

Disambiguation of Korean Names in References

Sungwon Kim*

Department of Library & Information Science, Chungnam
National University, Daejeon, Korea
E-mail: sungwonk@cnu.ac.kr

ABSTRACT

One of the characteristics of academic writing is the inclusion of citations and references. As the development of reference styles used for international scholarly communication has mostly been led by Western academic societies, the reference styles developed in Western nations do not reflect the characteristics of Korean names. As a result, it is hard to distinguish Korean authors through citations based on Western reference styles, which in turn decreases the retrieval efficiency of relevant authors and ultimately the efficiency of scholarly communication. This paper intends to analyze author name disambiguation of Korean authors indicated according to Western reference styles. It aims to suggest the necessity for enhancing name disambiguation of Korean authors and revision of reference styles. Its ultimate goal is to increase the efficiency of scholarly communication through the improvement of name disambiguation of Korean authors. For this purpose, this study collected and analyzed name data of Korean researchers and compared name disambiguation of authors by reference style. Based on research results, this study confirmed a necessity for revising reference styles to improve name disambiguation of authors and suggested a necessity for research into the improvement of plans for revision.

Keywords: bibliographic style, citation style, reference style, Korean personal names, author name disambiguation

Open Access

Accepted date: June 12, 2018
Received date: May 2, 2018

*Corresponding Author: Sungwon Kim
Professor
Department of Library & Information Science, Chungnam National
University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea
E-mail: sungwonk@cnu.ac.kr

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

1. INTRODUCTION

Academic writings include citations and references. Citations are a significant part of scientific research in all disciplines (Ardanuy, 2013; Bisman, 2011; Serenko & Dumay, 2015; Small, 2010). Academic researchers can increase credibility of their research by mentioning previous studies that are referred to through citations and reference lists. Similarly, readers can retrieve and access the materials for their research through the citations and references. Academic associations define description protocols for the standardization of citations and references used in scholarly communication so that the readers can access cited materials without confusion (American Medical Association Editors, 2007; American Psychological Association, 2009; Gibaldi, 2003; University of Chicago Press, 2003).

Such description protocols are called reference styles, citation styles, or bibliographic styles. Although the names are different, the major elements that such styles contain are the protocols for the descriptions of in-text citations and references.

Since the research areas of each academic association are diverse, and each area prioritizes certain pieces of information differently, reference styles vary depending on academic area and academic society. It is assumed that there are about 6,000 kinds of reference styles (Thomson Reuters, 2014, 2016). One thing that all reference styles have in common is the selection of the author's name as a primary description element. The author's name is the first descriptive element in both in-text citations and references. Since the author's name is selected as the first descriptive element in citations and references, the author's name in citations and references is used as a keyword and access point. The reference styles for which the primary descriptive element is the author's name are assumed to be based on the old tradition of Western catalogues, which use the author's name as the main entry (Chan, 2007).

Most reference styles applied to international scholarly communication are those defined by Western academic associations. The description protocols for names as the primary descriptive element reflect the characteristics of Western names and indication conventions (ANSI/NISO, 2010; British Standards Institution, 1990). However, they did not reflect characteristics of personal names and indication conventions in East Asian countries, including Korea. Western convention-oriented reference styles have been used for a long time without any serious difficulties. While research results produced by Western researchers have dominated international academic activities, Asian

In-text	<p>Within the computer science field, research specifically focused on Korean names primarily deals with automatic transliteration, with various studies being done on improving transliteration performance (Hong, Kim, Lee, & Rim, 2009; Song & Park, 2011). Kim (2008) suggested that the problems with Anglicizing Korean language, including names, required that Anglicization be considered as a form of translation, thus offering a new basis for understanding Anglicization. Kim (2006) considered the problems with the phonetically based McCune-Reischauer system, commonly used when cataloging Korean materials in American libraries, and proposed several methods for improvement. Kim (2001) analyzed the consistency of Anglicization in the names of Korean researchers, compared the name order and format of the Anglicized names of 263 researchers of English literature with the names of 818 other citizens. Kim & Kim (2012) patternized diverse Romanization formats of Korean personal names and suggested a format for the Romanization of Korean personal names which is considered to be the best.</p>
Reference	<p>Kim, H. S. (2001). Unity and consistency in the Romanization of Korean personal names. <i>Korean Journal of English Language and Linguistics</i>, 1(3), 417-435.</p> <p>Kim, J. W. (2008). Korean orthography of loanwords and Romanization of Korean from the viewpoint of translation. <i>The Journal of Translation Studies</i>, 9(2), 67-93.</p> <p>Kim, S. (2006). Romanization in cataloging of Korean materials. <i>Cataloging & Classification Quarterly</i>, 43(2), 53-76.</p> <p>Kim, S. W., & Kim, J. W. (2012). A research on the format for Romanization of Korean personal name. <i>Journal of Information Management</i>, 43(2), 199-222.</p>

