Mapping research trends
in the field of knowledge management

Mehri Sedighi and Ammar Jalalimanesh
Iranian Research Institute for Information Science & Technology (IranDoc),
Tehran, IRAN
e-mail: sedighi@irandoc.ac.ir; jalalimanesh@irandoc.ac.ir

ABSTRACT

This paper aims to map the research trend in the field of knowledge management (KM) by presenting a systematic and analytical scientometrics approach based on data from the Web of Science (WoS). The method for science mapping includes the following steps: Defining the domain; identifying keywords related to KM field and its subfields; conducting searches and collecting the publication and citation data from WoS; drawing the structure of scientific productions using scientometrics tools; enriching the science maps by adding new attributes; and analyzing the results. In this study, we provide a visualization overview of the wide distribution of KM publications. The analysis of clusters of the historiographical maps, based on Local Citation Score (LCS) and Global Citation Score (GCS), indicated the most frequent thematic trends. The co-word occurrence analysis for mapping KM research topics shows that the structure of fundamental subject areas within the field of KM has changed and expanded dynamically during 2004-2010. This study could be useful for researchers and subject specialists as well as policy makers as they may view and study the history of a discipline by drawing the structure of its scientific productions, in order to do strategically plan and determine the research priorities in the discipline.

Keywords: Knowledge management; Research trend; Science mapping; Domain discovery; Information visualization; Scientometrics.

INTRODUCTION

In bibliometrics and scientometrics research, much attention has been paid to the analysis of networks of documents, keywords, authors, or journals. Mapping and clustering techniques are frequently used to study such networks. The techniques are used to address questions such as what are the main topics or the main research fields within a certain scientific domain and how do these topics or fields relate to each other (Waltman, Jan van Eck and Noyons 2010). Science mapping analyses the networks of links between articles (citations, co-authorship), patents, or other information entities to understand the structure of science (Borner, Chen and Bonyak 2003), and can be used as a tool for science strategy and evaluation. A variety of methods have been used in science mapping, such as journal citation analysis, co-citation analysis, bibliometrics coupling, and co-word analysis. In recent years visualization tools have been improved to make the maps more informative and easier to understand (Besselaar and Heimeriks 2006).
Visualization mapping is used to explore large amounts of data and to derive new insights by identifying trends, or clusters, in the data associated with a field of study (Lee and Chen 2012). The map created through citation analysis provides a series of historical data, which cover the literature year by year (Garfield 1955) and (Small 1993). One of the earliest attempts to pictorially represent scientific development was Garfield’s historiograph (Garfield 1979). This is a diagram of citation patterns depicting the linking of papers forward and backward in time to trace the lineage of ideas over several generations. In a landmark study (Garfield, Sher and Torpie 1964), a historical account of the discovery of the genetic code was correlated with a citation network. Forty years later, his HistCite™ tool automatically generates chronological tables and historiographs of topical paper collections. It assists researchers and librarians in the following areas: identifying core papers on a topic in question; understanding the impact of specific authors, papers, and journals; and making sense of the history of old and new research topics.

Since then, new methods of information retrieval and new techniques for the analysis, visualization and spatial positioning of information studied based on techniques for visualizing the structure of small scientific domains begin to proliferate (Borner, Chen and Bonyak 2003). Ding et al. (2000) used bibliometrics techniques to break down an area of knowledge into its main elements, and represent the areas and sub-areas graphically. Knowledge domain visualization (KDV) detects and visualizes emerging trends and transient patterns in the scientific literature (Chen and Xie 2005). Some research works in knowledge discovery and data mining systems perform analysis of the engineering domain (Mothe and Dousset 2004; Mothe et al. 2006). Lin, Soergel and Marchionini (1991) developed a self-organizing map (SOM) that represents the semantic relationships among documents and can be used as a bibliographic interface for the retrieval of online information. Braam, Moed and Van Raan (1991a) proposed the combined use of co-citation with co-word analysis for the generation of science maps, emphasizing their structure and dynamic aspects. Borner, Chen and Bonyak (2003) concluded that since domain visualizations are typically based on reference key works in a field, they are a good tool to enable the novice to become familiar with a field through easy identification of landmark articles and books, as well as members of the invisible college or specialties.

