Research Journal of Applied Sciences, Engineering and Technology 14(9): 334-340, 2017
DOI:10.19026/rjaset.14.5073
ISSN: 2040-7459; e-ISSN: 2040-7467
© 2017 Maxwell Scientific Publication Corp.

Submitted: May 4, 2017 Accepted: July 6, 2017 Published: September 15, 2017

Research Article

Graph-Based Text Representation: A Survey of Current Approaches

[1]Geehan Sabah Hassan, [1]Asma Khazaal Abdulsahib and [2]Siti Sakira Kamaruddin
[1]College of Education for Human Science-Ibn Rushd, University of Baghdad, Baghdad, Iraq
[2]School of Computing, Universiti Utara Malaysia, 06010 UUM Sintok, Malaysia

Abstract: Lately, we have seen the problem of sparsity data has increased due to the increase in the amount of available documentation on the Internet, to take care of this issue need to choose the best strategy for the representation of the content. In recent years, scientists have been switched to the representation of the content graphically. Because the results of previous studies proved that the represented data as graphs reduce the problem of sparse data. So this study aims to review the sorts of graphs used to represent the content of documents. Were the exploratory outcomes recommended that our methodologies are superior to other methodologies in each of the synthetic global data sets and the real.

Keywords: Concept Frame Graph (CFG), Conceptual Graphs Model (CGM), Dependency Graph (DG), Formal Concept Analysis (FCA), sparsity problem, text representation schemes

INTRODUCTION

The way in which the text is represented has a significant impact on the accuracy of the results of clustering and classification. This is according to the previous studies conducted by the researchers where the researchers used the graphs to represent the documents and data and compare them with studies that represented the text using other methods such as (BOW, TF-IDF and N-GRAM) Document representation using the graphs gave high results in accuracy and thus reduced the data scattering ratio (Abdulsahib and Kamaruddin, 2015), So the key component for improving the performance of clustering or classification methods is choosing an appropriate representation of the contents of documents Hassan *et al*. (2015) and consequently will strengthen the clustering outcomes, in Bloehdorn *et al*. (2005) become clear when utilization of Bag Of Words (BOW) influences the clustering outcomes, where In many instances the outcomes are unsuitable as it ignores the relation between essential terms that don't coinciding are literal. So also, another conclusion supports this subject as Harish *et al*. (2010) shows that the (BOW) method It has its disadvantages, which incorporates high dimensionality of the representation, the loss of a semantic relationship that exist between the terms in the content of documents and loss of relation with neighboring words. Wang *et al*. (2011) explained that the graph model is more effective as it has the advantage of capturing the relation between words in the text, therefore the Researchers used the graph rather than other ways to represent the text such as BOW, TF-IDF.

One of the studies that dealt with this topic is Rajani *et al*. (2015). Where one of the problems of the representation of texts using the graphs is that they are expensive computationally so the researchers in this study focused on the development and improvement of the graphs and work of the algorithm uses k-d trees and ball trees to implement nearest neighbor graph construction through a focus on exploiting these underlying tree-based data structures to optimize parallel execution of the nearest neighbor algorithm. In Holder (2009) two approaches were compared. The first approach was based on logic and the second approach was based on graphs. Any representation of data using graphs the study showed many limitations in the logic-based approach, but these limitations can be overcome using a graph-based approach.

There are many sorts of graphs to represent the content incorporates a formal investigation of ideas, graphs identified with the theoretical casing and graphs based on thoughts. For activity graphs, the methods rely on graphs are reliant on the division of the graph. In this situation, clusters are determined by the cutting edges of the graph in a manner that is minimizing the total weights of the edges (cutting edge). It represents every document as a node. In this situation, when representing the text in the form of a graph is represented every document in the form of a graph, where the words represents as anodes either the relationship between the

Corresponding Author: Geehan Sabah Hassan, College of Education for Human Science-Ibn Rushd, University of Baghdad, Baghdad, Iraq

words is represented by the edges. In different clusters if similarity happens between any documents, this leads to the appearance of the edge between the two nodes. In the event of interconnection between documents be significant in one cluster, have the weight of the edges in this cluster more than the weight of the edges of the other clusters. Each algorithm is based on graph give different results, as algorithms partitioning may use graph differently every time.

Different reviews have demonstrated that the representation of texts utilizing graph produce a superior outcome when clustering the content of documents utilizing any method of clustering algorithms for instance k-means, DBSCAN. etc.