Fig. 1. Examples of indications for Korean authors in Science Citation Index/Social Sciences Citation Index journals.

researchers mainly participated as users utilizing research results, and their contributions as producers were limited.

Today, contributions and publications of scholars in Asia and developing nations have greatly increased due to the globalization of academic fields. The number of articles published by researchers in three East Asian countries, Korea, China, and Japan in Science Citation Index (SCI), Social Sciences Citation Index (SSCI), and Science Citation Index Expanded (SCIE) journals in 2013 was 348,779, which is about 24.88% of all the articles (Thomson Reuters, 2015). The referencing of their research results by Western researchers has also been increasing (OCLC, 2008).

In order to retrieve their research results effectively, a revision of reference styles is required to improve the ambiguity of Asian researchers' names that have different conventions in indication from Western countries.

1.1. Description of Korean Personal Names According to International Reference Styles

For in-text citation according to Western reference styles, authors are indicated by number or only by their surname. Their bibliographic information, including the author's name for the citation, is described in a note and/or in the reference list. According to a specific reference style, an author's name is either indicated in the full format (e.g., Kim, Yunhee) or in the form of 'surname, initial(s) of given names' (e.g., Kim, Y.). The following examples are an indication of in-text citation and references of Korean researchers cited in an article. As shown in Fig. 1, names of Korean researchers are indicated with only their surnames for in-text citation, and 'surname, initial(s) of given names' in references. In the case of in-text citation according to such

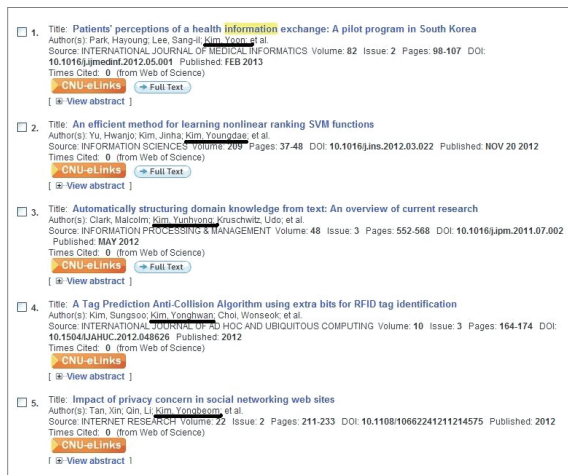


Fig. 2. Examples of retrieval results with the search term 'Kim, Y.'

protocols, five different authors are all indicated as 'Kim' without any distinction, making it impossible to identify each individual. To identify each author, the publication date and number of authors should be referred to in the reference list.

There are associated difficulties in retrieval, due to the characteristics of Korean personal names. Specifically, if a researcher found research results of the Korean researcher 'Kim, Y.' in the reference and tried to retrieve the author's other research results using the search term 'Kim, Y.' in a database, he/she could get the result shown in Fig. 2. Multiple different researchers with the same surname and first initial were retrieved at the same time, as in 'Yoon Kim,' 'Yongdae Kim,' 'Yunhyung Kim,' 'Yonghwan Kim,' 'Yongbeom Kim,' etc. The retriever cannot identify the correct 'Kim, Y.' from the search results, and additional information is required to identify the author 'Kim, Y.' When Korean authors' names are indicated according to the above mentioned Western reference style, it would decrease name disambiguation greatly.

When the author's name is indicated according to reference styles with low name disambiguation, the result is difficulty in identifying the cited author and accessing relevant articles by the author. Another of the bigger potential problems is that it has negative effects on research results performed with the author's name as a basic identifier such as citation analysis, network analysis among authors, or big data analysis. Author name ambiguity greatly distorted the results of co-citation analysis, and it was found that the English indication of Korean and Chinese names is a primary reason for the distortion which decreased disambiguation of indicated names (Kim, Diesner, Kim,

Aleyasen, & Kim, 2014; Song, Kim, & Kim, 2015; Strotmann & Zhao, 2012; Milojevic, 2013). In order to address the lack of author name disambiguation (AND) in reference styles, an algorithm was developed and applied to increase AND (Ferreira, Goncalves, & Laender, 2012; Kim et al., 2014).

