

# Knowledge Mapping for Research Papers

Hafiz Muhammad Faisal<sup>1</sup>, Zahida Shaheen<sup>1</sup>, Atta-ur-Rahman<sup>\*2</sup>,  
Gohar Zaman<sup>3</sup>, Anas Alghamdi<sup>2</sup>, Nawaf Abdulrhman Alowain<sup>2</sup>

<sup>1</sup>University Institute of Information Technology (UIIT), PMAS Arid Agriculture University Rawalpindi, Pakistan

<sup>2</sup>Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia

<sup>3</sup>Faculty of Computer Science and Information Technology, FSKTM, Universiti Tun Hussein Onn Malaysia

## Abstract

One of the most well-known techniques for recognizing knowledge in organization is knowledge mapping. It can help decision makers with better grasping the knowledge flow inside the organization. Mapping organization knowledge particularly, especially in research development institute, has gained much attraction from senior management in recent years. Data mining, among the most fundamental parts of research institutions, has an important part in scientific advances. Due to this important part, various data operations take place with data mining. Data mining operations include database segmentation, predictive modeling and link analysis in data mining information structures. Proposed new strategy for drawing knowledge map, based on data mining information structure logs. Our proposed framework contains five phases including data gathering and developing data warehouse, data preprocessing and transformation, applying knowledge mapping algorithm for extracting input information data for mapping, drawing data map finally analyzing the results. Knowledge map developed utilizing strategies from content mining and essential calculations from complex systems are utilized to recognize the most efficient knowledge. Subsequently, the strategy can be effortlessly executed from an innovation point of view.

### Key words:

*Knowledge Mapping, Knowledge Representation, Information Extraction, Semantic Network, Corpus*

## 1. Introduction

Knowledge is one of the most vital resources of organizations (Aizawa 2003) [2]. Today, due to the significance of knowledge, knowledge management has become a challenging issue for organizations. Knowledge management requires procurement, generation, identification, dissemination and capturing of advantages which give important benefits to the organization. Knowledge mapping and identification is an important phase in knowledge management and is essential for other steps. One of the most known methods for recognizing knowledge in any organization is knowledge mapping.

Knowledge mapping can help decision makers to comprehend the knowledge flow in a better way inside the organization. These maps can be built to show the knowledge sources, sinks, and requirements (Liebowitz, 2005) [19]. There has been a huge amount of research which

concentrates on the significance of knowledge in mapping and society (Brahmi, et al 2013) [7]. A few of these researches show attempts made to structure method selections and others concentrated on characterizing practical and conceptual techniques for drawing knowledge maps (Kebede 2010) [15]. (Kim et al 2003) [16] has classified knowledge maps into five categories including knowledge source, resource, structure, application, and advancement maps. This classification of Eppler knowledge source maps, questions the capacity of organizations for handling the project. Knowledge resource maps visually exhibit the existing load of knowledge on an individual, a group, a unit, or the whole organization. The application maps show which kind of knowledge to be applied to a particular business situation. Lastly, knowledge improvement maps can be utilized to show the important stages to build up a specific capability.

Knowledge map developed by utilizing strategies from content mining and essential calculations from complex systems are used to recognize the most effective knowledge. Moreover, the strategy can be executed effortlessly from an innovation point of view. A knowledge map is generally a visual representation of “knowledge about knowledge” rather than of knowledge itself doing this, like assigning scores, depends on the distance within the sentences. And how can formulate the words. The formulation of two words is called the pairing of words. These formulated words are keywords. Accurate knowledge will show after these extracted keywords. Knowledge map shows the nodes. These nodes are keywords and also show association or links. Total number of extracted keywords and total number of links are quantified knowledge in map. This research will devise a new method of drawing a knowledge map.