Co-word analysis is also an important method of information metrology, proposed as early as the late 70s in 20th century by the French bibliometricians. Currently, mature visualization skills of co-word analysis have been applied in many subjects and disciplines, such as nanotechnology (Kostoff et al. 2006), knowledge management (Ponzi 2003; Hou et al. 2006), the international scientific studies (Hou et al. 2006), human genome (Doisneau-Sixou et al. 2003), bioinformatics (Law and Courtial 1988) and medical informatics (Wagner and Leydesdorff 2005).

This paper presents a new approach for schematic visualization applied to the analysis of scientific domains. The scientific domain chosen is knowledge management (KM), and a total of 50,862 KM research articles published from 2001 to 2010, covered in the Web of Science (WoS) database was analysed.

**LITERATURE REVIEW**

KM publications in general focus on knowledge in organizations, knowledge-based, theory of the firm, strategy, and knowledge creation. Even though KM discipline is relatively a new research discipline, it has already boasted a number of scientometrics research with the

Guo and Sheffield (2008) who studied KM theoretical perspectives, research paradigms and research methods revealed that KM research covers the positivist, interpretive and critical pluralist paradigms. Nie, Ma and Nakamori (2009) explored six essential issues regarding KM research field, which include: why the research field is necessary; what enables its birth or triggers actions on it; what it deals with; how to implement it; how to support it; and where it has been applied. Lee and Chen (2007) addressed the topical content in knowledge engineering, semantic web and artificial intelligence related sub-areas. Serenko et al. (2010) conducted citation analysis of individuals, institutions, and countries in KM and intellectual capital fields. The results indicated the publications from several leading authors and foundations are referenced regularly. Dwivedi et al. (2011) found organizational and systems context-based KM research are the most widely published topics. Chen and Lee (2012) built an intellectual structure by examining a total of 10,974 publications in the knowledge management (KM) field from 1995 to 2010. Document co-citation analysis, pathfinder network and strategic diagram techniques were applied to provide a dynamic view of the evolution of knowledge management research trends.

The current study introduces a method to visualize the research trend and derive the intellectual structure of the domain KM based on a combined use of cited references, co-authorship and co-word occurrence. This method is used to provide a visualization representation of the structure map of any scientific field or related subfield obtained from publications appearing in different time periods.

**OBJECTIVES**

The aim of the present work is to map research trends in the field of KM by presenting a systematic and analytical scientometrics approach based on WoS data. Two research questions are posed:

a) What are the international research trends in the field of knowledge management and each of its sub-domains in the last ten years?

b) What are the most important scientific clusters formed in the historiographical map of knowledge management publications indexed in WoS during 2001-2010? What are their subject areas?
METHOD

This research involved the following six steps for scientific mapping of KM literature:

a) Domain discovery: We defined some keywords representing the domain clearly. We used these keywords to find and retrieve articles relevant to KM.

b) Keyword extracting: We extracted the keywords by scanning materials gathered in step (a) and under expert supervision.

c) Data gathering and preprocessing: We used Web of Science as it is a quality-controlled database of scientific articles and has a unifying research tool which enables the user to acquire, analyze and process the information in a timely manner.

d) Drawing basic science map: Different science maps can be drawn from different viewpoints and for different purposes. In this study we drew maps based on co-authorship, co-word occurrences, and citation historiography. Co-word analyses and co-word occurrences were used. When two professional terms expressing a particular research topic appear in the same article these two words have certain intrinsic relationship. And the more the co-occurrences between these two words, the closer their relationship is. According to this “distance”, the important keywords of a subject are classified further to sum up the research focus, structure and paradigm of a discipline by modern statistical techniques, such as factor analysis, cluster analysis, multidimensional scaling analysis or multivariate analysis methods. There are other tools which can be used in drawing the structure of science in each field. Among them, HistCite™ has more capability in drawing the map of science and the structure of a field, like its ability to provide detailed information about authors, journals, cited references, keywords, yearly output and other data. Added to these applications, HistCite™ can draw historiographs based on Local Citation Score (LCS) and Global Citation Score (GCS) to show the important works and history of science in a field or in an organization, so we decided to use HistCite™ for this research. Also, we used the co-authorship network to answer a variety of questions about collaboration patterns in KM field, such as the numbers of papers authors write, how many people they co-author with, and the times cited of the co-authored papers.