2. BACKGROUND OF DIFFICULTIES

As stated above, indication conventions of Korea for personal names are largely different from Western indication conventions, but Western reference styles do not reflect this fact. This study investigated the background of this problem.

There are many differences in personal name indication between Korea and Western countries. The following are the characteristics of personal name indication of Korea: 1) Koreans do not use middle names; 2) Koreans indicate surname first, before the given name; 3) In most cases, Korean names have a typical structure that consists of a one-syllable surname and two-syllable given name; 4) Owing to its typical structure, the three syllables of a Korean name can be written as all attached or spaced; and 5) There are population concentrations by surname (Kim & Cho, 2013).

The fundamental differences which cause low AND are number of surnames and population concentration by surname. First, there are far fewer Korean surnames compared to Western surnames. Hence, there is much lower surname disambiguation when compared with Western surnames. There are about 75,000 surnames in Finland,¹ and more than 1.45 million surnames in the UK (Cheshire, Longley, & Singleton, 2010). Although it varies according to country and language, Western surnames are assumed to number at least in the tens of thousands by country. On the other hand, Korean surnames are reported to be only 286 in total (Statistics Korea, 2010). The 286 surnames are based on indication in Chinese letters, but different Chinese letters share the same Korean letters and pronunciation. As a result, when the surnames are written using the Korean alphabet (Hangeul) and pronunciation, the number is reduced to about 180.

That is, while Western personal names can be individualized to at least tens of thousands, Korean surnames are only about 180. The difference in the number of surnames has a great influence on AND.

The other reason for name indication ambiguity in Western reference style is population concentration by

¹ <http://pasikarppanen.net/names/about.htm>

In-text citation ^o	<p>표영란(2004)은 외국의 정보활용능력 교과과정을 비교 분석하여 우리나라에 적용할 수 있는 정보활용능력 교수-학습모형을 제시한 바 있으며, 이현실(2004)은 델파이 기법을 활용하여 정보활용능력의 증진을 위한 교육항목을 도출하여 제시하였다. 이경연과 경동철(2005)는 정보활용능력의 평가를 위한 통합적 평가모형을 제시한 바 있으며, 이정연과 최은주(2006)는 정보활용능력 교육 효과에 관한 실험적 연구를 수행한 바 있다. 김순희(2006a, 2006b)는 디지털분야의 대학생들에게 필요</p>
Reference ^o	<p>[4] Park, Yongmin et al. 2014. "Two-Level Clustering for Sub-Topic Labeling of Social Media-Data." <i>Journal of KIISE</i>, 41(3): 225-232.^o</p> <p>[5] Seo, Deok Hee. 2011. "A Method for Discourse Analysis." <i>Education Review</i>, 4: 218-239.</p> <p>[6] Son, Heungsuk. 2013. "Exploring the Research Method of Education Policy: Fairdough's Critical Discourse Analysis." <i>Korean Journal of Educational Research</i>, 51(1): 163-189.^o</p> <p>[7] Oh, Dong-Geun, Bae, Yeong-Hwal, and Yeo, Ji-Suk. 2014. <i>Understanding and Application of Korean Dewey Decimal Classification, 6th Edition</i>. Daegu: Taeil.^o</p> <p>[8] Oh, Dong-Geun. 2007. <i>Understanding of DDC</i>. Daegu: Taeil.^o</p> <p>[9] You, Ho-Jong. 2010. "A Study on the Web PR Strategy using Social Media: The Case of Chungcheong Tourism." <i>e-Business Research</i>, 11(5): 97-116.^o</p> <p>[10] Lee, Yu-Sun. 2008. <i>Pragmatism</i>. Salim Knowledg Series, 324. Seoul: Salim Publisher.^o</p>

Fig. 3. Examples of full indications for Korean authors in domestic Korean journals.

