)

The result of automatic subject classification for Korean journals using WoS subject categories make it easier to compare the status of Korean scientific output and global scientific output(KSCI vs. SCI(E)). Fig. 2 shows the differences of the analysis result of the structures of scientific output (KSCI 718 journals, SCI(E) 7,995 journals in 2010 publication) by 14 subjects, as Leydesdorff suggested, and we mapped the WoS subject categories of the 14 subjects [5]. The considerable differences of subjects between KSCI and SCI can be found in below figure 4. First, Biomedical Sciences shares the highest portion with 19.1% in SCI(E) publication. On the other hand, Computer Sciences shares the largest portion with 17.9% in KSCI publication. Second, the disciplines such as Engineering, Environmental Sciences and Agriculture share much higher portions comparatively in KSCI than SCI(E). These meaningful differences could support discussion materials or form an important basis for study concern about scholarly activities. All equations must be typed or written neatly in black. They should be numbered consecutively throughout the text. Equation numbers should be enclosed in parentheses and flushed right. Equations should be referred to as Eq. (X) in the text, where X is the equation number. In multiple-line equations, the number should be given on the last line.
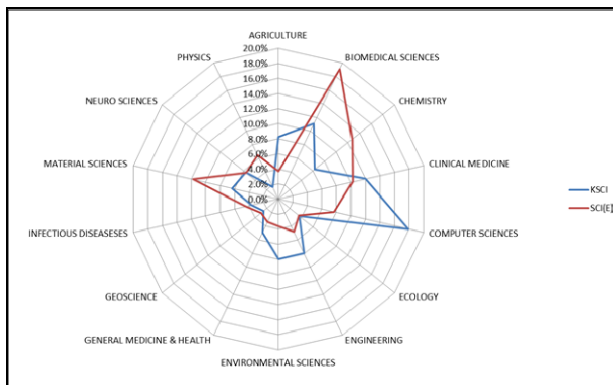

Fig. 4. Comparison of Scientific Output between KSCI and SCI(E)

5. CONCLUSION AND FUTURE WORK

This paper suggested the method of automatic subject classification for Korean scientific journals using KSCD. Korean researchers cite global journals, such as SCI, much more than domestic journals. The assumption that the relationship between cited journals would be strong provided the motivation and background to this research.

In particular, the idea for automatic subject classification is to use SCI(E) journals with WoS subject categories cited by Korean journals. The result of using this suggested method showed 87% matching rate through the verification of 90 Korean SCI(E) journals. Also, the considerable differences of subjects between KSCI and SCI(E) were revealed through comparison between KSCD vs. SCI(E) databases, on which the result of automatic subject classification were based.

This paper has limitations due to the utilization of the top 3 cited SCI(E) journals by each Korean journal. In further research, we plan to use all of the cited SCI(E) journals by each Korean journal to improve the precision of the result of subject classification and analyze the specific differences by sub-subjects between KSCI vs. SCI(E) in depth.

REFERENCES

[1] B. K. Kim, M. Y. Kang, S. H. Choi, S. Y. Kim, B. J. You, and J. D. Shin, "Citing behavior of Korean scientists on foreign journals in KSCD," Journal of the Korean Society for information Management, vol. 28, no. 2, 2011, pp. 117-133

[2] S. H Choi, B. K. Kim, M. Y. Kang, B. J. You, J. W. Lee, and J. W. Park, "A study of citing patterns of Korean scientists on Korean journals," Journal of the Korean Society for information Management, vol. 28, no. 2, 2011, pp. 97-115.

[3] H. N. Choi, B. K. Kim, Y. I. Jung, and S. H. Choi, "Korean scholarly information analysis based on Korea Science Citation Database (KSCD)," Collnet Journal of Scientometrics and Information Management, vol. 7, no. 1, 2013, pp. 1-33

[4] M. Dewey, "Dewey Decimal Classification (DDC)," Online Computer Library Center (OCLC).

[5] L. Leydesdorff and I. Rafols, "A global map of science based on the isi subject categories," Journal of the American Society for Information Science and Technology, vol. 60, no. 2, 2009, pp. 348–362.

[6] A. E. Mahdi and A. Joorabchi, "Automatic Subject Classification of Scientific Literature Using Citation Metadata," Digital Enterprise and Information Systems, Springer Berlin Heidelberg, 2011, pp. 545-559.

[7] KSCD: Korea Science Citation Database, http://ksci.kisti.re.kr

[8] KSCI: Korea Science Citation Index, http://ksci.kisti.re.kr

[9] Web of Science, http://apps.webofknowledge.com

[10] Journal Citation Reports (JCR), http://thomsonreuters.com/journal-citation-reports/

Seon-Hee Choi

She is the head of Department of Domestic Information at Korea Institute of Science & Technology Information (KISTI), Republic of Korea. She received the B.S., M.S in library and information science from Yonsei university, Korea in 1992, 1995 respectively. Her main research interests include advanced information service and Science Citation analysis in Korea.

Byung-Kyu Kim
He received the B.S., M.S in computer science from Chungnam National university, Korea in 2001, 2003 respectively. Since then, he has been with Department of Domestic Information, Korea Institute of Science & Technology Information. His main research interests include metadata management system and applications for metadata processing.