

## Analysis of Research status based on Citation Context

Byungkyu Kim, Seon-heui Choi, Muyeong Kang

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

Ji-Hoon Kang\*

Computer Science and Engineering  
Chungnam National University, Daejeon, South Korea

### ABSTRACT

*A citation analysis utilizes the relations among citations and is the most popular bibliometric method. This analysis is based on 1) the evaluation by paper, journal and researcher of the research output, 2) the identification of emerging research topics, 3) the production of a map of the intellectual structure of the research domain and 4) various services for academic information. However, this approach has a limitation in that a citation is treated in a very simple manner, even though the purpose of citation can vary greatly. To address this problem, new approaches have been studied that take into account the citation context. This research separates the citations according to the citation functions and tries to conduct an analysis according to the newly classified citations. Furthermore, research on the citation summarization and visualization based on both the citation context and the citation function of the citations was also attempted. However, since there are very few studies related to citation context in South Korea, more research and development is needed in this area. This study analyzes the status of the research in terms of the citation context. For this, we utilized social network analysis methods.*

**Key words:** Citation Context, Citation Function, Citation Summarization.

### 1. INTRODUCTION

Traditional citation analysis based on the number of citations is limited because it does not provide the context of citations. Therefore, scholars have investigated diverse citation analysis methods in order to overcome the limitation of the traditional citation analysis. This research focuses on how citation analysis can be improved by understanding citation contexts. The prior citation analysis cannot show the complicated citation-network. Recent studies have used context citation analysis and provided qualitative analysis of research articles as well as quantitative analysis of them. Therefore, these research findings can suggest important ideas in the areas of research article searches and relevant services.

The traditional citation-based bibliometrics have several limitations in terms of evaluation of research outcomes and its research trend service. In order to overcome the limitations of bibliometrics, recent studies have investigated the phrases in research articles to understand the context of citation, which can provide better evaluation of research

outcomes and research trends. The context or phrase-based citation analysis categorizes the diverse formats of citations into specific types or functions, and this helps us to understand more detailed functions of each citation and reorganize the relationships of research articles based on specified citation categories [3], [32]. This new context-based bibliometrics analysis can also better visualize the citation relationships based on the contexts of citations, and eventually it will improve the citation information service.

There are very few studies related to citation context in South Korea. Therefore, more research and development for this area is needed. This study analyzed the status of research for citation context. For this, we utilized the method of the social network analysis.

### 2. LITERATURE REVIEW

Citation analysis, which utilizes the relations among citations, is the most popular method of bibliometric methods [19]. This analysis is based on 1) the evaluation by paper, journal and researcher of research output, 2) the identification of emerging research topics, 3) the production of the map for the intellectual structure analysis by research domain and 4) various services for academic information[9,10,15,16]. However, this approach has a limitation since a citation is

---

\* Corresponding author, Email: [jhkang@cnu.ac.kr](mailto:jhkang@cnu.ac.kr)

Manuscript received Mar. 02, 2015; revised Mar. 16, 2015;  
accepted Mar. 23, 2015

treated very simply, even though the purposes of citation are various. To complement this problem, new approaches based on citation context have been studied. This research separates citations by citation functions and tries to analyze based on the newly classified citations [11], [18], [28]. Furthermore, research on citation summarization and visualization based on both citation context and citation function of citations has been tried.

Garfield categorized the purposes of citations into 15 categories (e.g., giving credit for related work, providing background reading, disclaiming work or ideas of others) [7]. Kroon reviewed the different categories of citations between 1965 and 2008 [14]. Digital libraries and information services did not consider the diverse purposes of citations, but they assumed that all the citations have only one purpose. Therefore, researchers need to investigate the specific purposes of citations by themselves, and this has left a significant burden to the researchers. Due to the limitations of the traditional bibliometrics, recent studies have focused on automatic categorization of citations based on their purposes and further providing summaries of relevant citations (citation). Those studies investigated automatic detection of specific definition of citation, citation context, citation sentence, and citation area; then, they provided automatic summaries of research articles based on the previously detected and categorized citations (citation).