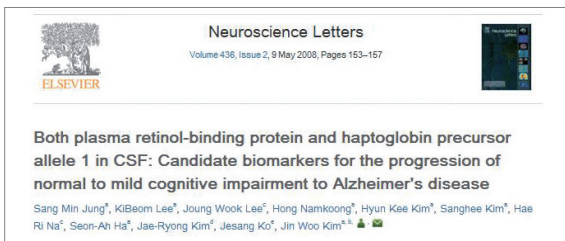


Fig. 4. Indication example of Korean researchers' names in the title page of an international academic journal.

surname. In the USA and the UK, the most common surname is 'Smith.' Americans and Britons whose surname is 'Smith' make up 0.881% and 1.183% of the population of their country, respectively.² Around 1% of the whole population is 'Smith.' In Norway, the most common surname is 'Hansen,' which makes up 0.7% of the population (Aksnes, 2008). Though it is the most common surname, it is only about 1% of the population, and you can identify individuals using their surname without much difficulty. On the other hand, Korea's top 5 surnames take up 54% of the whole population, with Kim at 21.6%, Lee at 14.8%, Park at 8.5%, Choi at 4.7%, and Jeong at 4.3%. More than 90% of the whole Korean population uses 37 surnames (Kim & Cho, 2013). Due to the concentration of surnames in the Korean population, up to 21.6% of researchers in a certain research area have the identical surname Kim stochastically, and it is not easy to identify an individual with his/her surname only.

Such characteristics of Korean personal names, numbers of surnames, population concentration of certain surnames, and Western reference styles that do not reflect such things

lead to low AND and eventually decrease efficiency of scholarly communication.

When those Korean authors are indicated according to Western reference styles, it is hard to retrieve and identify a specific author's research results. Name ambiguity of Korean authors in scholarly communication is very serious.

2.1. Korean Reference Styles

Due to the differences in number of surnames and population distribution by surname between the West and the East, Korean names cannot be identified with surname only. When indicating Korean personal names in English, the names should be fully indicated in order to identify the Korean authors correctly (Kim & Cho, 2013).

Such characteristics of Korean personal names, which require indication of full names, are reflected in Korean reference styles as shown below (Fig. 3). The reference styles used in Korea require that the researcher's full name should be indicated for in-text citations and reference lists in academic journals if the author cites materials produced by a Korean author (Korean Society for Library and Information Science, 2013). Domestic reference styles of Japan and China³ also require full indication of the names of Asian researchers (Kimura, 2013, 2014).

3. DATA COLLECTION FOR THE ANALYSIS OF DISAMBIGUATION

For this study, about 6,300 names of Korean researchers written in English were collected. The data collection was restricted to Korean researchers who published their articles in international academic journals and participated in international scholarly communication. In order to reflect the actual English name indication that they selected by themselves, data were collected directly from the title page of SCI/SSCI journals for this study.

One more thing to consider in collecting data is that English names indicated by the authors themselves should be collected because English name indication style varies depending on researchers. Fig. 4 is the title page of an article written by Korean authors and posted in *Neuroscience Letters*, an SCI journal.

In Fig. 4, authors commonly indicate their names in the order of given name followed by surname but the two syllables of the given name are indicated in diverse ways, including writing them together (e.g., KiBeom), spacing (e.g.,

² <http://www.britishsurnames.co.uk/surnames/SMITH>

³ <http://www.cnki.com.cn/index.htm>

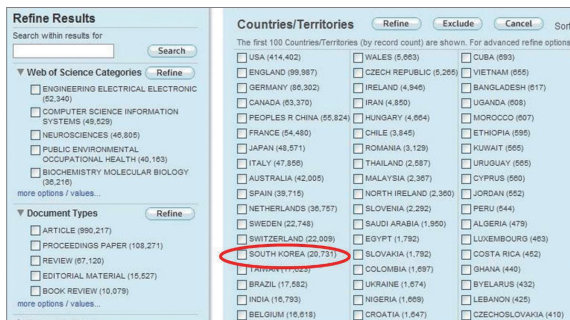


Fig. 5. Captured screen to collect English author names of Korean researchers.

Sang Min), or hyphen (e.g., Seon-Ah). In most SCI/SSCI journals, authors' names on the title page can be indicated in the authors' own style. For this reason, this paper restricted the data collection of the study to the Korean researchers who posted an article to international academic journals and participated in international scholarly communication. In order to analyze English names indicated by authors, English names of Korean researchers posted to SCI/SSCI journals were collected from title pages.

