A Bibliometric Analysis of the Literature on Information Literacy
정보활용능력 주제영역의 계량분석 연구

Myung-Kyu Park*
Heejung Kim**

ABSTRACT

This paper aims to find out the kinds of sub-topics that were researched in relation to Information Literacy (IL). The text mining method was applied to the articles with ‘information literacy’ in the fields of the descriptor, title and abstract in the LISA Database. Also, out of 214 journals that published these articles, those with the top ten highest frequencies were listed and analyzed. Research results show that 908 articles on information literacy were published in 214 journals and ‘User training’ and ‘Students’ were major descriptors in the sub-topic area of information literacy. Also, Reference Services Review and The Journal of Academic Librarianship are two key journals in IL research as they have the highest frequency of related articles and have shown increasing trends.

초 록

본 논문은 정보활용능력과 관련된 연구의 세부 주제영역을 분석해내고자 하는 데에 그 목적이 있다. 이를 위하여 LISA 데이터베이스에서 디스크립터, 표제 그리고 초록에 ‘information literacy’가 포함된 논문들을 대상으로 텍스트마이닝 기법을 적용하였다. 분석 대상이 된 논문들은 총 908건이었고, 이 논문들을 게재한 학술지는 총 214개였는데, 이 중 상위 10개 학술지를 대상으로 한 분석은 수행하였다. 연구결과 ‘information literacy’ 연구영역의 하위 주제 영역의 주요 주제영역은 ‘User training’과 ‘Students’였다. 또한 ‘information literacy’ 주제영역과 가장 밀접한 학술지는 Reference Services Review와 The Journal of Academic Librarianship으로 나타났다.

Keywords: bibliometric analysis, text mining, descriptor analysis, journal analysis, information literacy 계량분석, 텍스트마이닝, 디스크립터분석, 학술지분석, 정보활용능력

* Professor, Graduate School of Education, Yonsei University(parkmk@yonsei.ac.kr)
** Director of Library & Information Services, International Vaccine Institute(heejung@ivi.int)

Received : 16 May 2011 • Revised : 24 May 2011 • Accepted : 13 June 2011
[DOI:10.3743/KOSIM.2011.28.2.053]
1. Introduction

Since Paul Zurkowski (1974) introduced the concept of “information literacy” in 1974, Information Literacy (IL) has been defined and explored by various organizations and researchers to date. As described in the Alexandria Proclamation of 2005, IL is essential for individuals to achieve personal, social, occupational and educational goals (Garner 2005). Therefore IL has been recognized as a core literacy for the twenty-first century and the number of publications related to IL has grown phenomenally over the last thirty years (Garner 2005). While a body of literature on IL has developed, the nature and growth of IL research has been studied by a number of researchers in the field. First, Bruce (2000) reviewed the history of IL research and identified four phases that represent patterns and general developments of the IL research field. She described the first phase (1980s) as precursors, the following phase (1990-95) as the experimental phase, the third phase (1995-99) as the exploratory phase, and the fourth phase (2000- ) as the evolving phase (Bruce 2000).

In addition, selected bibliographies on library instruction and IL have been provided by authors such as H. Rader (2002). She reviewed more than 5000 publications related to library user instruction and IL and then listed selective publications (Rader 2002).

Her study, however, lacked any statistical analysis of the trends. Johnson and others provided a selected bibliography on IL with annotations (Johnson and Jent 2007; Johnson et al. 2009). Major sources well known in the school library media field have also been reviewed and the importance of context has been highlighted to underscore the complexity of IL (Loertscher and Woolls 2002; Lloyd and Williamson 2008).

But an objective analysis of research trends in a completely statistical manner which would reveal certain patterns in this area of study is comparatively meager.

Edwards (1994) analyzed the journal literature on bibliography instruction for the years of 1977 to 1991 to find that 29.9 percent of the articles in her study were research based, and the effectiveness of instruction was the leading topic of research articles.

Crawford and Feldt (2007) examined the literature on instruction in academic libraries. Aharony (2010) reviewed the different publications dealing with information literacy and the emerging trends reflected from 1999 to 2009 in the Web of Science (WOS) database.