Teufel suggested specific citation purposes and automatic categorization of citation functions. Prior studies focusing on citation purposes employed natural language processing, text mining, and machine learning in order to detect diverse purposes of citations in research articles. Many scholars have used Murugesan categorization for analyzing citation context, which focuses on conceptual/operational use, evolutionary/juxtapositional group, organic/perfunctory group, and confirmative/negational group [17]. Simone proposed an automatic reference analysis method based on the sentences which include citations and 12 categories of citations based on their purposes [25]. Tuarob conducted a similar research study by investigating the purposes of citations based on Simone's method [25], [26]. Similar to Simone, Tuarob also reported that 61% of citations were difficult to categorize, and most of citation purposes are both "mention" and "argument" [26]. Hu investigated how the location of citations are related to the characteristics of references, and visualized the locations of citations across research articles, which include introduction, method, result, and conclusion [12]. Boyack examined the co-citations in research articles and clustered them to understand how each cited reference is related to the other references cited together [4]. Abu-Jbara and Radev proposed natural language processing methods in order to know the qualitative factors in research articles [1], [22]. Ulrich and Weitz analyzed the context-based citations based on citation sentences and typed citation graph (TCG), which can enable authors to understand the relationships among articles such as research trends and background information easily [21], [27]. Agarwal suggested a framework for the summary of a research article based on the citation sources and their clustering analyses [2]. Qazvinian provided a citation summary method for a research

article based on the relationships among citation sentences [6], [20]. Chen suggested an impact-based summary method, which can also provide relevant information with regards to the research article identified [5]. Sugiyama proposed the supervised learning method, which can detect citation sentences and their contexts and suggest which sentence needs to have appropriate citation(s) [24]. Huang suggested the citation semantic link network (CSLN) based on citation purpose category, sentiment analysis, and keyword extraction, and it can provide users with visualized citation relationship services [13]. The characteristics of CSLN include frequency, multi-dimension, location, and opinion and sentiment. Sendhilkumar proposed a method that can quantify the quality of citations based on the assumption that each research article would have different quality of citations [23].

### 3. RESEARCH METHOD

This research identified a total of 43 research articles focusing on citation-context, and it identified their journal publications, keywords, author(s), their institutions, nations, and research groups and then analyzed their relationships based on the social network analysis method. In particular, this research used KISTI KnowledgeMatrix and Cyram NetMiner in order to visualize its analysis. Prior studies in citation-context were mainly published in the conferences of natural language processing and information retrieval, and they were also appeared in convergence journals in information science, linguistic journals, and bio-informatics journals.