To collect the data, the advanced search function was utilized to restrict research results of Korean researchers as shown in Fig. 5 from the database Web of Science, which indexes SCI/SSCI journals. The data collection period was set to one year and the list of articles posted by Korean researchers to international academic journals during that period was collected.

3.1. Categorization of Reference Styles According to Name Indication Protocols

Academic societies in various research areas have developed reference styles proper to each area (American

Medical Association Editors, 2007; American Psychological Association, 2009; Gibaldi, 2003; University of Chicago Press, 2003). Reference styles used by each research area and association are assumed to be in the thousands. Endnote, which is a bibliographic management tool that is frequently utilized by researchers, supports more than 6,000 reference styles for in-text citation and list of references (Thomson Reuters, 2016).

This study selected several reference styles that are widely used among various reference styles, and reviewed indication styles of personal names and name disambiguation of Korean researchers. Generally, reference style defines two description protocols for in-text citations and for references. However, in-text citation protocol defines indication of serial number or the surname of the author. This study focused on the description protocols of the author's name contained in the reference. In the description of the author's name in reference, there are largely two groups of indication: One is to indicate 'surname, initial(s) of given names' and the other is 'surname, and full description of given name.'

Table 1 shows representative reference styles adopted in Western academic circles. After analyzing reference styles, the indication protocols of the author's name are categorized into four groups. In addition, 'surname, and first initial' was added to the name indication protocols as group 5. In the case of a commercial database, the index is made with 'surname, and first initial of given name' apart from reference styles. Thus, it is assumed that they have different name disambiguation.

Indication protocols of author name were defined by reference style based on indication protocols and categorized into five groups. For citation of Korean authors' research results, indication examples of author names are

Table 1. Protocols of author's name indication and indication examples by reference style

Group	Reference style	Sample authors & description example		Relevant reference style
	Description protocols for author's name	Sample author 1: Hee-sook Shin	Sample author 2: Kim, Jae Kyeong	
1	Surname, initial(s) of given names	Shin, H.	Kim, J. K.	APA 6th Harvard
2	Surname, initial(s) of given names (without period)	Shin H-S	Kim JK	AMA Vancouver
3	Surname, given names (the first name and middle initial)	Shin, Hee-sook	Kim, Jae K.	ASA 3rd
4	Surname, given names (full)	Shin, Hee-sook	Kim, Jae Kyeong	Chicago 16th MLA 7th
5	Surname, first initial of given names	Shin, H.	Kim, J.	Commercial DB

APA, American Psychological Association; AMA, American Medical Association; ASA, American Sociological Association; MLA, Modern Language Association.

also summarized. As shown in Table 1, even though authors' names are indicated according to the same reference style, names of authors may be indicated differently based on the indication style of the author.

3.2. AND Analysis by Reference Style

AND of Korean personal names was compared between five groups by reference style. This study developed a disambiguation index in order to compare disambiguation of Korean personal names according to reference style. Converting the collected data by reference style, data of personal name indication were analyzed and a disambiguation index was calculated by reference style.

3.3. Development of Author Name Disambiguation Index

This study intended to show AND of Korean researcher names in Western reference styles. To show AND more clearly, the author decided to develop a method which can demonstrate this in a quantifying way. Quantifying differences in disambiguation with specific figures allows for differences in disambiguation to be compared more clearly. For the purpose of this study, an AND index was developed to quantify disambiguation. The AND index quantifies the degree of disambiguation of Korean personal names indicated by reference style. It was calculated as follows.

AND index A

$$= \frac{\text{Number of English names indicated by reference and bibliographic style A}}{\text{Number of collected Korean author samples (or number of English indication)}} \times 100$$

3.4. Conversion of English Personal Names

In order to find out how English names of Korean authors are converted according to each reference style, and how much disambiguation the converted Korean author names have, this study collected 6,335 Korean personal names and converted them using the description protocols of each reference style. For this purpose, this study converted the author names indicated on the title page according to description protocols defined in each reference style (e.g. 'Kim, Y.' for 'Yunhee Kim' according to American Psychological Association [APA] style, 'surname, initial(s) of given names'). As most reference styles specify that author name is indicated with serial number or surname for in-text citation, this study was based on the description protocols for reference lists.