In this research proposed new strategy for drawing knowledge map, based on data mining information structure logs. Our proposed framework contains five phases including data gathering and developing data warehouse, data preprocessing and transformation, applying knowledge mapping algorithm for extracting input information data for mapping, drawing data map finally analyzing the results. There are many ways for representing mined knowledge in literature that could be listed as decision tables, decision trees, classification rules, association rules, instance based

and clusters. Hierarchical knowledge map, so-called concept maps provide one model for the hierarchical organization of knowledge: top-level concepts are abstract with few characteristics. A similar type of knowledge map technique using Data mining approach is used. The technique takes a published research paper in PDF format and converts it into an equivalent knowledge map without any dependency of prior knowledge about its corpus etc. Such techniques are encouraged due to their independent nature.

Rest of the paper is organized as follows: section 2 contains the literature review, section 3 contains the details of proposed technique. The proposed idea is implemented in section 5. The results are demonstrated in section 5 while section 6 concludes the paper.

## 2. Literature Review

According to (Kebede, 2010) [15] nowadays organizations are much attracted by knowledge management. It can be due to the reason that organizations now hope to create knowledge with the help of knowledge management process. They hope so to increase the innovation in firms and organizations. Representation of a Knowledge Map presented a knowledge map is a graphical or visual representation of numbers, texts, abstract symbols, models, or stories which is designed purposefully to bridge the communication between map users and makers (Lee et al. 2012)[10]. Probably, organization charts are the simplest types of relationships between the map users and map makers. The idea of map makers behind mapping is to show the users that how their job is related to others in an organization. It is very important that the relationship between the map makers and users should be dynamic and smooth. Knowledge maps must be such that Sumathy et al. (2014) [29] explore that the knowledge mapping is the visualization of knowledge using graphics and symbols, their relationships and positions are defined and depicted with the help edges or arcs. Knowledge maps are successful when they are not spatial but depicting ideas in progression along with relationships and also simplify the representation of reality. They are made up of large facts interconnected and organized together, thus, show the connection between information. The advantage of using knowledge maps is that they are easily usable with other methods like Knowledge Outlines, Analogies, and Frames. According to Liebowitz, (2005) [19] Idea management can be a form of the knowledge management that will permit the system to capture suggestions and ideas of workers which then can be shared online. For example, Imaginatik's Idea Central is an efficient way of gathering ideas and sharing it within the organization. It also helps in better evaluation of ideas by making review process efficient. In [30-40], various approaches have been made for sake of

automatic text categorization. For example, key-phrase extraction, ontological approaches, deep learning and softcomputing etc.

In order to make knowledge flow better among the people so that innovative thinking can be stimulated, knowledge audit should be developed by the organizations. It can be done by mapping the sinks, flows, and sources of knowledge with the help of knowledge map in the organization. To perform a knowledge audit there is a main step of managing knowledge as it helps in determining the flow of knowledge across the organization by (Liebowitz, 2005) [19]. Rcciardi et al.(2005)[25] also stated that the knowledge mapping is an essential work on comprising of review, audit, and synthesis.

Therefore, knowledge map is the vital output given by the audit process of knowledge, as it provides the insight to organizations for improving their processes. Knowledge maps help to determine the points where the flow of knowledge stops or is lost in the organization. According to (Liebowitz 2005) [19], if knowledge maps are well-structured, they help in identifying intellectual capital, socializing newcomers, and help in Machine Learning Approaches

On the contrary, machine learning's approaches address the issues related to dimensions of features by reducing them in targeted documents. These approaches effectively help to apply subsequent learning algorithms by avoiding the problems of over-fitting. Statistical measures used in machine learning are Information Gain, Chi-Squared Statistics, Document Frequency, Expected Cross Entropy, and Odds Ratio by (Aizawa, 2003)[2]. Many comparative studies have been done regarding each field including empirical and theoretical both views. In addition to this, knowledge mapping is also implemented on open learning. Such maps can structure knowledge in multiple ways like learning path planning and problem solving, on online learning and distance education, and learning design. Some people categorized knowledge management into four steps which are knowledge identification maps, knowledge creation, and development, application and assessment maps. However, such classification does not offer a wide-range, accurate, and multipurpose solution for knowledge management by (Balaid et al. 2013) [4]. Moreover, the knowledge map is advantageous for different fields, like information retrieval, information visualization, business process re-engineering, and strategic decision-making support whereas different situations bring out different uses by (Hao et al. 2014) [12].