LITERATURE REVIEW

Text representation schemes: BOW model ("bag of word") and n-grams Formerly, are used as ways or methods for clustering texts without any consideration of the relationship between words in the text. Attempted the Previous research to represent the content of documents utilizing graphical schemes, for example, "Formal Concept Analysis (FCA), Dependency Graph (DG), Conceptual Graphs (CGs) and Concept Frame Graph (CFG)". The factor is very important and decisive for the outcome of the clustering, is to choose a suitable model for the representation of texts models mentioned above. In general, in this study texts are represented in a single document using the type of graphs that will detail each type later, where words are considered as the node for graph either the lines between the nodes or (Edges) it represents a correlation between words.

Concept Frame Graph (CFG): The analysts utilize a few sorts of graphs in the text representation. Some scientists have made a proposition a learning method for building a CFG information base from the content of documents. Mining text Additionally, addressing the issue of content with the concept is based through conceptual knowledge and knowledge discovery by the base frame construction. Determine a novel method called the Concept Frame Graph (CFG), An intuitive idea outline system is introduced in a client guided information disclosure from the learning base. The researchers find during empirical studies on the genuine documents and the realities of life that the suggested approach is promising for mining further information (Rajaraman and Tan, 2002). The examination to assess the execution of mining with and without graph based text representation was done in Rajaraman and Tan (2003). Algorithms in which the use of other methods of graph is being accuracy ratio, which is not good compared to the method that utilized CFG; where the change in Recall and Precision are 18% and 35% respectively.

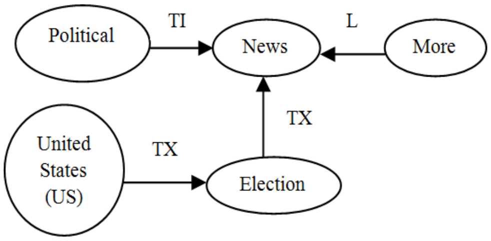

Fig. 1: An example of concept frame graph

This outcome demonstrates that enhanced mining execution can accomplish by representing the contents of documents as a graph.

To determine the terms in the document must first the application of the stems algorithm, lemmas *etc*.,. Using the stem or other techniques to normalize language-specific algorithm, At that point, every term shown in a document becomes a node in the graph. Every node in the graph is distinctive and unique as each node has its own term, even in the case of repetition of the same term in one document. The second task, if a word B an instantly goes after a word A some place in an "area" (content substance, title, or connection and so on.) S of the document, then there is a coordinated edge from the node relating to term A to the node comparing to term B with an edge mark B. An edge is not made among two words on the off chance were isolated by particular punctuation (Quynh and Napoli, 2012).

With the current representation, the graph can catch basic data of content (the place and the relative place of the word). There are three segments specifically to represent the standard, including title, link and text. Title contains the content identified with the archives title and any gave keywords (metadata). The anker text that shows up in hyperlinks on the document called Link. Text involves content of the document (this incorporates hyperlinked content, however, not the content in the document title and keywords). Example of a CFG for a short English Web document having the title "POLITICAL NEWS", a link that reads "MORE NEWS" and text containing "UNITED STATES ELECTION NEWS", is Fig. 1 illustrates, where "TL" signifies the title area, "L" demonstrates a hyperlink and "TX" remains in the noticeable content. There are five words happened in the document: “POLITICAL”, “NEWS”, “MORE”,“ UNITED STATES”, “ELECTION”, which relate to five nodes in the graph. Four edges in the graph demonstrate the relations between words in the document: For example, there is an edge from "POLITICAL" to "NEWS" named by "TI" implying that "POLITICAL" quickly goes before "NEWS" in the title area.

Formal Concept Analysis (FCA): Some specialists utilize another sort of graphs in the text representation, in particular, "Formal Concept Analysis" (FCA) founded by "Rudolf Wille in the mid 80s" (Wille, 1982) toward the start of the eighties of this century, have been developed over the recent ten years in the application of the international community in an extensive variety of disciplines, for example, in Psychology, AI, data recovery, Linguistics and Software Engineering. FCA is the basic method used in the concept of hierarchy or formal ontology from an arrangement of items and properties. In the hierarchy, every concept is represented as a collection of items that share similar qualities for a particular group of properties. The sub-concept in the hierarchy incorporates a subset of the articles in the ideas above it. The technique was gotten from applying lattice and order hypothesis created by Garrett Birkhoff In the 1930s. This scientific strategy includes analysis information to clarify to determine conceptual structures in the set of data. In FCA the similarity measure depends on "Tversky's model" and unpleasant set hypothesis. The items of a similar sort in FCA are named "formal objects," either the items of an alternate kind dubbed "formal properties". The adjective "formal" is utilized to affirm the concept formal. The formal objects do not have to necessarily be "objects" for any kind of logical meaningful of "object". However, utilizing " object" and " attribute " gives a sign in light of the fact that in numerous applications, it might be helpful to choose an object like elements as formal items and their components or properties as formal traits.