There are two reasons for such conversion. First, even

Table 2. Examples of Korean personal names converted by reference style

Author name	Reference style			
	Group 1	Group 2	Group 3	Group 4
Yunhee Kim	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
Yunhee Kim	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
Yunhee Kim	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
Yoonhee Kim	Kim, Y.	Kim, Y	Kim, Yoonhee	Kim, Yoonhee
Yoonhee Kim	Kim, Y.	Kim, Y	Kim, Yoonhee	Kim, Yoonhee
kim yunhee	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
KIM YUN HEE	Kim, Y. H.	Kim, YH	Kim, Yun H.	Kim, Yun Hee
Yun Hee Kim	Kim, Y. H.	Kim, YH	Kim, Yun H.	Kim, Yun Hee
kim yun hee	Kim, Y. H.	Kim, YH	Kim, Yun H.	Kim, Yun Hee
Yun-Hi Kim	Kim, Y.	Kim, Y-H	Kim, Yun-H.	Kim, Yun-Hi
YunHee Kim	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
Younhee kim	Kim, Y.	Kim, Y	Kim, Younhee	Kim, Younhee
Yunhee Kim	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
Yunhee Kim	Kim, Y.	Kim, Y	Kim, Yunhee	Kim, Yunhee
Yoonhee Kim	Kim, Y.	Kim, Y	Kim, Yoonhee	Kim, Yoonhee
Yun Hee Kim	Kim, Y. H.	Kim, YH	Kim, Yun H.	Kim, Yun Hee
kim yun hee	Kim, Y. H.	Kim, YH	Kim, Yun H.	Kim, Yun Hee

if Korean personal names are identical, their English name may be indicated in different forms depending on researchers. For example, even if the same APA style is applied, if the author name on the title page is 'Gildong Hong,' it is indicated as 'Hong, G.,' and if the author name on the title page is 'Gil Dong Hong,' it is indicated as 'Hong, G. D.'

Second, even if Korean authors with the same name indicate their names in the same way in English, they also can be different according to the reference style of an academic journal. For example, an author whose name is indicated as 'Yunhee Kim' on the title page can be indicated as 'Kim, Y.' in APA style, while as 'Kim, Yunhee' in Chicago style. The following table gives examples of English indication of 17 different Korean authors who have the same Korean name when their English names are converted according to Western reference styles. For clear demonstration of the AND index in Table 2, the author uses the researcher database of the National Research Foundation instead of the experiment dataset because the National Research Foundation set contains huge numbers of researcher data and it can provide a better sample.

3.5. Comparison of AND by Reference Style

In the examples in Table 2, 17 Korean authors who have the same Korean name are indicated with 7 different English names. The authors are indicated in five unique

Table 3. Author name disambiguation index by reference style

	By reference style				By indication style		
	Group 1	Group 2	Group 3	Group 4	Surname only	Surname, first-initial (group 5)	Surname, all-initials
Relevant reference style	APA, Harvard	AMA, Vancouver	ASA	MLA, Chicago			
No. of unique authors	2,52	3,681	5,934	6,166	285	1,235	3,076
No. of total authors	6,335	6,335	6,335	6,335	6,335	6,335	6,335
Disambiguation index	39.78	58.1	93.67	97.33	4.5	19.49	48.56

APA, American Psychological Association; AMA, American Medical Association; ASA, American Sociological Association; MLA, Modern Language Association.

formats according to Chicago style (group 4) while they are indicated in two unique formats according to APA style (group 1), which means there is a difference in AND by reference style.

In order to compare differences in AND by reference style, this study analyzed indication of Korean researcher names converted according to Western reference styles, and calculated disambiguation index by reference style. For example, there were 17 Korean authors with the same name in Table 2, who used a total of seven English names. The AND of the English name indication was 41.2 (7/17×100). In the case of indication according to Chicago style, five unique name indications are categorized and the AND index for Chicago style was 29.4 (5/17×100). According to APA style, two unique names are indicated in English and their AND index was 11.8 (2/17×100).

Calculating AND index by reference style through comparison of the number of author indications by reference style with the original author's name data, we can quantify the AND of Korean authors by reference style. Such an AND index by reference style can be utilized as a criterion for evaluating effectiveness and accuracy in Korean name disambiguation of each reference style.