e) Enriching science map: As some of the maps were unclear or too complex to analyze, we used some pre-processing to simplify the maps. This step, which is one of the main contributions of this study, is enriching the maps. Adding more attributes such as colours, different shapes or varying in size and thickness are among common techniques to enrich the basic maps. Excluding elements based on specific filters also helps neglect unnecessary data and simplify the maps.

f) Analyzing the results: In the last step we will try to find the answers to our research questions by exploring the maps and analyzing the data.

Figure 1 illustrates the six steps involved.
RESULTS

Domain Discovery
KM has existed as a separate field of scientific research for almost a decade. In the midst of finding and studying fundamental papers, books and other related reference works on KM, as well as obtaining experts’ opinions to partition KM to smaller sub-categories, we found a useful document i.e. a knowledge management encyclopedia, which proposes a comprehensive classification of this domain. This encyclopedia classifies KM into six logical categories. We used five of these six categories as a basis for explaining our domain and extended it in the next step. According to our extracted domain, we can look at the KM from theoretical, procedural, managerial, technological and organizational viewpoints. Therefore we tried to find keywords which represent these categories clearly in the next step.

Keyword Extraction
Figure 2 depicts the final categories and the associated keywords for drawing science maps. Firstly, these keywords were extracted from papers related to each category, experts’ opinions and the knowledge management encyclopedia. Then the glossary with hundreds of terms was processed to find similar phrases and was reduced by cutting very specific words. The final list was given to the experts and we received their feedback about the glossary in a couple of rounds.

Data Gathering and Pre-processing
We need to identify an appropriate database for searching and retrieving documents to draw science maps based on these criteria: comprehensiveness of materials, metadata standards, supportive software for analyzing and mapping, and ease of use. After investigating different data sources and scientific collections and applying the mentioned criteria, we decided to extract data from WoS. We searched WoS database in March 2010 using the extracted keywords in Figure 2.

The final dataset consists of a total of 50,862 documents indexed in WOS during 2001-2010 (Figure 3). As can be seen, the lowest number of records is in 2001 and the most records is in 2009. The documents in 2010 have not been completely covered by WoS during that time. The yearly growth rate for these publications in WOS was 10.9% per year. As we needed to process extracted data wholly, the search results should be integrated in one database. A total of 65,696 author keywords were found in the retrieved articles. Table 1 presents the top 20 keywords based on publication year of articles. This confirms the importance of these words in KM field.
Figure 2: Major and Minor Categories Related to Knowledge Management
Figure 3: The Number of Knowledge Management Documents Indexed in WOS (2001-2010)

