ISSN (Online) : 0974-5645
ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(S1), 452–456, January 2015* DOI: 10.17485/ijst/2015/v8iS1/59413

Automatic Subject Classification of Korean Journals based on KSCD

MuYeong, Kang[1], Jae-Do, Shin[2] and Byungkyu Kim[3*]

[1]Information Service Center, Korea Institute of Science and Technology Information, Daejeon, 305-806, South Korea
[2]Department of Computer Engineering, ChungNam National University, Daejeon, 305-764, South Korea
[3]Division of Policy Research, Korea Institute of Science and Technology Information, Daejeon, 305-806, South Korea; bk.kim@kisti.re.kr

Abstract

Journal subject classification is important in terms of being used for scholarly information service and foundation of disciplinary research analysis. Subject classification by subject matter experts or journal information takes a significant amount of time or does not provide accurate information about subject respectively. In order to overcome these current problems, this research suggested automatic subject classification method by using SCI journal information cited by domestic science and engineering journals, and it also investigated the classification results. We found that using the entire cited academic journals has a better accuracy rate than using the most cited three or five academic journals, and this research showed that the more academic journals included in the analysis the more accurate the rate. Especially, this research utilised the subject category of Web of Science as the standard of subject classification and provided foundations for comparing subject category structure in academic research results in KSCI and SCI.

Keywords: Automatic Subject Classification, Korean Journals, KSCI

1. Introduction

Traditionally, library system literature was classified by its subjects and utilized as important indexing information in order to manage literature efficiently and provide information service. In digital scholarly information service, subject classification information can provide efficient information retrieval service but also it can provide foundational information for understanding diverse research activities and trends in research disciplines.

Subject classification can be divided into two main categories: (1) classification by subject matter experts, and (2) simple classification based on basic literature information such as title and key words. The subject classification by subject matter experts can provide more precise classification information; however, it may take a significant amount of effort and costs in finding appropriate experts and their actual works in subject classification. The simple classification using the journal title can expedite the subject classification process; however, it does not provide detail and accuracy in subject classification. Therefore, it is necessary to develop a new subject classification method which can provide prompt and accurate subject classification service.

This research suggests an automatic subject classification method based on the cited literature in academic journals, and it also analyzes the research activities in Korea by their subjects and further compares the subject structure with SCI(E) database.

2. Related Literature

A well-organised subject classification table is necessary to perform the subject classification of scholarly articles. Libraries have utilised DDC (Dewey Decimal Classification) in order to categorise literature, and

*Author for correspondence

publishers, organisations, and services all have their own subject classification tables[1]. For example, Web of Science which is a well-known scholarly information citation indexing service categorises the research areas of science and engineering into 176 groups and classifies each individual article into its designated category[2]. If there is a journal that has more than one subject area it can be categorised into different research categories.

Leydesdorff and colleagues in 2009 categorised the specific research areas in Web of Science into 14 major subject categories and created a scientific research activity map based on the data in Web of Science[3]. Mahdi and colleagues (2010, 2011) suggested a subject classification method based on citation information, and they used DDC as their subject classification[4,5]. Glänzel and colleagues (1999) employed a reference analysis method in order to supplement the existing subject classification method in which some major journals are categorised into different research areas due to their diversities in research subjects[6]. Gómez-Núñez and colleagues (2011) conducted research to improve the subject classification table of the SJR (SCIMAGO Journal & Country Rank), which is SCOPUS's citation indexing service, by using reference analysis[7–9].

The prior study by Choi[19] for this research performed automatic subject classification research with Korean academic journals by using the cited journal information in SCI, and it compared the subject structure between the KSCD (Korea Science Citation Database) database and SCI database[10]. The prior research by Choi[19] employed only three SCI(E) journals which were mostly cited in each journal for subject classification; however, this research employed all the SCI journal information in order to secure the accuracy of this study[10].

3. Data Collection And Method

In order to automatise the subject classification for Korean academic journals, we employed the KSCD, which is the KSCI (Korea Science Citation Index) database provided by KISTI (Korea Institute of Science and Technology Information)[11]. We used KSCD since it has the most reliable and high-quality database in the fields of science and engineering among citation indexing services in Korea[12,13]. Lee and colleagues conducted an experiment in which they applied KSCD citation information into JCR (Journal Citation Reports), and they demonstrated that the ranking of JCR had changed by the application of KSCD citation indexing[14]. KSCI as a well-known citation indexing service includes major journals in the field of science and engineering, and it also provides KJCR (Korea Journal Citation Reports) such as Web of Science and JCR[15,16]. KSCD's citation index covers from 2002 to 2013, and it currently has 834 science and engineering journals in Korea. There are some changes depending on the selection criteria in each year[12]. Figure 1 describes the relationship among KSCD, KSCI and KJCR.