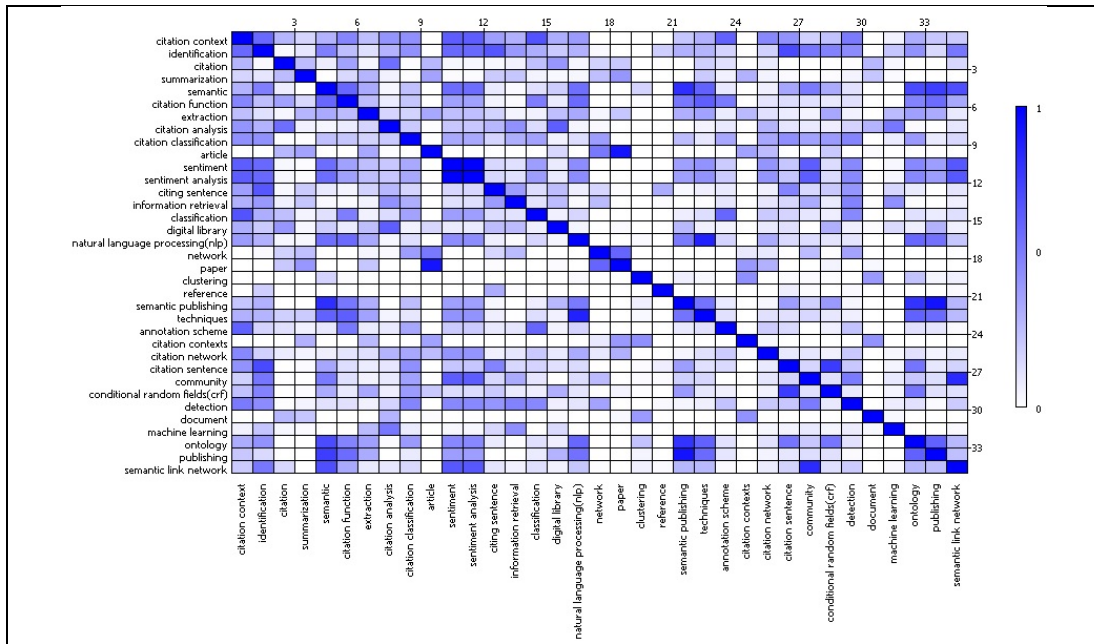
### 4. DATA ANALYSIS OF RESEARCH BASED ON CITATION CONTEXT

#### 4.1 Analysis of author keyword

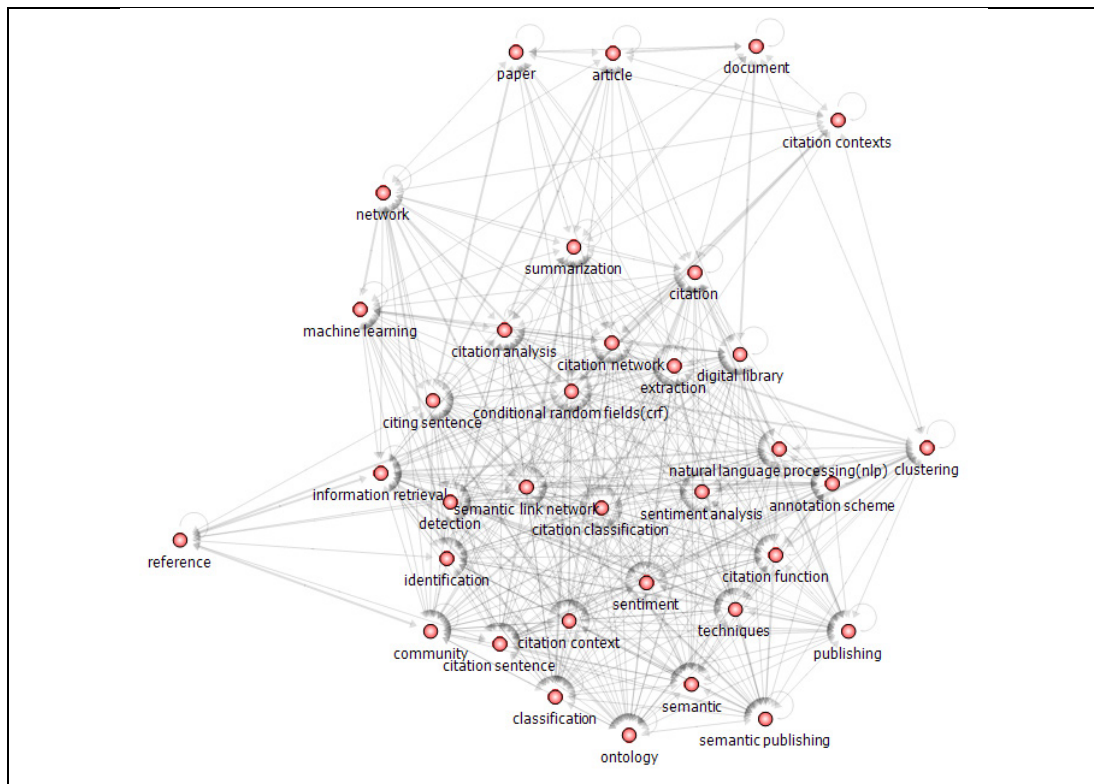
In library and information science, studies on both purpose of citations and utilization have been accomplished. Since the year 2000, advanced research has increased in computer science such as research group for natural language processing and text mining. The studies based on citation context can be separated as 1) the research on citation function as both identification of citation context and classification scheme design for citation function, 2) the research for the summarization of papers utilizing citation contexts 3) the research for performance improvement of information retrieval using the keywords from citation contexts and 4) the research of new service models based on citation context and citation function. We analyzed the articles related to the research for citation context since the year 2000, and specifically selected the article which Simone Teufel published in 2006 as the key paper [25]. This paper suggests the method for the automatic classification of citation function. Over 90 papers cited this key paper in their research. These studies usually utilized the computer-based technology such as natural language processing, text mining and machine learning. Also, as a result of analysis of author information, such as author occupation and research paper's

source, there is a lot of research output in mainly computer science, information science and partly medical science. We produced a proximity matrix of related papers and keywords networks based on this such as below Fig. 1 using author keywords or those manually extracted in the case of a paper without author keywords by ourselves. Netminer of Cyram inc., which is the software for the social network analysis,

was used to produce matrices and the networks for this study. Fig. 1(a) is the matrix produced with the Pearson correlation coefficient and Fig. 1(b) is the network of keywords based on the matrix. Through these figures, we can see how key words such as citation context, citation sentence, citation function, sentiment analysis, citation intention analysis, citation classification, and citation summarization are related.



