American Journal of Mobile Systems, Applications and Services
Vol. 5, No. 1, 2019, pp. 1-5
http://www.aiscience.org/journal/ajmsas
ISSN: 2471-7282 (Print); ISSN: 2471-7290 (Online)

The Process of Research and Development in Management Information Systems Based on Data Mining of Articles

Abdullah Naralan[1], Hasan Asil[2, *]

[1]Department of Business, Faculty of Economics and Administrative Sciences, Recep Tayyip Erdogan University, Rize, Turkey
[2]Depertment of Computer, Branch Azarshahr, Islamic Azad University, Azarshahr, Iran

Abstract

Millions of articles and studies are annually presented and published in different conferences and journals. One of the considered methods is to study the process of research development and evolution in different branches and their interrelationships. Another method is to distinguish research differences during different years. One of the considered methods is to study the process of research development and evolution in different branches and their interrelationships. This research was an attempt to study a collection of articles published in digital libraries and to cluster them. This study examined more than 10,000 of articles published in the field of management systems. Information Systems is the expression used to describe an Automated System, be it manual, which covers people, machines or organized methods to collect, process, transmit and disseminate data representing information for the user or client. The objective of this study was to provide different research clusters during the past several years. This research used the K-means Algorithm for Clustering. After The Analyze, 100 useful Subject was chosen. The Analyze show the process of research in this Papers. The information, Information technology, Information sharing is a 3 useful subject for research in past five years.

Keywords

Management, Data Mining, Information Systems, Monitoring, Process of Research and Development

Received: November 9, 2018 / Accepted: January 19, 2019 / Published online: February 13, 2019

1. Introduction

Nowadays, digital libraries are among the most important sources for carrying out studies and research in the web environment. Millions of articles and studies are annually presented and published in different conferences and journals [1]. The number of these articles is increasing each day. Considering the vast amount of information in these libraries, managing them with traditional tools is almost impossible and modern tools and methods are required for their management [2]. One of the considered methods is to study the process of research development and evolution in different branches and their interrelationships. Another method is to distinguish research differences during different years [4-6].

Data has become one of the important resources of organizations because of information systems. Therefore, methods and techniques are required for efficiently accessing and sharing data and extracting information from data and using them. By creation and expansion of web and digital libraries and the dramatic increase in the volume of

* Corresponding author
E-mail address: h.asil@iauazar.ac.ir (H. Asil)

information, the need for these techniques and methods has become increasingly evident [7].

Currently, many organizations throughout the world have become aware of advantages of data mining for extracting the required knowledge from unstructured data. Text mining is a sub branch of data mining, which has mostly been used for recovering information and classifying texts in recent years [8].

These techniques could be utilized for identification of the required knowledge from texts and could also affect different texts [11, 12].

Classifying and clustering research and results achieved from different sciences are very important. This research was an attempt to study a collection of articles published in digital libraries and to cluster them. This study examined more than 10,000 of articles published in the field of management systems. The objective of this study was to provide different research clusters during the past several years.

A computer information system is a system that a branch of Science composed of people and computers that processes or interprets information [2-5]. The term is also sometimes used in more restricted senses to refer to only the software used to run a computerized database or to refer to only a computer system. Information Systems is the expression used to describe an Automated System, be it manual, which covers people, machines or organized methods to collect, process, transmit and disseminate data representing information for the user or client.

There are two methods for categorizing data mining information: classifying and clustering. In classification, there are predefined classes of concepts and the goal is to design a system for assigning each new document to one of the classes. This is known as data classification. On the other hand, data clustering is creating these categories automatically. In fact, the goal of document clustering is to determine the focus of concepts in the collection of texts. There are no predefined categories here. In this study, clustering was used for categorization.

Clustering is a very powerful technique for discovering groups and natural dependence in a data set and also for understanding the contained structural and thematic models, without having any prior knowledge about data characteristics [11].

As one of the unsupervised machine learning methods, document clustering is widely used in different fields of processing natural languages such as data recovery, automatic multi text summarization, ….

For example, in search engines, data clustering of search engine results has significant impact on improving the accuracy of data recovery [12, 17].

The basis of clustering is to create and categorize similar documents. In fact, clustering is a method that takes a large collection of documents and automatically divides it into several smaller collection of similar documents. Therefore, documents in a cluster are thematically or conceptually similar. There are two general methods for data clustering [13, 18]:

1. Hierarchical clustering:

In the hierarchical method, first each document is considered as a cluster and then the distance between cluster pairs is calculated.

2. Partitional clustering

In the partitional method, documents are somehow divided into several parts and then documents are assigned to these parts based on their distance.