Matsuo et al., (2004) [20] provided that the computational linguistics has always been attracted by co-occurrence matrix technique. Terms are clustered with respect to their distribution regarding specific syntactic contexts. Tanaka and Iwasaki (1996) translated an uncertain term by the amalgamation of two languages and used  $N*N$  matrices for combining them. Dagan, Lee, and Pereira (1999) explained

a way of probability estimation of word combinations which were unseen previously. Luhn studied co-occurrence for term weightage in 1957 thus; its implementation goes back to the 1950. By calculating term frequency more detailed ways of term occurrence have emerged. Kageura and Umino (1996) briefed five measures of weighting: (1) index term is thought of a word which shows in the document. (2) A repeating word in a document can be an index term. (3) A word that seems just in a predetermined number of archives is probably going to be an index term for these archives. (4) A word that shows up more as often as possible in a document than in the entire database is probably an index term for that record. (5) A word present in the database and depicts a particular distributional characteristic is probably an index term for that database. Here index term means keyword, keyterm, or keyphrase.

### 3. Research Methodology

To find the keywords from documents, the semantic net technique is used to build a knowledge map and to find the frequency of occurrence of a keyword N\*N matrix or co-occurrence matrix is applied. The language used is English.

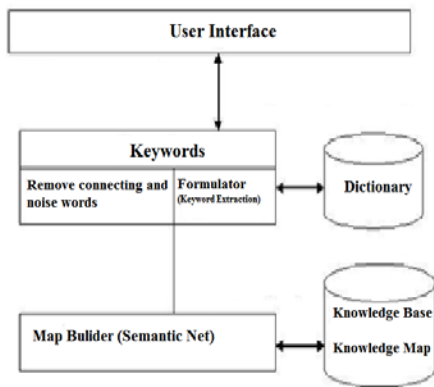


Fig. 1 System Architecture

Figure 1 is representing the system architecture. It gives the overview of the application. With the help of User Interface, the user will select a PDF document; keywords will be selected by the application while removing connecting or noise words. Dictionary of stop words or noise or connecting words is maintained. Then the map builder will transform the fetched keywords using semantic networks into knowledge map.

Figure 2 shows the flow diagram of the proposed system. It shows the complete process of the application that how keywords are extracted and their frequency is calculated. Firstly, the document in PDF format is selected and uploaded in the application. The program then processes the

text in that document and compares it with the database labelled as a dictionary.

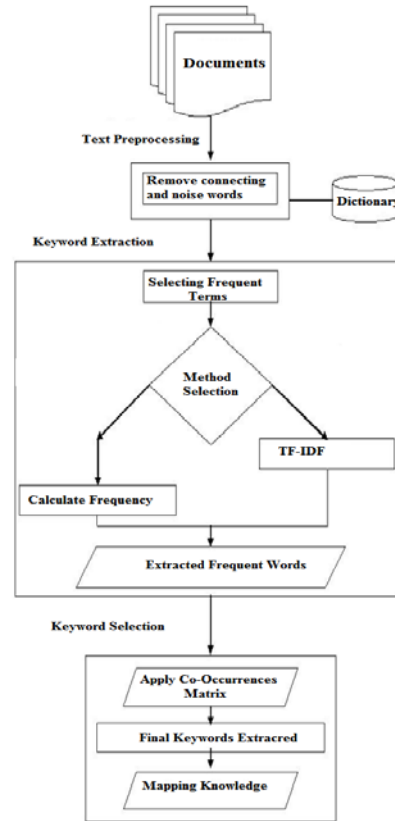


Fig. 2 Flow Diagram

The words stored in the dictionary are noise or stopwords. After comparison, the application is left behind with words occurring frequently. It then chooses a method to calculate the frequency of those words i.e. TF/IDF algorithm. It then extracts words whose frequency has been calculated and arrange them in N\*N or co-occurrence matrix. Co-occurrence matrix single outs the repeating words and map them into knowledge by using semantic networks.