(a) proximity matrix



(b) keyword network

Fig. 1. Analysis result of author keyword



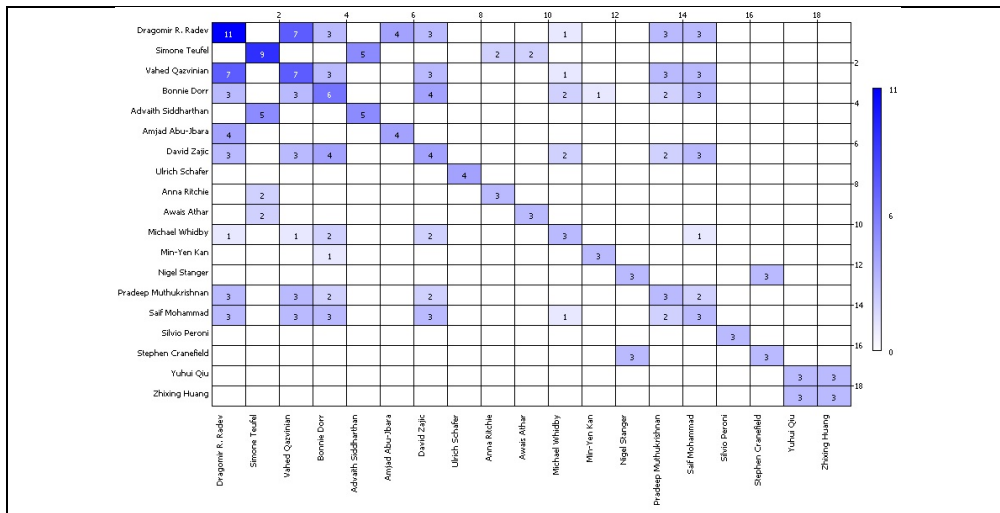
Fig. 2. Tag Cloud based on the Keyword of research relating Citation Context

Fig. 2 below shows the tag cloud based on the keywords extracted from the research articles. It suggests that citation phrase, citation sentence, citation function, sentiment analysis, citation summary, extraction, identification, categorization, network, and information searching are emphasized around the citation.

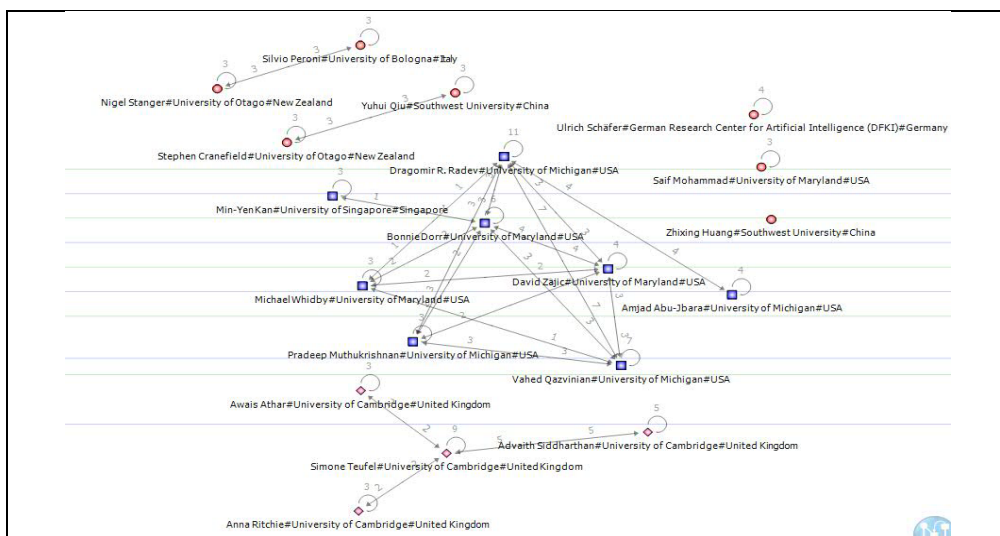
**4.2. Analysis of researchers**

Based on author information such as the author’s name, occupation and co-authorship, we identified the central author and groups.

Fig. 3(a) shows the co-occurrence matrix and Fig. 3(b) is the network of co-authors produced by the matrix.