2. Method

This study examined the process of research and development for articles published in digital libraries and utilized the results for categorizing articles.

2.1. Data Collection

Data were collected from articles published in the digital Internet library of Science Direct. In order to create a database, engine related articles were implemented according to the web data transfer protocols.

This engine was responsible for retrieving articles from the considered library. Based on the text engine, more than 10,000 of articles published during the past five years in the field of information systems were downloaded.

The following figure shows the available articles of the case study during the past five years.

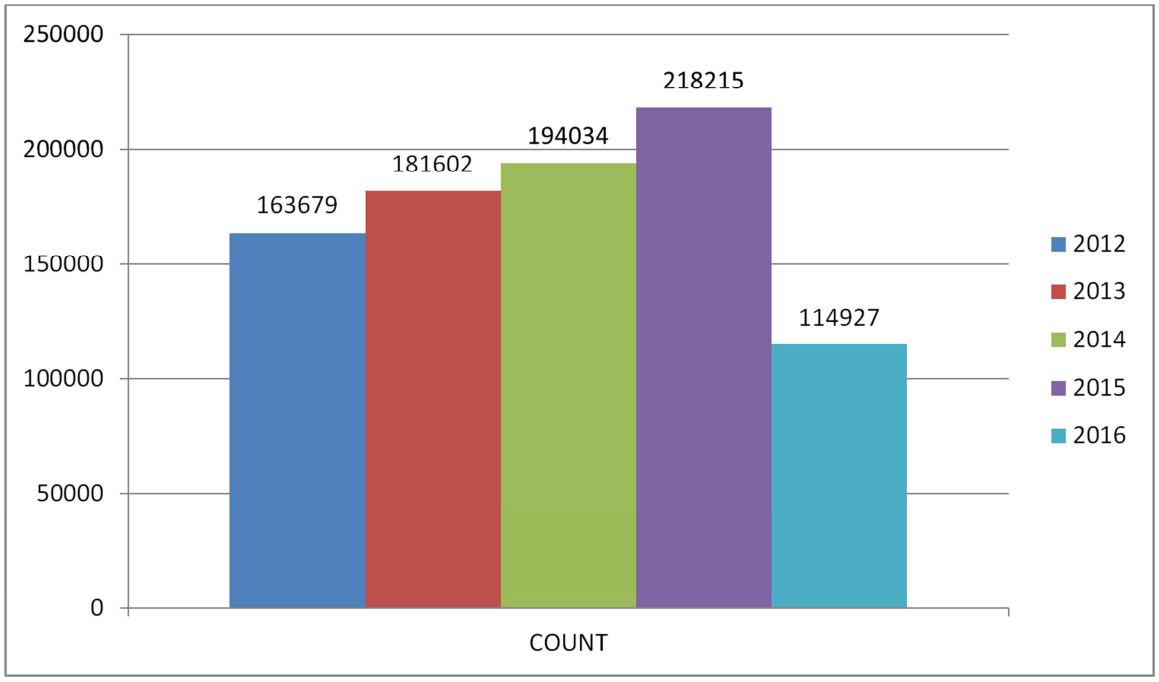

Figure 1. Count of articles in five years.

As it is evident, the article trend has been increasing in this field for the past years.

2.2. Web Mining

Generally, web content include various information such as textual and visual data, audio data, and metadata. Exploring web content is considered an exploration of multimedia data as well. Generally, web data include unstructured data such as free texts or semi-structured data such as HTML pages and somewhat more structured data such as tables or HTML pages created by databases [9]. In order to realize objectives of the study, data retrieved from the library were distinguished according to the web mining method and the redundant data were removed. Information items needed for data analysis were:

1. Article title
2. Abstract summary
3. Keywords

The number of recognized keywords in this datamining was approximately 18260 words.

2.3. Article Clustering

In this database, there are no specified collections for categorizing articles. As mentioned earlier, clustering is one of the unsupervised learning methods. Which means there is no prior knowledge regarding classes in clustering and the responsibility of the system is to distinguish the specifications of these classes by studying data and clusters.

This research used Clustering. The k-means clustering method used in this research. k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

3. Result

The objective of this study was to examine the process of thematic research and development in the field of information systems during the past several years.

After the Clustering Paper, the systems select the Useful keywords. Pig 2 shows Useful keywords.