FCA The way to analyze the information and to represent the knowledge and data administration (Stumme, 2002). Additionally, built up an FCA-based approach of breaking down the impact of a smoothing system on data sparsity. Can be regarded the documents in the application of retrieval of information as "object-like", while the terms can seen an "attribute - like" (Cimiaon *et al*., 2005). More cases a group of formal items and their formal qualities are the tokens and the kinds, qualities and information kinds, information-is drove actualities and speculations, words and implications and so on (Priss, 2006). FCA has pragmatic applications in the ranges of information mining, content mining, learning administration, machine learning, programming improvement, science, semantic web, et cetera (Wang and Liu, 2008). Further review utilized FCA by proposing an instrument to enhance IR in a website in light of FCA. The said instrument makes semantic relations through questions and permits the revamping of ideas in the state of a lattice. The answers are then put together by a web index (Qadi *et al*., 2010). Introduce an approach in view of Formal Concept Analysis, a strategy in light of request hypothesis and for the most part utilized for the examination of information, specifically to discover natural connections between objects portrayed through an arrangement of traits from one perspective and the characteristics themselves on the other. So as to get qualities from a specific corpus, additionally contrast this approach and various leveled agglomerative grouping and in addition to BiSection K-means as a case of a divisive clustering method.

Conceptual Graphs Model (CGM): The third type of graphs that we have outlined is the "Conceptual Graphs" (CG) Were Sowa and Way (1986) proposes "Conceptual Graph Model" (CGM) which is a greater ability to be envisioned for comprehension. In Montes-y-Gómez *et al*. (2000) and Schenker *et al*. (2003) The specialists gave great importance to extract the features of texts or for classification work by utilizing this type of graph. Consisting language for knowledge representation. It is well established in the field of psychology, philosophy, linguistics. In CGS could be representing the knowledge structure on the semantic level. The CGs include two components: concept and relation. The CGs are thusly bipartite, associated and limited. An arrangement of edges and nodes of vertices is incorporated on a diagram. The CGs disentangle the relations of any arity to whatever remains of the system dialects that utilization a named circular segment.

The CGs are like diagrams utilized as a part of typical regular dialects. Exactly and profoundly organized data can be enough spoken to by CGs. Built CG is Often regularly utilized for chart coordinating; it produces outcomes that are dependable for different purposes. The strategy of looking at information in the content is utilized by representing the content of the document with CG formalism and perform the CG match. In Chu and Cesnik (2001) and Carninci *et al*. (2005) been utilized the CG on Different works to capture the normal structure of the text. The majority of the works is utilized parsing projects to take on the linguistic structure of the content before changing over them into CGs. In view of their ability, worked CG in this examination to catch effectively the semantics and structure of the extracted data. In hospitals are utilized the conceptual graphs to catch the structure and semantic data/information set out in medicinal document free text. Requesting and self-sorting out methods ("lattice techniques and knowledge space") were utilized to enhance the association of ideas from standard medicinal classifications and extensive arrangements of free content therapeutic archives (Chu and Cesnik, 2001). As Hensman and Dunnion (2004) utilized CG representation for ordering "Extensible Markup Language (XML)" archives. Data are included is installed as a meta-tag in the archive. This approach introduced two stages; defining semantic parts, then utilizing these parts with the knowledge to build specific CGs.

Likewise, Projection algorithm is centered around basic resemblance amongst CG and the implementation time is best fully NP. At Siti (2011) built the graphs in view of two propositions ; to start with, there is a

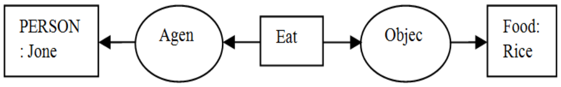

Fig. 2: An example of conceptual graphs

relation among the words in a sentence inside a modulo frame estimate with an ideal size of 6 (if the separation between terms is equal six tokens or less than, the edges of the graph are created), second, the nearest word considers the powerful association. A few reviews concentrate on robotized thematic marking that can convey advantages to clients going for breaking down and understanding report accumulations and also for web indexes focusing on the linkage between gatherings of words and their latent themes.