Acquisition of keywords has two categories: keyword extraction and keyword assignment. Extraction of a keyword is done by source document whereas assignment gets keywords from an already defined dictionary or corpus. Extraction of keywords can be extended from words to phrases; however, identification of phrase is a major issue.

### 4. Implementation

Acquisition of keywords has two categories: keyword extraction and keyword assignment. Extraction of a keyword is done by source document whereas assignment gets keywords from an already defined dictionary or corpus.

Extraction of keywords can be extended from words to phrases; however, identification of phrase is a major issue (Huang, 2006) [14].

### A. Keyword Extraction

Keywords can be thought of a concentrated version of text documents and their summaries in concise forms (Menaka, 2013) [23]. Nowadays, retrieval of documents and webpages through keywords is a new trend hence; keyword extraction is an important method of retrieval. Tasks of text mining like summarization and document clustering also use keyword extraction. The ideology behind keyword extraction is to find a relevant keyword in the text (Menaka, 2013) [23]. The first step towards extraction is the selection of the desired document. In the implemented system, Figure 3 shows the document browser. Suppose, a PDF document is selected out four in the list.

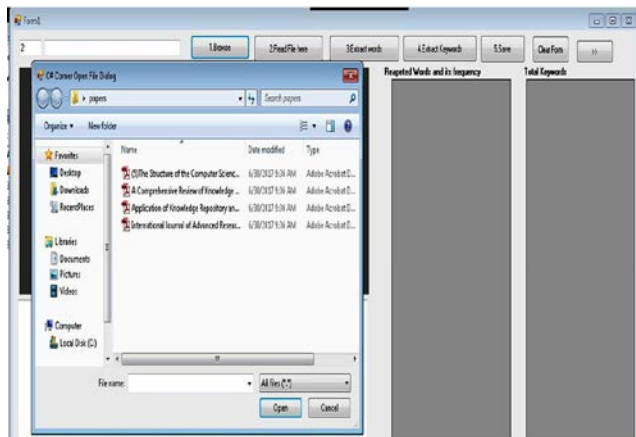


Fig. 3 Document selection

After selection of the paper, its title will be appeared in the dialogue, as shown in Figure 4.

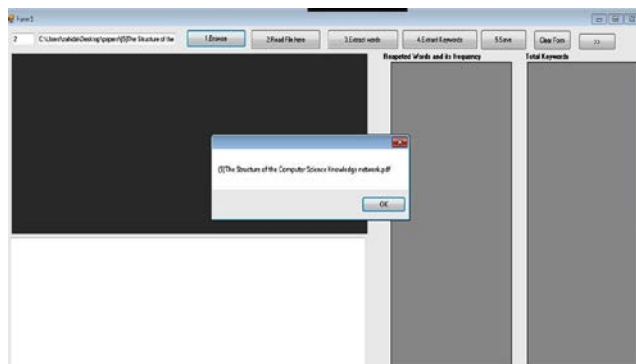


Fig. 4 Displaying paper title

Next dialog given in Figure 5 narrates the number of pages in the document.

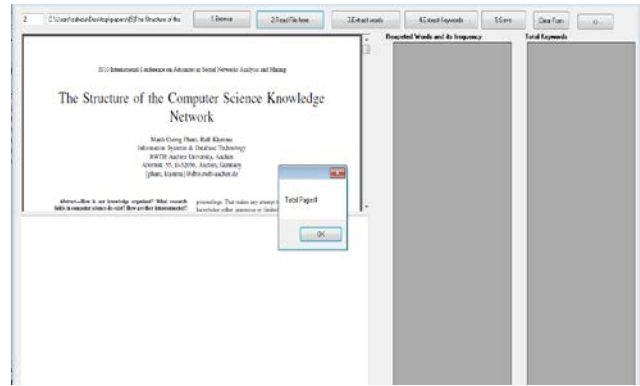


Fig. 5 Displaying total pages of the paper

Next step is to let the program read file by pressing the “Read File here” button. A popup window given in Figure 6 will show a total number of words in the paper.