(a) Co-occurrence matrix



(b) Co-author network

Fig. 3. Analysis result of Co-author

Through the analysis of co-authorship among authors, authors and institutes that play a main role in this research area, were revealed to be University of Michigan (Department of EECS and School of Information: Dragomir R. Radev, Vahed Qazvinian) and University of Maryland (Human Language Technology Center of Excellence : Bonnie Dorr, David Zajic) in USA and University of Cambridge (Natural Language and Information Processing Group: Simone Teufel, Anna Ritchie) in UK. Specifically, the research partnership between University of Michigan and University of Maryland was strong as a result.

## 5. DISCUSSION

The citation function analysis can help us understand the meaning and purpose of citations included in research publications by extracting them with diverse methods, and this result can be utilized to further understand the research articles, evaluate the quality of publications, and enable the systematic search of articles. The citation function analysis can also allow us to analyze the research trend and any possible relationship among different research fields. Therefore, it is very important to understand the exact intention of citations in his/her publication. In order to better understand this, it is necessary to examine the physical and logical locations of citations, as well as analyze the citation sentences and phrases.

## 6. IMPLICATION

This research suggests that the citation context analysis method can help us understand the current trends of research with better contextual information. In addition, the future research can suggest how the citation analysis based on citation context can be applied into research information services. Specifically, the citation summary service can enable scholars to access the citation summary directly without reading each citing research article, and it can expedite the process of research work further.

## 7. CONCLUSION

The traditional quantitative bibliometrics by counting the number of citations have limitations, and the context-based citation analyses methods emerged as important bibliometrics methods that analyze research articles based on citation analysis, citation summary, and qualitative evaluation.

This research analyzed the relevant research articles to identify major scholars, keywords, and their key findings. This research shows that there are a number of research approaches to context-based citation analyses; however, the existing studies in context-based citation analyses are in the beginning stages, and many of them only provide prototypes rather than reliable services. This research also provided the processes of actual applications of context-based citation analyses based on the prior research findings.

We believe that context-based citation analyses services for citation purpose and summary will be available in academic research databases, and it would be possible to base the citation search service on citation relationships in the future. This citation analysis service focusing on citation purposes and summary enables researchers to increase their research productivity by providing effective information searches.

## REFERENCES

- [1] A. J. Amjad, J. Ezra, and D. Radev, "Purpose and Polarity of Citation: Towards NLP-based Bibliometrics," Proc. NAACL-HLT'13, 2013.
- [2] A. Nitin, et al., "Towards multi-document summarization of scientific articles: making interesting comparisons with SciSumm," proc. ACL HLT '11, 2011, p. 8.
- [3] A. Bader, et al., "Document clustering of scientific texts using citation contexts," Information Retrieval, vol. 13, no. 2, 2010, pp. 101-131.
- [4] K. W. Boyack, H. Small, and R. Klavans, "Improving the accuracy of co-citation clustering using full text," J Am Soc Inf Sci Tec., vol. 64, no. 9, Jul. 2013, pp. 1759-1767.
- [5] C. Chong, et al., "Design and implementation for literature search and impact-based summaries," Proc. Intelligent Systems and Knowledge Engineering (ISKE), 2010.
- [6] D. Cody, et al., "Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization," Journal of the American Society for Information Science and Technology, vol. 63, no. 12, 2012, pp. 2351-2369.
- [7] E. Garfield, "Can citation indexing be automated?," Proc. Statistical Association Methods for Mechanized Documentation, vol. 269, Dec. 1965, pp. 189-192.
- [8] F. Lawrence D., Y. Aphinyanaphongs, and C. F. Aliferis, "Computer models for identifying instrumental citations in the biomedical literature," Scientometrics, vol. 1, no. 12, 2013, pp. 1-12.
- [9] G. Wolfgang, A. Schubert, and H. J. Czerwon, "An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis," Scientometrics, vol. 44, no. 3, 1999, pp. 427-439.
- [10] G. N. Antonio J., et al., "Improving SCImago Journal & Country Rank (SJR) subject classification through reference analysis," Scientometrics, vol. 89, no. 3, 2011, pp. 741-758.
- [11] C. Donald O. and G.M. Higgins, "How can we investigate citation behavior? A study of reasons for citing literature in communication," Journal of the American Society for Information Science, vol. 51, no. 7, 2000, pp. 635-645.
- [12] H. Zhigang, C. Chen, and Z. Liu, "Where are citations located in the body of scientific articles? A study of the distributions of citation locations," Journal of Informetrics, vol. 7, no. 4, 2013, pp. 887-896.
- [13] H. Zhixing and Yuhui Qiu. "Construction and aggregation of citation semantic link network," proc.