However, techniques of bibliometric analysis have rarely been utilized to investigate IL research.

This research examined the literature on IL to find out the research trends in publications using the text mining method. Research questions for this study were as follows:

1. What kinds of sub-topics were researched in relation to IL and which sub-topics were most prevalent?
2. Which journals were leading the research on IL and what kind of trends in publication were displayed over time?
2. Data Collection and Methodologies

The steps for selecting a data collection and research procedure are as follows:
1. Identify the articles with ‘information literacy’ in the fields of the descriptor, title and abstract in the LISA Database.
2. Formulate keywords by stemming and removing stop words from titles, descriptors, abstracts.
3. Apply the log TF X IDF formula to the frequency of each keyword and produce a weight (Salton and McGill 1983).
4. Extract the controlled descriptors assigned to each article in the LISA.
5. Draw a Cosine similarity matrix of descriptors, based on the matrix of the weighted keywords and descriptors.
6. Produce the PFnet results and map, based on the Cosine similarity matrix of descriptors.

The LISA Database provided an ideal database for this study due to its comprehensiveness in the library and information science field. The availability of abstracts and descriptors for most articles was also an advantage. The details of the data are as follows:

The 908 articles that were retrieved were published between 1991 and 2009 even though no year limitations were included in the retrieval formula, and the results for 2009 were incomplete at the time of retrieval. The descriptor ‘information literacy’ alone was used in the search, even though the authors are fully aware that the exclusion of bibliographic instruction and library instruction preclude finding many articles on the same topic.

3. Data Analysis

3.1 Descriptor Analysis

The results from the analysis explained above showed that a total of 3495 keywords and 745 controlled descriptors were extracted from the 908 articles. ‘User training’ was the descriptor with the highest number of occurrences (appearing 310 times) after excluding proper nouns, country names, geographical names, as well as information literacy. The 52 descriptors of that occurred nine or more times from 1991 to 2009 were analyzed for the study and the descriptors and their number of occurrences are as shown in table 2. The descriptors with the highest number of occurrences were ‘User training’ followed by ‘Univer-

<table>
<thead>
<tr>
<th>Retrieved DB</th>
<th>Accessed date</th>
<th>retrieval formula</th>
<th>Retrieved results</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>LISA</td>
<td>Oct 12 2009</td>
<td>DE=(information literacy) and TI=(information literacy) and AB=(information literacy)</td>
<td>908 articles</td>
<td>1991 ~ 2009</td>
</tr>
</tbody>
</table>

(Table 1) Data Collection
It also shows that 22 out of 52 descriptors are directly related to school and higher education and they are marked with an asterisk. Of particular note is that ‘University libraries’, ‘Academic libraries’, ‘College libraries’, and ‘Community college libraries’ appeared a total of 326 times. This shows that IL research has been predominantly in the educational sector, especially the academic library context.

The Nearest Neighbour Cluster program was applied to the results from the PathFinder Network for grouping appropriate clusters and the visual map was drawn using Pajek software (Lee 2006; Kim and Lee 2008).

As shown in Figure 1, 52 descriptors were depicted in five major clusters. First, Cluster A occupied the largest portion in the map, and the networks were formed around two hubs, ‘User training’ and ‘Students’. ‘User training’ was identified as an especially core landmark node here and connected to the other major node. Sub-topics of Students, Curricula, Libraries, Academic libraries, and Higher education in one cluster, and Computer assisted instruction in Cluster E were derived from User training. Curricula was linked to Universities and Schools after Education, Public libraries, School libraries and Education activities were centered around Libraries, with School children especially connected to Public libraries. Academic libraries was linked to Evaluation and Assessment. Higher education was linked to Lifelong learning and Information society concept as well as Standards in Cluster D. The other ten descriptors were linked to User training as a single node.

This map shows that topics on user training have been explored actively and make up the majority of the available material. Cluster D, which included...
Standards and Competencies, and Cluster E, which included Computer assisted instruction based on the Internet and the web were formed as a separate cluster but still connected to User training. In addition, Curricula was researched in relation to schools and universities, assessment and evaluation were explored in academic library settings.