Prior studies researching Korean scholars' citation behaviours using KSCD found that academic journals are the most cited scholarly work. In addition, the analysis of cited journals shows that the portion between international journals and Korean journals is approximately 8:2, and a good number of those international journals are registered under SCI and SCOPUS, whose subject classification information can be accessed through their websites[17,18].

This result provided the motivations and background information for this research. Particularly, this research employed SCI academic journals, which were widely cited and used internationally. The subject classifications in the science and engineering fields for Web of Science consist of 176 subject categories, and SCI academic journals have 1.55 subject categories on average (JCR, 2013). Figure 2 describe the research method and automatic subject classification procedure.

This research utilised the entire cited SCI academic journals compared to the prior study by Choi[10], which only used three of the most cited journals for their subject classification in academic journals. This research also conducted subject classification for five of the most

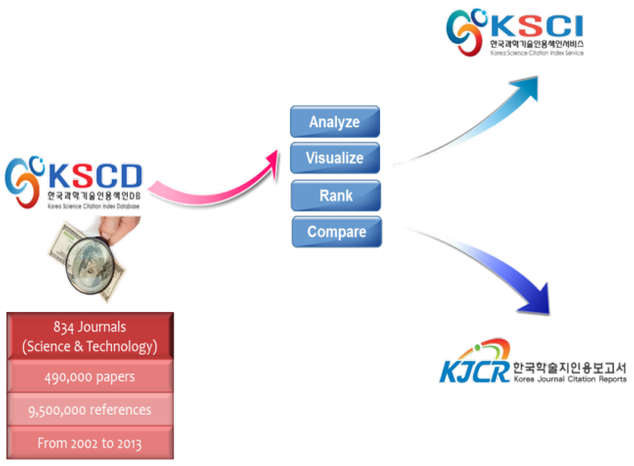
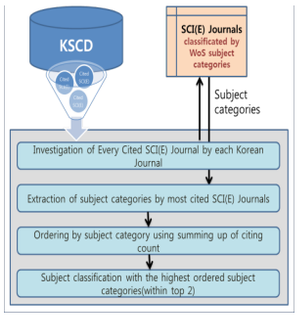


Figure 1. KSCD, KSCI and KJCR.



1. Investigation of WoS subject categories for SCI(E) Journals.
2. Investigation of most cited SCI(E) journals with cited count of each KSCI journal
3. Search 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)

Figure 2. Process of automatic subject classification.

cited journals in order to capture more accurate trend as the number of cited academic journals has recently increased[19].

4. Results

By using the automatic subject classification method suggested in this research, we developed the subject classification table for the 654 academic journals from KJCR 2013 based on the subject classification table by Web of Science (WoS). In order to examine the reliability of the automatic subject classification results, we used 94 Korean SCI(E) academic journals whose subjects were already defined by WoS. We compared the subject classification by WoS to the automatically selected subject classification by evaluating whether the WoS subject classification of SCI(E) Korean academic journals (1.37 on average) were included in the automatically selected subject classification.

We found that they match with each other in 100.0% (Table 1). In addition, we found that using the entire cited academic journals has a better accuracy rate than using the most cited three or five academic journals, and this research showed that the more academic journals included in the analysis the more accurate the rate.

By using the automatic subject classification result based on WoS subject categories, we can compare the result to global scholarly research activities. Based on the WoS subject category for each academic journal, we provided the portion map of the subject component for KJCR 2013 and JCR 2013 (percentage based on the number of publications—KSCI: 654 journals, and SCI(E) 8,411 journals) according to the 14 subject categories suggested by Leydesdorff (Figures 3 and 4).

Figure 3 shows the comparison between KJCR 2013 and JCR 2013, and it also provides an additional analysis about Korean scientists' SCI(E) articles by subject[20]. In the case of SCI(E), Biomedical Sciences have 20% of the total research publications; however, in Korea Computer Sciences have 22% of total publications. In addition, KSCI has bigger portions than SCI(E) for Engineering, Environmental Sciences, and Agricultural Sciences. Figure 3 shows that Korean scientists' SCI article analysis

Table 1. Accuracy rate comparison in different automatic subject classification methods

Automatic Subject Classification Method	Accuracy Rate (Based on 94 Korean SCI(E) Academic Journals)	Reference
3 mostly cited academic journals	87.8%	Result from (Choi et al., 2014)
5 mostly cited academic journals	90.4%	
Entire cited academic journals	100.0%	