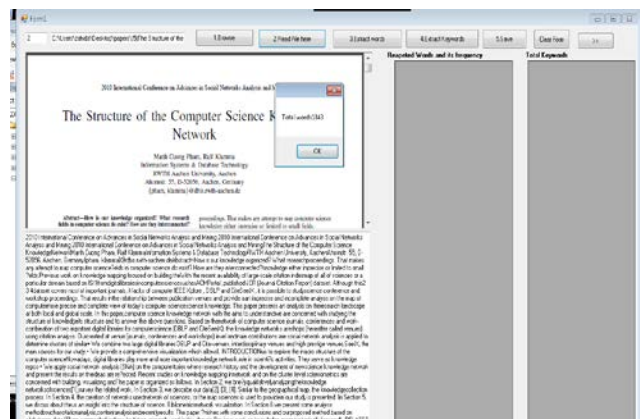


Fig. 6 Displaying Total Words

### B. Stop Words

Words which are not used for retrieving but are part of text or language are called as stop words. There is a consistent need to eliminate these stopwords because they condense the text making it less worthy for analysts. Term space’s dimensionality is reduced when stop words are removed (Menaka, 2013) [23]. These are the words which do not give the understanding of the document for example articles, pronouns, prepositions etc. text mining techniques do not treat such words as keywords. In my project, I made a database of stopwords using XPO6 and Ranks NL databases which are available online. Combining these two databases resulted in one comprehensive database of stopwords.

### C. Frequency-Based Single Document Keyword Extraction

There are many ways of extracting keywords which propose to compare the required document to a corpus in order to



are computerised and algorithms can be applied. Thirdly, they are represented in diagrams by (Hartley, 1997) [13]. Information reserved impacts greatly on the performance of extraction in a specified document, if the information in the given document is rich the performance will be better and if the richness of information in given document is less than performance is compromised. In order to store the information, we opted for Semantic networks as a knowledge organisation system (KOS). We opted for semantic networks because it can hold multiple types of lexical units and their relationships by (Huang, 2006) [14]. It will help us to see the original text may it be a term, a word, or a phrase rather viewing linear or isolated points. A term will be considered or viewed as a node while the relationship between them will be viewed as an edge. Nodes store the information regarding the frequency of a term whereas network's structure holds the information about the dependency by (Huang, 2006) [14]. This stored information will help us in weightage algorithm to find the frequency.

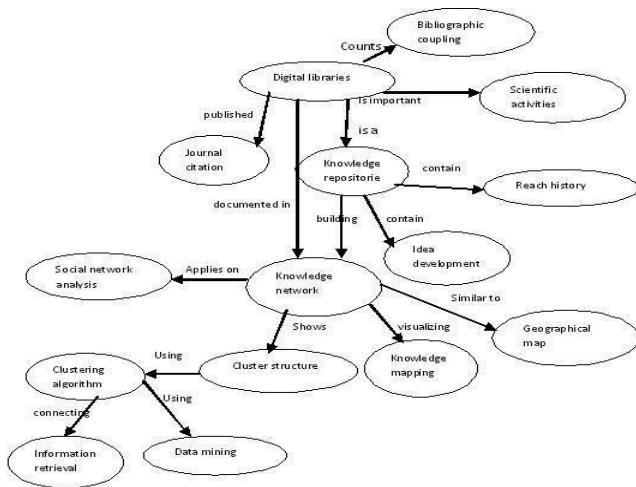


Fig. 8 Semantic Network Model

Figure 8 represents a sample text for which semantic model is built. Nodes are showing the concrete terms and relationship is shown by edges. These edges are also called the associations. The formula that is used to build the knowledge map can be given as:  
 Quantity of knowledge=No. of keywords + No. of links

## 6. Conclusion

In this research work, tried to develop a technique using a combination of existing methods to find keywords from one document at a time. Initially, the format of the document is PDF. The major benefits of this technique are that it is simple in nature and does not require a whole corpus. Its performance is equivalent to TF/IDF algorithm. It's believed this program will be very helpful in retrieving

keywords from the single document while remaining domain independent.