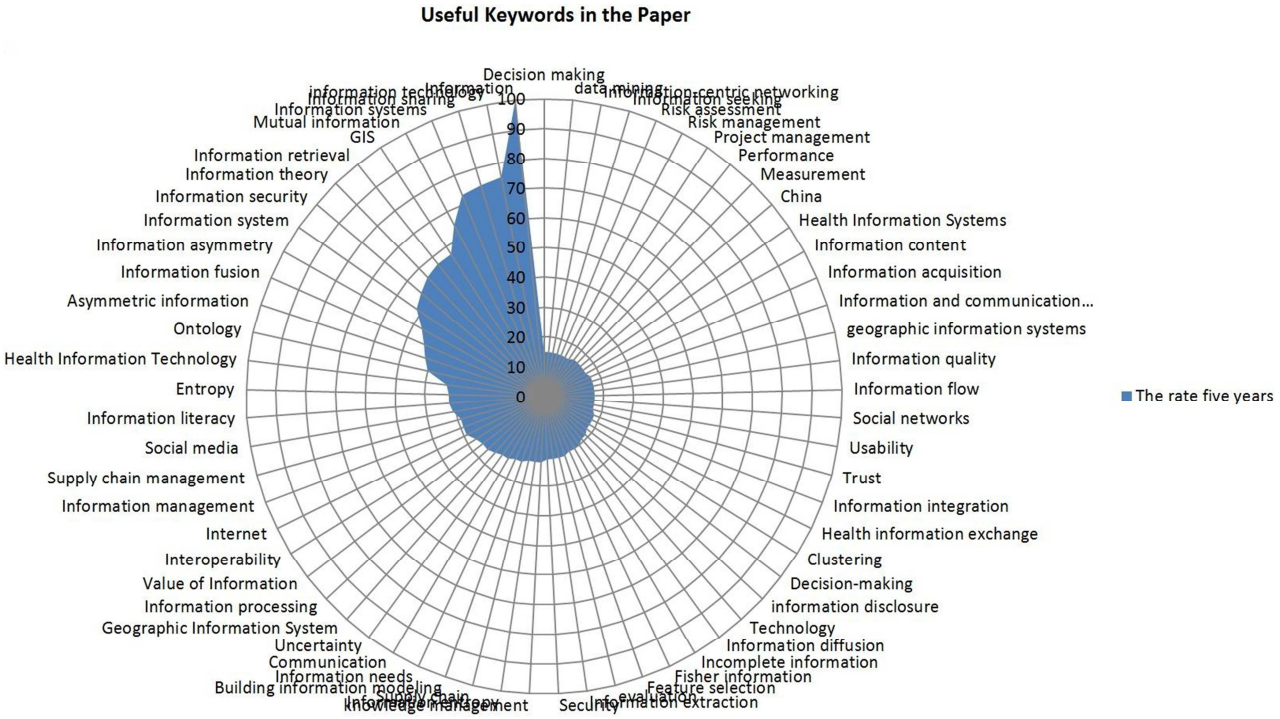

Figure 2. The rate of useful keywords.

The results of the study showed the information keywords are useful keywords. The Figure 2 shows the Distribution of useful used words in the papers.

The count of keywords is the papers are high. For sample, the number of words for 2013 is 199587. This number is calculated after cleaning and removal of redundant words. For the Analyze Papers, select 100 useful keywords. The Figure 3 compares rate use of these subjects in five years.