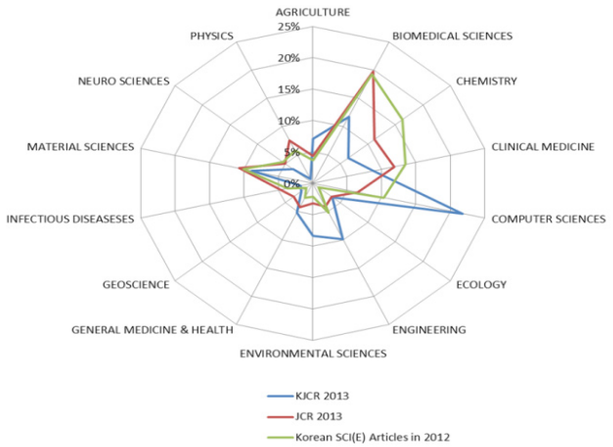


Figure 3. Comparison of scientific output among KJCR, JCR and Korean SCI(E) articles.

by subject based on the comparison between KJCR 2013 and JCR 2013 is similar to JCR subject structure, but it is not identical. Figure 4 shows the research article analysis from 94 Korean SCI(E) academic journals as well as the comparison between KJCR 2013 and JCR 2013.

Similar to Figure 3 and 4 shows that the portion of Korean SCI journal articles for each discipline is similar to JCR; however, it is similar to KJCR in the case of physics. As a result, KJCR and JCR are significantly different in terms of the portion of subject categories as you can see the analysis results from Figure 3 and Figure 4, and this significant difference can be used for the future discussion and/or important evidence for future studies. Figure 5 shows the analysis results by specific subject in order to further analyze the results of Figure 3, and we also conducted the comparison study with top 10 specific subjects for each. This suggests that Multidisciplinary Sciences for JCR and Engineering, Electrical & Electronic for KJCR have a significant portion for each.

The Figure 6 shows the comparison analysis results by specific subject for Computer Sciences and Engineering, where KJCR has the bigger portion than JCR in terms of classified journals. This result shows that KJCR and JCR have different subject categories in terms of specific subject.

5. Conclusion

This research suggests the automatic subject classification method for Korean science and engineering academic journals based on the KSCD, which is the well-known

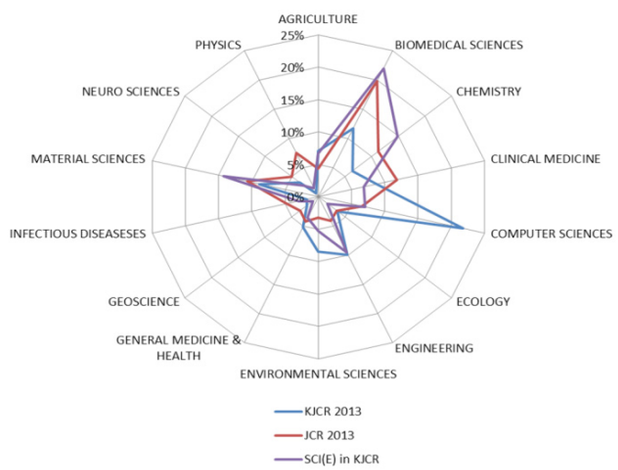


Figure 4. Comparison of scientific output among KJCR, JCR and Korean SCI(E) journals.

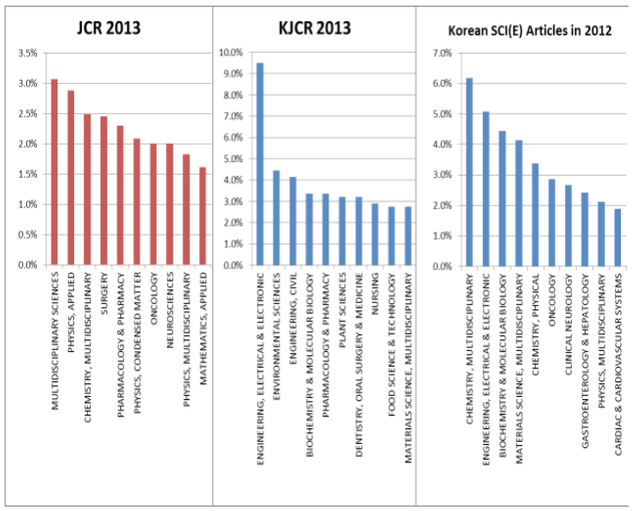


Figure 5. Comparison of scientific output among KJCR, JCR and Korean SCI(E) articles.