At first, terms occurring frequently are extracted then co-occurrences are counted of single or frequently occurring terms. A term is considered to have a significant meaning when it comes with a specific subset of repeatedly occurring terms. The extent of co-occurrence placement/distribution bias is calculated by x2. This research shows that keyword extraction can be done without the need of a corpus or without using TF/IDF algorithm. In coming sections, I have explained the main working of the N\*N matrix to extract keywords. Details and evaluation of the algorithm are discussed.

## References

- [1] Abilhoa, W. D., and De Castro, L. N. 2014. A keyword extraction method from twitter messages represented as graphs. *Applied Mathematics and Computation*, pp. 308-325.
- [2] Aizawa, A. 2003. An information-theoretic perspective of tf-idf measures. *Information Processing & Management*. pp. 45-65.
- [3] Armstrong, T. 2009. Multiple intelligences in the classroom. *Ascd*.
- [4] Balaid, A. S. S., Zibarzani, M., and Rozan, M. Z. A. 2013. A comprehensive review of knowledge mapping techniques.
- [5] Balaid, A. S. S., Zibarzani, M., and Rozan, M. Z. A. 2013. A comprehensive review of knowledge mapping techniques. *Journal of information systems research and innovation*, pp. 61-66.
- [6] Beliga, S., Mestrovic, A., and Martincic-Ipsic, S. 2015. An overview of graph-based keyword extraction methods and approaches. *Journal of information and organizational sciences*, pp. 1-20.
- [7] Brahmi, M., Atmani, B., and Matta, N. 2013. Dynamic knowledge mapping guided by data mining: application on healthcare. *Journal of Information Processing Systems*, pp.1-30.
- [8] Chan, K., and Liebowitz, J. 2005. The synergy of social network analysis and knowledge mapping: a case study. *International journal of management and decision making*. pp. 19-35.
- [9] Gupta, B., Iyer, L. S., & Aronson, J. E. (2000). *Knowledge management: practices and challenges*. Industrial Management and Data Systems.
- [10] Gupta, P., Mehrotra, D., and Singh, R. (2012) Achieving excellence through knowledge mapping in higher education institution, *IJCA proceedings on international conference on recent advances and future trends in information technology*.
- [11] Gupta, A. and N. T. Deotale. 2014. A Mining Method to Create Knowledge Map by Analyzing the Data Resource, *International Journal of Engineering Trends and Technology (IJETT)*, Vol. 9, pp. 430-435.
- [12] Hao, J., Y. Yan, L. Gong, G. Wang and J. Lin. 2014. Knowledge map-based method for domain knowledge browsing, *Decision Support Systems*, Vol. 61, pp.106-114.
- [13] Hartley, R. T., and Barnden, J. A. 1997. *Semantic networks: visualizations of knowledge*. *Trends in Cognitive Sciences*, pp.169-175.
- [14] Huang, C., Tian, Y., Zhou, Z., Ling, C. X., and Huang, T. 2006. *Keyphrase extraction using semantic networks*