Table 1: Top 20 Keywords based on the Publication Year of Articles

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Year 2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data mining</td>
<td>158</td>
<td>189</td>
<td>268</td>
<td>291</td>
<td>271</td>
<td>359</td>
<td>409</td>
<td>443</td>
<td>546</td>
<td>441</td>
<td>3375</td>
</tr>
<tr>
<td>Risk management</td>
<td>87</td>
<td>110</td>
<td>140</td>
<td>142</td>
<td>154</td>
<td>207</td>
<td>212</td>
<td>215</td>
<td>118</td>
<td>323</td>
<td>1708</td>
</tr>
<tr>
<td>Knowledge management</td>
<td>83</td>
<td>102</td>
<td>130</td>
<td>127</td>
<td>158</td>
<td>156</td>
<td>177</td>
<td>200</td>
<td>299</td>
<td>243</td>
<td>1675</td>
</tr>
<tr>
<td>Social capital</td>
<td>56</td>
<td>67</td>
<td>100</td>
<td>97</td>
<td>136</td>
<td>151</td>
<td>193</td>
<td>232</td>
<td>263</td>
<td>241</td>
<td>1536</td>
</tr>
<tr>
<td>Human capital</td>
<td>52</td>
<td>58</td>
<td>80</td>
<td>93</td>
<td>74</td>
<td>105</td>
<td>114</td>
<td>160</td>
<td>162</td>
<td>168</td>
<td>1066</td>
</tr>
<tr>
<td>Knowledge representation</td>
<td>38</td>
<td>44</td>
<td>56</td>
<td>73</td>
<td>59</td>
<td>68</td>
<td>44</td>
<td>61</td>
<td>57</td>
<td>49</td>
<td>549</td>
</tr>
<tr>
<td>Organizational culture</td>
<td>13</td>
<td>18</td>
<td>47</td>
<td>32</td>
<td>37</td>
<td>56</td>
<td>70</td>
<td>79</td>
<td>92</td>
<td>98</td>
<td>542</td>
</tr>
<tr>
<td>Expert system</td>
<td>22</td>
<td>41</td>
<td>64</td>
<td>47</td>
<td>47</td>
<td>54</td>
<td>68</td>
<td>74</td>
<td>63</td>
<td>57</td>
<td>537</td>
</tr>
<tr>
<td>Organizational learning</td>
<td>38</td>
<td>38</td>
<td>42</td>
<td>37</td>
<td>36</td>
<td>50</td>
<td>38</td>
<td>45</td>
<td>63</td>
<td>58</td>
<td>445</td>
</tr>
<tr>
<td>Classification</td>
<td>12</td>
<td>32</td>
<td>35</td>
<td>39</td>
<td>41</td>
<td>60</td>
<td>45</td>
<td>45</td>
<td>52</td>
<td>69</td>
<td>59</td>
</tr>
<tr>
<td>Knowledge discovery</td>
<td>38</td>
<td>36</td>
<td>36</td>
<td>40</td>
<td>45</td>
<td>41</td>
<td>45</td>
<td>52</td>
<td>44</td>
<td>30</td>
<td>407</td>
</tr>
<tr>
<td>Innovation</td>
<td>13</td>
<td>20</td>
<td>18</td>
<td>28</td>
<td>21</td>
<td>36</td>
<td>43</td>
<td>57</td>
<td>76</td>
<td>78</td>
<td>390</td>
</tr>
<tr>
<td>Risk assessment</td>
<td>23</td>
<td>29</td>
<td>44</td>
<td>43</td>
<td>34</td>
<td>48</td>
<td>36</td>
<td>50</td>
<td>15</td>
<td>59</td>
<td>381</td>
</tr>
<tr>
<td>Clustering</td>
<td>8</td>
<td>22</td>
<td>30</td>
<td>37</td>
<td>51</td>
<td>35</td>
<td>51</td>
<td>58</td>
<td>44</td>
<td>44</td>
<td>380</td>
</tr>
<tr>
<td>Machine learning</td>
<td>27</td>
<td>25</td>
<td>41</td>
<td>38</td>
<td>41</td>
<td>51</td>
<td>34</td>
<td>42</td>
<td>36</td>
<td>45</td>
<td>380</td>
</tr>
<tr>
<td>Knowledge</td>
<td>22</td>
<td>23</td>
<td>30</td>
<td>33</td>
<td>23</td>
<td>37</td>
<td>42</td>
<td>53</td>
<td>52</td>
<td>40</td>
<td>355</td>
</tr>
<tr>
<td>Knowledge sharing</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>18</td>
<td>25</td>
<td>26</td>
<td>38</td>
<td>70</td>
<td>78</td>
<td>71</td>
<td>355</td>
</tr>
<tr>
<td>Knowledge acquisition</td>
<td>31</td>
<td>33</td>
<td>34</td>
<td>27</td>
<td>18</td>
<td>43</td>
<td>38</td>
<td>41</td>
<td>39</td>
<td>36</td>
<td>340</td>
</tr>
<tr>
<td>Ontology</td>
<td>9</td>
<td>9</td>
<td>24</td>
<td>24</td>
<td>25</td>
<td>34</td>
<td>39</td>
<td>50</td>
<td>62</td>
<td>55</td>
<td>331</td>
</tr>
<tr>
<td>Education</td>
<td>16</td>
<td>18</td>
<td>21</td>
<td>19</td>
<td>26</td>
<td>30</td>
<td>36</td>
<td>50</td>
<td>60</td>
<td>47</td>
<td>323</td>
</tr>
</tbody>
</table>
**Drawing and Enriching Historiographical Map**

HistCite™ has been used in the drawing of historiographs based on Local Citation Score (LCS) and Global Citation Score (GCS). The 50,862 retrieved documents on KM received a total of 59,153 local citations and 338,377 global citations. We drew the LCS map with only 150 nodes, due to the high number of links, and the need to have a clearly presented graph. Figure 4 shows the main clusters of this structure. There are four clusters and each cluster consists of a number of documents.