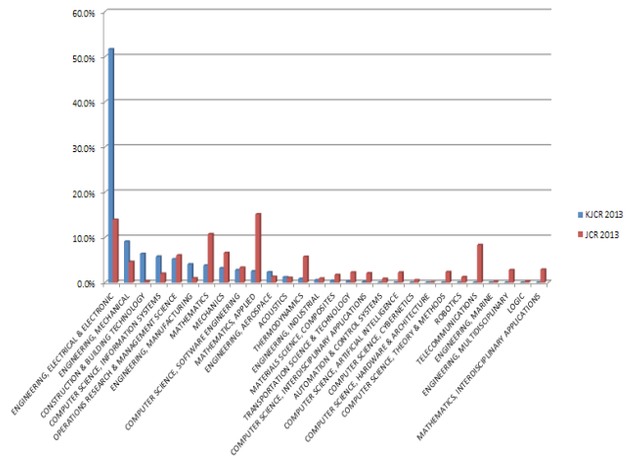


Figure 6. Comparison of computer sciences & engineering output between KSCI and SCI(E).

Korean citation index database. The assumption that Korean academic journals would cite the international academic journals, especially those registered under SCI, provided the motivation and background for this research. In particular, the automatic subject classification is focused on the SCI academic journals that were mainly cited in the Korean academic journals and used their subject classification information in Web of Science. As suggested above, we conducted the automatic subject classification for Korean academic journals and compared the results to 94 Korean SCI(E) academic journals which were already classified by WoS subject categories. The comparison result shows that they exactly match each

other (100%), and this shows the accuracy and reliability of the proposed automatic subject classification method.

The proposed method has a higher accuracy rate than the prior method, which only used the most cited SCI academic journals. In addition, based on the automatic subject classification result, we confirmed that there are significant differences between KSCI and SCI in terms of their subject categories.

The future research can extend the automatic subject classification based on KSCD to measuring the convergence rate among disciplines in terms of journals, major subjects, and specific subjects and further apply the results into scholarly information service.

6. References

1. Dewey M. Dewey Decimal Classification (DDC). (Online Computer Library Center (OCLC); 1876–2010. Available from: http://www.oclc.org/us/en/dewey
2. Web of Science. Available from: http://apps.webofknowledge.com
3. Leydesdorff L, Rafols I. A global map of science based on the ISI subject categories. J Am Soc Inform Sci Tech. 2009; 60(2):348–62.
4. Mahdi AE, Joorabchi A. A citation-based approach to automatic topical indexing of scientific literature. J Inform Sci. 2010; 36(6):798–811.
5. Mahdi AE, Joorabchi A. Automatic subject classification of scientific literature using citation metadata. Digital enterprise and information systems; 2011; p. 545–59.
6. Glanzel W, Andras S and Czerwon HJ. An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. Scientometrics. 1999; 44(3):427–39.
7. Gomez-Nunez AJ, Vargas-Quesada B, de Moya-Anegon F, Glanzel W. Improving SCImago Journal & Country Rank (SJR) subject classification through reference analysis. Scientometrics. 2011; 89(3):741–58.
8. Scimago Lab. Scimago Journal & Country Rank. 2007; Available from: http://www.scimagojr.com.
9. Elsevier; Scopus. Available from: http://www.scopus.com/home.url.
10. Choi SH, Kim BK. Automatic Subject Classification of Korean Journals. International Journal of Contents. 2014; 10(1):43–6,.
11. KSCI: Korea Science Citation Index. Available from: http://ksci.kisti.re.kr.
12. KSCD: Korea Science Citation Database.
13. Yang K, Lee JW. Comparative analysis of citation indexes in Korea. Proceedings of the 8th International Conference on WIS; 2012; p. 756–62.
14. Lee JW, Yang KD, Kim BK, You BJ. Analysis of Korea science citation database's effect on JCR. J Inform Manag. 2012; 43(3):23–41,.
15. KJCR: Korea Journal Citation Report. Available from: http://kjcr.kisti.re.kr.
16. Journal Citation Reports (JCR). Available from: http://thomsonreuters.com/journal-citation-reports.
17. Choi SH, Kim BK. Kang M, You BJ, Lee J, Park JW. A study of citing patterns of Korean scientists on Korean journals. J Korean Soc for Informat Manag. 2011; 28(2):97–115.
18. Kim BK, Kang MY, Choi SH, Kim SY, You BJ, Shin JD. Citing behaviour of Korean scientists on foreign journals in KSCD. J Korean Soc for Informat Manag. 2011; 28(2):117–33.
19. Choi SH, Kim BK. Automatic subject classification of Korean journals. Internat J Cont. 2014; 10(1):43–6.
20. Kim BK, So MH, Choi SH. Korea's STEM research analysis based on publications in the web of science. J Informat Sci Theo and Pract. 2014; 2(1):35–47.