The results of the analysis of Korean name disambiguation by reference style are shown in Table 3. The AND index of group 1 that adopted 'surname, initial(s) of given names' for Korean author names was 39.78. Group 2 uses a reference style with no periods. In group 2 reference lists, when the author's given name is hyphenated, it is indicated as two initials with a hyphen between them (e.g., Kim, Y-H). Its disambiguation index was 58.1. In group 3, the first element (syllable) of the given name was fully described and the second syllable was indicated as an initial. The disambiguation index of group 3 was 93.67, which was greatly improved. Group 4 indicates author name according to the author's indication as it is, and its disambiguation index was 97.33.

In addition to the AND analysis by reference style, disambiguation by other indication styles was also analyzed. The AND index of a Korean author's name when the name was indicated by only surname was 4.5, while in group 5, when it was indicated by surname and first initial of the given name, which is a common style in databases, it was 19.49. When it is indicated with surname and all initials of a two-syllable given name, the AND index was 48.56. AND index with surname only indicated for in-text citation was 4.5. This quantitatively proves that it is almost impossible to identify Korean authors in the list with surname only. The 'surname, first-initial' style (group 5), which is generally adopted in commercial databases, was also found to have a lack of disambiguation.

Indication style of 'surname, all-initials' greatly improved the disambiguation. However, for about 35% of researchers, the two syllables of a Korean given name are written as one combined word, without a space (Kim & Cho, 2013), and the indexer might fail to accurately report the first letter of the second syllable. For example, the two-syllable Korean given name 'Sungeon' can be indicated as 'Sun Geon' or 'Sung Eon'. Depending on how the second syllable is divided, the initial could be indicated 'G' or 'E' for 'Sungeon'. Without additional checking in Korean, the indexer cannot ensure the accuracy of the initial for the second syllable (G or E).

4. CONCLUSION AND IMPLICATIONS

This study analyzed AND of Korean researchers according to the reference styles adopted by major Western scholarly communication channels such as SCI and SSCI. As mentioned in the analysis results, reference styles based on number of Western surnames, population distribution, and conventional practices to indicate personal names cannot fully identify East Asian authors, including Koreans. Therefore, with author names indicated according to

reference style, we cannot distinguish different authors. Difficulties in name disambiguation not only have a negative influence on retrieval, but may distort diverse related research results such as citation analysis, co-citation analysis, network research, and big data analysis that use author names as their basic data.

For these reasons, in the case of implementation of related research that uses authors' names as basic materials for analysis, an algorithm to complement AND is employed or even manual clarifying work is accompanied. Though significant revision can be made through additional complementary methods, the fundamental solution for effective retrieval and various related problems is to enhance AND itself.

The best way to enhance AND of Korean authors, which is different from the population distribution of Western surnames and indication practices, is a full description of the authors' surname and given name. According to the above analysis, the reference style with a full description of the author's name has an AND index for Korean authors of 97.33, meaning almost perfect disambiguation of authors. This proves the effectiveness of the full description of given names.

The AND index of reference styles in group 3 was 93.67, which also showed significant disambiguation, though it was lower than group 4. However, the reference styles in group 3 and 4 do not provide separate protocols for specifically Korean names; a clear and detailed guide is required for description of Korean names.

Reference styles in group 1 and 2 have low name disambiguation of Korean authors and should enhance AND of authors with similar naming conventions. The best solution would be to introduce full description of names of all authors. However, it is unnecessary and a waste of resources to include a full indication of Western authors' names because they have significant name disambiguation with initials using the current description protocols. Thus, the alternative solution is to select countries that require enhancement of name disambiguation and make the names of authors from those countries fully described, while retaining the current name indication for authors in countries where good name disambiguation is maintained as they are. In that case, the list of countries for full description should be clearly presented to prevent confusion among researchers.

Considering the fact that many difficulties in AND and distortion of related research results are caused by difficulties in name disambiguation of Koreans, it is a very urgent and important issue to revise reference styles for the

enhancement of Korean AND in the circumstances where international research activities of East Asian researchers, including Koreans, and their participation in international scholarly communication are increasing. It should be emphasized that such efforts for enhancing AND should be applied to the countries that have a serious population concentration in surnames.

ACKNOWLEDGMENTS

This work was supported by the research fund of Chungnam National University.