- structure analysis. In *Data Mining, 2006.ICDM'06. Sixth International Conference on IEEE*, pp. 275-284.
- [15] Kebede, G. 2010. Knowledge management: An information science perspective, *International Journal of Information Management*, pp. 416-424.
- [16] Kim, S., E. Suh and H. Hwang. 2003. Building the knowledge map: An industrial case study. *Journal of Knowledge Management*.
- [17] Le-Khac, N. A., Aouad, L. M., and Kechadi, M. T. 2007. Knowledge Map: Toward a new approach supporting the knowledge management in Distributed Data Mining. In *Autonomic and Autonomous Systems, Third International Conference, IEEE*, pp. 67-67 IEEE.
- [18] Lott, B. 2012. Survey of keyword extraction techniques. *UNM Education*, 50.
- [19] Liebowitz, J. 2005. Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. *Journal of knowledge Management*. pp. 76-86.
- [20] Matsuo, Y., and Ishizuka, M. 2004. Keyword extraction from a single document using word co-occurrence statistical information, *International Journal on Artificial Intelligence Tools*, pp. 157-169.
- [21] Matta, N., G.Ducellier, Y.Charlot, M.R.Beldjoudi and F.Tribouillas. 2011. Traceability of Design Project Knowledge using PLM, *IEEE proceedings of International Conference on Cooperation Technologies and sciences, Philadelphia*.
- [22] Martínez-Torres, M. D. R. 2014. Identification of intangible assets in knowledge-based organizations using concept mapping techniques. *R&D Management*, pp. 42-52.
- [23] Menaka, S., and Radha, N. 2013. Text classification using keyword extraction technique. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(12).
- [24] Pham, M. and R.Klamma. 2010. The Structure of the Computer Science Knowledge Network, *International Conference on Advances in Social Networks Analysis and Mining*.
- [25] Rcciardi, R.R., A.C.O.Barroso and J.L.Ermine. 2005. Knowledge Evaluation for Knowledge Management Implementation, 1st international Conference on Nuclear Knowledge Management, INAC.
- [26] Restall, G. 2007. *Symbolic Logic. International Encyclopedia of the Social Sciences*, Macmillan.
- [27] Shen, W., Wang, J., Luo, P., and Wang, M. 2012. Linden: linking named entities with knowledge base via semantic knowledge. In *Proceedings of the 21st international conference on World Wide Web. ACM*, pp. 449-458
- [28] Su, H. N., and Lee, P. C. 2010. Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in *Technology Foresight, Scientometrics*, pp. 65-79.
- [29] Sumathy, K.L. and M.Chidambaram. 2014. Application of Knowledge Repository and Mapping in Knowledge Management, *World Congress on Computing and Communication Technologies*.
- [30] Faisal, H.M., M. Ahmad, S. Asghar, Atta-ur-Rahman, "Intelligent Quranic Story Builder", *International Journal of Hybrid Intelligent Systems*, vol. preprint, pp. 1-8, 2017.
- [31] Shahzadi N., Atta-ur-rahman, Sawar M.J. "Semantic Network based Classifier of Holy Quran". *International Journal of Computer Applications (IJCA) Vol. 39(5): pp. 43-47, February 2012.*
- [32] Shahzadi N., Atta-ur-rahman, Shaheen A., "Semantic Network based Semantic Search of Religious Repository". *International Journal of Computer Applications (IJCA) Vol. 36 (9), pp. 1-5, December 2011.*
- [33] Atta-ur-Rahman, "Knowledge Representation: A Semantic Network Approach", *Handbook of Research on Computational Intelligence Applications in Bioinformatics*, Edition: 1st, Chapter: 4, Publisher: IGI Global, 2016.
- [34] J. Alhiyafi, Atta-ur-Rahman, Fahd Alhaidari, A. Alghamdi, "Automatic Text Categorization using Fuzzy Semantic Network", *SEAHF'19*, 2019.
- [35] Atta-ur-Rahman, F.A. Alhaidari, "The Digital Library and the Archiving System for Educational Institutes", *Pakistan Journal of Information Management and Libraries (PJIM&L)*, vol. 20 (1), pp. 94-117, 2019.
- [36] G. Zaman, H. Mahdin, K. Hussain, Atta-ur-Rahman, "Information Extraction from Semi and Unstructured Data Sources: A Systematic Literature Review", *ICIC Express Letter*, 2019.
- [37] D. Musleh, R. Ahmad, Atta-ur-Rahman, F. Alhaidari, "A Novel Approach to Arabic Keyphrase Extraction", *ICIC Express Letters* 10(10):875-884, 2019.
- [38] Atta-ur-Rahman, F.A. Alhaidari, "Querying RDF Data", *Journal of Theoretical and Applied Information Technology* 26(22):7599-7614, 2018.
- [39] M. Ahmad, U. Farooq, Atta-ur-Rahman, A. Alqatari, S. Dash & A.K. Luhach, "Investigating TYPE constraint for frequent pattern mining", *Journal of Discrete Mathematical Sciences and Cryptography*, 22:4, 605-626, 2019.
- [40] Atta-ur-Rahman, Fahd Abdulsalam Alhaidari, "An Electronic Data Interchange Framework for Educational Institutes", *ICIC Express Letters*, vol.13, No. 9, September 2019.