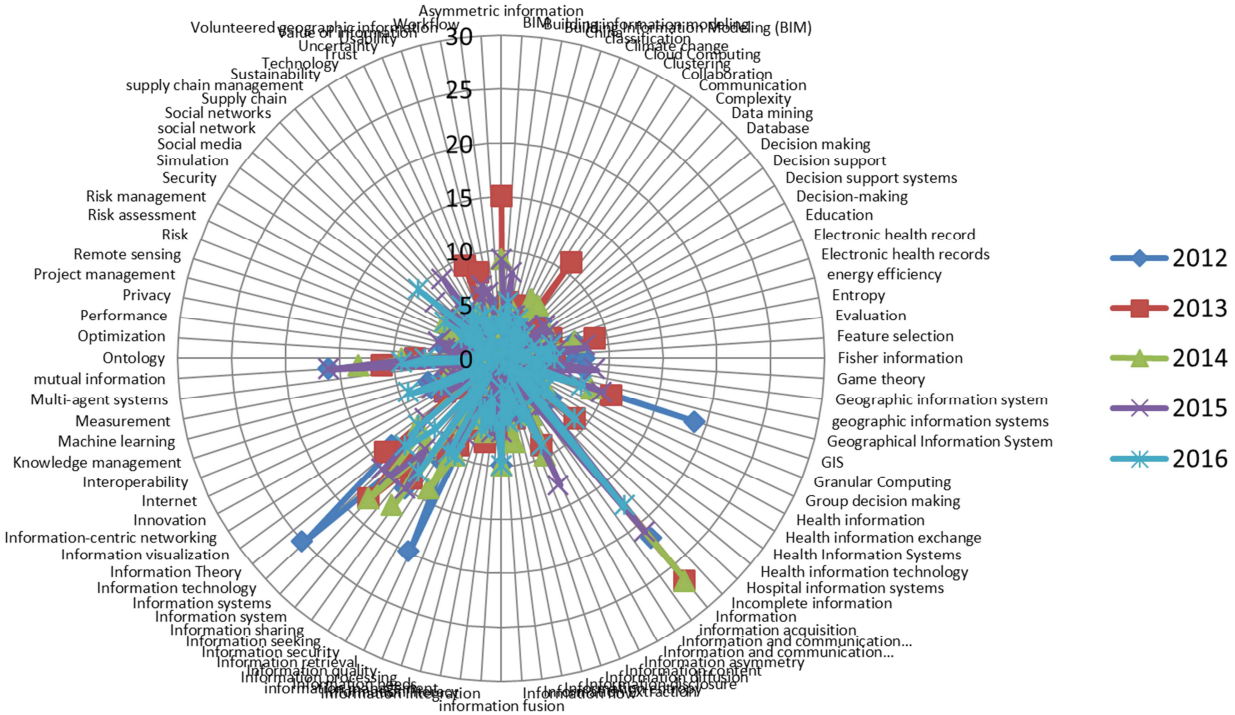

Figure 3. Compares rate use of these subjects in five years.

The information, Information technology, Information sharing is a 3 subject for research

4. Conclusion

Digital libraries are among the most important sources for carrying out studies and research in the web environment. Studying the process of research and development in various disciplines is one of the useful cases that could help researchers in decision making. Nowadays, data mining is one of the most important methods for extracting the required knowledge from unstructured databases. In this research, an attempt was made for categorization of articles published and presented in different journals and conferences during the past several years. The information, Information technology, Information sharing is a 3 useful subject for research in past five years.

References

[1] Miller, W. T., (2005). Data and Text Mining-A Business Applications Approach. Pearson Pentice Hall.

[2] Michael W. Berry, M. C. (2007), Survey of Tect Mining: Clustering, Classification, and Retrieval, Second edition.

[3] Cerrito, B. P., (2006). Introduction to Data Mining using SAS® Enterprise Miner. SAS Publishing.

[4] Battioui, C. (2008). A Text Miner analysis to compare internet and medline information about allergy medications. SAS Regional Conference.

[5] Jacob Kogan, C. N. (2003), Marc Teboulle, Clustering large and highdimensional data.

[6] Sanders, A., & De Vault, C. (2004). Using SAS® at SAS: The Mining of SAS Technical Support. SUGI29.

[7] D'Atri A., De Marco M., Casalino N. (2008). "Interdisciplinary Aspects of Information Systems Studies", Physica-Verlag, Springer, Germany, pp. 1–416, doi:10.1007/978-3-7908-2010-2 ISBN 978-3-7908-2009-6.

[8] Jump up^ Jessup, Leonard M.; Joseph S. Valacich (2008). Information Systems Today (3rd ed.). Pearson Publishing. Page? & Glossary p. 416.

[9] Alter, S (2013). "Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future". Journal of the Association for Information Systems. 14 (2): 72–121.

[10] Jump up^ Alter, S. (2006) The Work System Method: Connecting People, Processes, and IT for Business Results. Works System Press, CA.

[11] Jump up^ Beynon-Davies P. (2009). Business Information Systems. Palgrave, Basingstoke.

[12] Jump up^ Marc S. Silver, M. Lynne Markus, Cynthia Mathis Beath (1995). "The Information Technology Interactive Model: A Foundation for the MBA Core Course". MIS Quarterly: 361–390.

[13] Xu. Rui, W. D. I. (2005), Survey of clustering algorithms. Neural Networks. 2005.

[14] "Introduction to Text Miner." In "SAS Enterprise Miner Help." SAS Enterprise Miner 6.2. SAS Institute Inc., Cary, NC.

[15] Christopher Issal, M. E., Document Clustering. 2010 (Master of Science Thesis).

[16] Eamonn Keogh, S. L. (2005), Chotirat Ann Ratanamahatana, Towards parameter-free datamining.

[17] Hasan Asil, (2013), Innovation Strategies: Challenges or Opportunities in Software Development Teams, Journal of Software Engineering 7 (4).

[18] Hasan Asil, Abdulh Naralan, (2005), Impact of information technology on management in small and medium industries, Journal of Telecommunications System & Management 5 (3), 1-3.