Semantics, Knowledge and Grid, 2008 SKG'08 Fourth International Conference on IEEE, 2008.

- [14] K. F. William, "Finding Communities in Typed Citation Networks," 2008.
- [15] L. Leydesdorff and I. Rafols, "A global map of science based on the ISI subject categories," *J. Am. Soc. Inf. Sci.*, vol. 60, no. 2, Feb. 2009, pp. 348-362.
- [16] A. E. Mahdi and A. Joorabchi, "A citation-based approach to automatic topical indexing of scientific literature," *Journal of Information Science*, vol. 36, no. 6, Nov. 2010, pp. 798-811.
- [17] M. J. Moravcsik and P. Murugesan, "Some Results on the Function and Quality of Citations," *Social Studies of Science*, vol. 5, no. 1, Feb. 1975, pp. 86-92.
- [18] H. Nanba, N. Kando, and M. Okumura, "Classification of research papers using citation links and citation types: Towards automatic review article generation," *Advances in Classification Research Online*, vol. 11, no. 1, Nov. 2011, pp. 117-134.
- [19] A. Pritchard, "Statistical bibliography or bibliometrics," *Journal of documentation*, vol. 25, 1969, pp. 348-349.
- [20] V. Qazvinian and D. R. Radev, "Scientific paper summarization using citation summary networks," *Proc. The 22nd International Conference on Computational Linguistics - COLING '08*, 2008, pp. 689-696.
- [21] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, "The ACL anthology network corpus," *Language Resources and Evaluation*, vol. 47, no. 4, Jan. 2013, pp. 919-944.
- [22] R. Dragomir and A. Abu-Jbara, "Rediscovering ACL discoveries through the lens of ACL Anthology Network citing sentences," *Proc. the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries, Association for Computational Linguistics*, 2012, pp. 1-12.
- [23] S. Sendhilkumar, E. Elakkiya, and G.S. Mahalakshmi, "Citation Semantic Based Approaches to Identify Article Quality," *Proc. ICCSEA*, 2013, pp. 411-420.
- [24] K. Sugiyama, T. Kumar, M.-Y. Kan, and R. C. Tripathi, "Identifying citing sentences in research papers using supervised learning," *proc. 2010 International Conference on Information Retrieval & Knowledge Management (CAMP)*, 2010, pp. 67-72.
- [25] S. Teufel, A. Siddharthan, and D. Tidhar, "Automatic classification of citation function," *Proc. the 2006 Conference on Empirical Methods in Natural Language Processing - EMNLP '06*, 2006, pp. 103-110.
- [26] T. Suppawong, P. Mitra, and C.L. Giles, "A Classification Scheme for Algorithm Citation Function in Scholarly Works," *Proc the 13th ACM/IEEE-CS joint conference on Digital Libraries*, 2013, pp. 367-368.
- [27] W. Benjamin and U. Schafer, "A Graphical Citation Browser for the ACL Anthology," *Proc. LREC*, 2012, pp. 1718-1722.
- [28] M. D. White and P. Wang, "A Qualitative Study of Citing Behavior: Contributions, Criteria, and Metalevel Documentation Concerns," *LIBR QUART*, vol. 67, no. 2, Apr. 1997, pp. 122-154.



#### **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 Dept. of S&T Information Center, Korea Institute of Science & Technology Information. His main research interests include Korea Citation Database and Citation Analysis.



#### **Seon-Hee Choi**

She received the B.S., M.S in library and information science from Yonsei university, Korea in 1992, 1995 respectively. Since then, he has been with Dept. of S&T Information Center, Korea Institute of Science & Technology Information. Her main research interests include Korea Citation Database and Information Service.



#### **Muyeong Kang**

He is the head of Dept. of Information Infrastructure at Korea Institute of Science & Technology Information. He received the M.S., Ph.D., in computer engineering from Hannam university, Korea. His main research interests include National Academic Content Development and Data Management.



#### **Ji-Hoon Kang**

He received the B.S. in computer science and statistics from Seoul National university in 1979 and the M.S., Ph.D. in Computer Science from KAIST in 1981, 1996 respectively. Since then, he has served as Professor of Dept. in Computer Science & Engineering of Chungnam National University. His main research interests include Situation Awareness, Semantic Web, Ontology, XML, XQuery, Digital Library Systems and Logic Databases.