![Figure 4: Historiograph of Documents in Knowledge Management Field based on Local Citation Score (LCS) with 150 Top Nodes](image)

The first cluster, which is the largest cluster in this structure, consists of 52 documents during 2001-2006. In this cluster, document number 3529 received the most local citations (289) in comparison with the other documents, followed by document number 987 (187 LCS), 2163 (123 LCS) and 4460 (93 LCS) respectively. The subject area of the first cluster is "the organizational and social aspects of knowledge management". A small second cluster with 5 documents established during 2002-2005 can be seen and document number 11963 under the subject "process of knowledge management" received the most local citations (114). This cluster is associated with several other documents on "tacit knowledge" and "spatial clustering". The third cluster with 22 documents was established during 2001-2006 and is related to the first cluster. The document that received the highest LCS (101) is document number 13,552, which has established many links to other documents in this cluster. The forth cluster with 19 documents was established during 2001-2005. In this cluster, document
number 977 received the most local citations (417 LCS) in comparison with other studied documents. This article, published in *MIS Quarterly* belonged to the "conceptual foundations of knowledge management" and is one of the most effective article in this field and in this time range. The other important documents in this cluster, all published in the *Journal of Management Information Systems* are included: 1722, 1716, and 8789. Table 2 presents the bibliographic data of the documents in the mains clusters with their respective LCS and GCS.

Table 2: Bibliographic Data of the Documents in the Main Clusters

<table>
<thead>
<tr>
<th>Document number</th>
<th>Bibliographic data</th>
<th>LCS</th>
<th>GCS</th>
</tr>
</thead>
</table>

Therefore, in a general classification based on LCS, the fours clusters can be classified by the following subjects respectively: (a) organizational and social aspects of knowledge management; (b) tacit knowledge and spatial clustering; (c) role of social capital in knowledge management; and (d) conceptual foundations of knowledge management.

In the GCS map, the classification of the four clusters by subject are as follows: (a) role of social capital in knowledge management; (b) social-practice perspective of knowledge; (c) conceptual foundations of knowledge management; and (d) tacit knowledge and spatial clustering.
Co-authorship Network
A co-authorship network is used to answer a broad variety of questions about collaboration patterns, such as the numbers of papers authors write, how many people they co-author with, the typical distance between authors through the network, and how collaboration patterns vary between authors and over time. We used network workbench (NWB) software to draw the co-authorship network of KM publications. In this network, the nodes represent authors and the size of the nodes indicates the number of articles each author has written. The lines or edges indicate the co-authorship and the line thickness represents the frequency of co-authorship. Figure 5 depicts the KM co-authorship network, composed of 200 top nodes and 128 edges. There are 125 weakly connected components. This network has two large nuclei and other nuclei contain smaller groups (usually 3 or 4 people). In this network, authors who have the most collaboration with others are highlighted: Chen, YM., Wang, CB., Cheu, HC., Lee, WB., Chen, YJ. and Gottschalk, P. These authors also have the highest number of articles in this network. As can be seen, most of them are also the most cited authors.