‘Students’ had five single nodes and was linked to University libraries, which had two single nodes and was expanded to Cluster C. Also the Students node was expanded to Cluster B, which included Teachers. That is, there has been plenty of research on Students in University libraries as a basic descriptor which was connected to Faculty, Librarians, Cooperation and Collaboration. Students as Users has been studies with regard to Teachers in Secondary schools, Teaching, and Teacher training.

3.2 Journal Analysis

Journals that publish articles on IL were also analyzed in this study. 908 articles, dated from 1991 to 2009, were identified from a total of 214 journals. The journals with the top ten highest occurrences of articles were chosen for the study. The name of these journals and their frequency by year are as found in Table 3.

Eleven journals were selected that had the ten most frequent occurrences of articles on IL. There were eleven journals because two different journals had articles appear fifteen times. These journals were grouped into three groups. The first group was the journals that had a low number of occurrences before 2000, but increased consistently after. The second group was made up of journals that had a high number of occurrences only during the specific time period.
<table>
<thead>
<tr>
<th>Journal Name</th>
<th>91</th>
<th>92</th>
<th>93</th>
<th>94</th>
<th>95</th>
<th>96</th>
<th>97</th>
<th>98</th>
<th>99</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Services Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>The Journal of Academic Librarianship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39</td>
</tr>
<tr>
<td>Research Strategies</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>College &amp; Research Libraries News</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Library Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>3</td>
<td></td>
<td>2</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>portal: Libraries and the Academy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Journal of Library Administration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>College &amp; Research Libraries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Journal of Librarianship and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Information Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian Academic and Research</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Libraries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Library + Information Update</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>
The third group had irregular occurrences without any pattern. 2009 data was excluded because of its incompleteness. These three groups were depicted as follows.

3.2.1 Journal group with consistent increase in frequency

*Reference Services Review* and the *Journal of Academic Librarianship* had the highest and second highest number of occurrences, respectively, and both journals had a consistent increase in frequency. Six journals out of the eleven highest journals showed the same pattern, including Portal: *Libraries and the Academy*, *College & Research Libraries*, *Journal of Librarianship and Information Science*, and *Library + Information Update*.

3.2.2 Journal group with abrupt high frequency period

Two journals were identified in this group: *Library
Review and Journal of Library Administration. These two journals were not very active in information literacy research in general, but during a specific time period, they published many articles on the topic.

3.2.3 Journal group with irregular frequency

Three journals were identified in this journal group: Research Strategies, College & Research Libraries News, and Australian Academic and Research Libraries.
4. Discussion and Conclusion

In this study, the LISA database was retrieved to identify the research trends in information literacy over time. A total of 908 articles were selected from the retrieval and they were analyzed using the text mining method. An analysis of the Descriptor Profiling Map showed that ‘User training’ and ‘Students’ were major descriptors in the sub-topic area of information literacy. User training was connected to school libraries, public libraries, academic libraries, education and computer assisted instruction.

Research topics on students were related to university libraries, faculty, collaboration, cooperation, and teachers, secondary schools, teacher training.

It was found that 908 articles were published in 214 journals, and eleven journals were listed as the top ten highest ranking journals that contributed to IL research. These journals were grouped into three groups. The first group included six journals that tended to increase steadily. The second group with
two journals had abruptly high frequency during a certain period of time, and the last group had three journals that failed to show any patterns. From the regression analysis, seven journals out of eleven journals had performed research on information literacy increasingly, and four journals decreased from 2000 to 2008. Reference Services Review and The Journal of Academic Librarianship are two key journals in IL research as they have the highest frequency of related articles and have shown increasing trends.

In conclusion, over many years, research on information literacy has covered a wide swath of topics and has become greatly comprehensive. The descriptors found in this study to appear most frequently in articles related to IL research confirms previous suggestions that this area of study has focused mostly on higher education, school libraries, and on the education sector in general. As Bruce (2000) mentioned, interest in workplace-based research on IL began to emerge in 1995-1999 era, but this study confirms that it is still scarce. This shows that there is a need to expand research territory of IL into areas beyond education into information literacy instruction for the diverse population in communities, in the work place, and other contexts.

References