REFERENCES

- Aksnes, D. W. (2008). When different persons have an identical author name: How frequent are homonyms? *Journal of the American Society for Information Science and Technology*, 59(5), 838-841.
- American Medical Association Editors. (2007). *American Medical Association manual of style: A guide for authors and editors* (10th ed.). New York: Oxford University Press.
- American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- ANSI/NISO. (2010). *Bibliographic references: ANSI/NISO Z39.29-2005 (R2010)*. Retrieved May 9, 2018 from <https://www.niso.org/publications/ansiniso-z3929-2005-r2010-bibliographic-references>.
- Ardanuy, J. (2013). Sixty years of citation analysis studies in the humanities (1951-2010). *Journal of the American Society for Information Science and Technology*, 64(8), 1751-1755.
- Bisman, J. E. (2011). Cite and seek: Exploring accounting history through citation analysis of the specialist accounting history journals, 1996 to 2008. *Accounting History*, 16(2), 161-183.
- British Standards Institution. (1990). *Recommendations for citing and referencing published material* (2nd ed.). London: British Standards Institution.
- Chan, L. M. (2007). *Cataloging and classification: An introduction* (3rd ed.). Lanham, MD: Scarecrow Press.
- Cheshire, J. A., Longley, P. A., & Singleton, A. D. (2010). The surname regions of Great Britain. *Journal of Maps*, 6(1), 401-409.

- Ferreira, A. A., Goncalves, M. A., & Laender, A. H. F. (2012). A brief survey of automatic methods for author name disambiguation. *ACM SIGMOD Record*, 41(2), 15-26.
- Gibaldi, J. (2003). *MLA handbook for writers of research papers* (5th ed.). New York: Modern Language Association.
- Kim, J., Diesner, J., Kim, H., Aleyasen, A., & Kim, H. M. (2014, October). *Why name ambiguity resolution matters for scholarly big data research*. Paper presented at the 2014 IEEE International Conference on Big Data. Washington, DC.
- Kim, S. & Cho, S. (2013). Characteristics of Korean personal names. *Journal of the American Society for Information Science and Technology*, 64(1), 86-95.
- Kimura, M. (2013). Differences in descriptions of Chinese personal and corporate name authority data: A comparison between China, Japan and South Korea [written in Japanese]. *Library and Information Science*, 69, 19-46.
- Kimura, M. (2014). Differences in representations of Japanese name authority data among CJK countries and the Library of Congress. *Information Processing & Management*, 50(5), 733-751.
- Korean Society for Library and Information Science. (2013). *Publication guides for Journal of the Korean Society for Library and Information Science* [written in Korean]. Retrieved May 9, 2018 from <https://kslis.jams.or.kr/>.
- Milojevic, S. (2013). Accuracy of simple, initials-based methods for author name disambiguation. *Journal of Informetrics*, 7(4), 767-773.
- OCLC. (2008). *Online catalogs: What users and librarians want*. Retrieved May 9, 2018 from <https://www.oclc.org/content/dam/oclc/reports/onlinecatalogs/fullreport.pdf>.
- Serenko, A. & Dumay, J. (2015). Citation classics published in knowledge management journals. Part I: articles and their characteristics. *Journal of Knowledge Management*, 19(2), 401-431.
- Small, H. (2010). Referencing through history: How the analysis of landmark scholarly texts can inform citation theory. *Research Evaluation*, 19(3), 185-193.
- Song, M., Kim, E. H. J., & Kim, H. J. (2015). Exploring author name disambiguation on PubMed-scale. *Journal of Informetrics*, 9(4), 924-941.
- Statistics Korea. (2010). *Results of the 2010 population and housing census*. Retrieved May 9, 2018 from <http://kosis.kr/>.
- Strotmann, A., & Zhao, D. (2012). Author name disambiguation: What difference does it make in author-based citation analysis? *Journal of the American Society for Information Science and Technology*, 63(9), 1820-1833.
- Thomson Reuters. (2014). *EndNote fact sheet*. Retrieved May 9, 2018 from <http://thomsonreuters.com/content/dam/openweb/documents/pdf/scholarly-scientific-research/fact-sheet/endnote-fact-sheet-ssr1303171.pdf>.
- Thomson Reuters. (2015). Welcome to the next generation of InCites. Retrieved May 9, 2018 from <http://about.incites.thomsonreuters.com/>.
- Thomson Reuters. (2016). *EndNote (version 7.5)* [Computer software]. Retrieved May 9, 2018 from <http://endnote.com/>.
- University of Chicago Press. (2003). *The Chicago manual of style* (15th ed.). Chicago: University of Chicago Press.