![Co-authorship Network of Knowledge Management Field](image)

Co-word Occurrence Maps
Co-word analysis enables the structuring of data as networks of links and nodes, and as distributions of interacting networks (Yang, Wu and Cui 2012). In this study, we applied co-word occurrence analysis for mapping KM research topics. Due to the complexity of co-occurrence word network extraction process, we focused on records that were retrieved using the keyword search “knowledge management”. The extracted data were analysed based on 4 time periods: 2004-2005, 2006-2007, 2008-2009 and 2010. The co-occurrence word networks were extracted on records original keywords and by the aid of Sci2 tools package. The outputs graphs were exported as GraphML XML files. The final representations were drawn using NodeXL package which is an add-in for Microsoft Excel. Figure 6 exhibits four snapshots of the mentioned networks filtered by important KM topics.
Mapping Research Trends in the Field of Knowledge Management

Figure 6: Co-occurrence Word Maps of Knowledge Management during 2004-2010

The size of the spheres depicts frequency of each keyword; edge depicts the connection relationship between two words; and the thickness of the lines indicates the strength of connection. The strongly related words are linked with the thick line and the weakly related words are linked with the thin line, that is, the thicker the line between the two nodes, the closer the relationship is.

From exploring the networks and comparing them with each other, we can see the stability and change in KM research topics. A number of topics are present in all years, whereas some topics have disappeared. New topics emerge as a recombination of existing topics and in interaction with technological developments. For instance, looking at the phrases such as “ontology” chronologically demonstrates the evolution of this concept and its relation with KM. In 2010 map some new topics emerged such as “trust”, “risk assessment”, “knowledge engineering” and “performance”. It is interesting that “strategy” is connected to “KM” via “Performance”. The relation between keywords such as “digital libraries” and “information retrieval” with KM concealed as time went by.

DISCUSSIONS AND CONCLUSIONS

Understanding and evaluating research is essential for research planners, policy makers and researchers. One of the most efficient methods in evidence-based research assessment is the use of scientometrics approach in examining scientific output covered by global citation databases. Considering the distinct status of knowledge management in modern organizations and its effective role in improving the efficiency and effectiveness of organizational processes, drawing a picture of the scientific publications structure produced by researchers in this area is necessary.
In this study, we have provided a visual overview of the wide distribution of KM publications by analyzing KM articles published during 2001–2010 periods covered by WoS. The yearly growth rate for KM publications in WoS was about 10.9%. We introduced a method based on a combined use of cited references, co-authorship and co-word occurrence to visualize the research trend in KM. The method can be applied to any science field to help understand research trends and their evolution. The resulted maps are mostly network representation of elements such as authors, subjects or papers. The resulted co-word occurrence maps give an insightful representation of the research topics within KM field. Based on the analysis of clusters of the historiographical maps, some of the major subject areas in KM field are organizational and social aspects of knowledge management such as “social capital”, “knowledge network”, and “analysis of social network”. A big cluster is allocated to these subjects in both maps based on LCS and GCS. The subject areas of the other clusters are such as “tacit knowledge and spatial clustering”, “conceptual foundations of knowledge management” and “the impact of social capital on knowledge management”. We also applied co-word occurrence analysis for mapping KM research topics, and the analysis shows that the structure of fundamental subject areas within the field of KM has changed and expanded dynamically during 2004-2010.

Using information visualization in different scientific disciplines could be useful for researchers and subject specialists as well as policy makers. The researchers and subject specialists at a glance can see which topics in their discipline have been under research by their peers, and which areas have been less attended to during a specific time period. The results of such studies would assist the policy makers in the allocation of research funding to specific topics and subject fields with more confidence (Osareh and Keshvari 2010).

We acknowledge that this study has a number of limitations. For instance, due to the huge number of extracted records, we could not do any data cleaning (such as removing spelling errors in keywords) or pre-processing (such as word stemming). Doing so would definitely improve the accuracy of information presented in the historiographical map of KM publications.

Another limitation is in using the co-citation method, for newly published papers may not have enough time to garner citations. However, we believe this study could be useful for a wide range of users, notably scientists, researchers and librarians. It can also help early career researchers gain useful and interesting insights into the exciting field of KM. Future studies in this area could be done using other analytical approach and the results be compared with each other.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Dr. Hamid Reza Jamali who has offered invaluable assistance, support and guidance in this research. He is Assistant Professor at the Department of Library and Information Studies, Faculty of Psychology and Education, Kharazmi University (Tehran, Iran); and Head of Information Science Research Centre, at Iranian Research Institute for Scientific Information and Documentation (IranDoc).
REFERENCES


China: WISE LAB, Institute of Science Studies and Management of Science and Technology, Dalian University of Technology.