The current ways to deal with accomplishing this endure in quality, nevertheless, have been improved by concentrating on the structure at data mode. From Hulpus *et al*. (2013) the outcomes. It was found it can recognize the ideas that represent the best themes by utilizing the graph-based approach. The benefits of this the kind of graphs capable of capturing the relation between words. However the inconveniences of these sorts of charts: Comparison of graphs is arithmetically complex. One disadvantage of this approach (CG) is that the correlation procedure gets to be distinctly polynomial and includes an extensive number of parameters.

Example of Conceptual Graphs (CGS): There are some rising methodologies of utilizing an entire representation of content than just words and straightforward relations between words. One of the regular techniques to catch the semantic relations between words is given by Conceptual Graphs (CGs), the model presented by Sowa and Way (1986). In CGs, there are two sorts of nodes which are Concepts and Relations. Among them, a Relation node shows the semantic part of the episode ideas. For example, the sentence "“Jone is eating rice” can be represented as a conceptual graph as in Fig. 2. The rectangles and circles in the chart are Concepts and Relations, separately. An arc pointing indicating a circle denotes the primary contention of the connection and a bend indicating far from a circle check the last contention. Here, Eat is a nonspecific idea while Jone and Rice are singular ideas of PERSON and FOOD. The relations from Eat to Jone and from Eat to Food are named Agent, Object, individually, in light of the fact that Jone and rice play Agent and Object semantic parts in the current context.

Conceptual Graphs contain rich semantic data, so they can be utilized as a part of information representation. A semantic significance of a sentence can be gotten by making an interpretation of CGs to predicate analytics. The official standard for conceptual graph linguistic structure and semantics is the ISO/IEC 24707 standard for Common Logic, which characterizes the semantics as far as a dynamic language structure and model-theoretic semantics1. Be that as it may, it is difficult to change natural language content to CGs structures (Ordoñez-Salinas and Gelbukh, 2010).

There are many works in building CGs which can be partitioned into manual advancement, deterministic methodologies and measurable methodologies. For instance of deterministic approach, Hensman and Dunnion (2005) portrayals Construction semi-automatic conceptual graphical presentations of texts utilizing a mix of existing language resources, for example WordNet and VerbNet. The fundamental thought of this strategy is that the creators utilized VerbNet and WordNet to distinguish semantic parts. To start with, all records were changed over into XML design. At that point, they utilized a syntactic parser to parse every one of the sentences and recognize parts utilizing VerbNet. For every proviso in the sentence, the primary verb was distinguished and a sentence example is constructed utilizing parse tree. In every verb in the sentence, remove from VerbNet all the conceivable semantic edges.

At last, the conceptual graph for each sentence was worked by standard principles of CGs. In another work, Ordoñez-Salinas and Gelbukh (2010) Proposed the utilization of linguistic use in light of the reliance formalism and the standard characterized for Conceptual Graphs. The researchers utilized noun pre-modifiers and noun post-modifiers and additionally verb outlines, separated from VerbNet, as a wellspring of the meaning of semantic parts to manufacture the dependency grammar, which included verb classification, their syntactic portrayal and frame depictions. This sentence structure was intended for the gotten trees to look like CGs. To sum things up, by utilizing CGs, a rich semantic data of a content can be caught to a graph, however the reality remains that constructing this kind of graph is not a straightforward errand.

Dependency Graph (DG): The latter type in this study is some kind of graph is "dependency graph". Dependency Graph is a coordinated graph that indicates the dependencies of many items. DG is a sort of the representation scheme of the content which can be characterized in linguistic terms, as a strategy to imagine the structure of a sentence to show how distinctive words associate with each other utilizing direct connection called dependencies. The current approach allowed of dependencies for modeling relations between words, among terms or whole words sectors. Have a decision that appreciation or failure to attend an assessment, in order complement the reliance of the graph is possiblein reality (Balmas, 2004). This graph is a fitting representation of the relation of dependency. This graph is an independent language, meaning that are applicable to text normalization in any language. The graph includes a package of proposals (nodes), a cover employment of affirmation (node values) and fixed a series of dependency connections (that connects the brackets) restrict waive secrets.

The confidences are completely decided (one value), a particular part (many values), or part Unknown (all values). The focal points of this sort of graph find causal connections amongst the words and optimizes the performance of a similarity measure between the texts (Wang *et al*., 2011). Defined a dependency graph As per the Zimmermann and Nagappan (2008) is a coordinated graph "G = (V, E)", as V is an arrangement of nodes (pairs) and "E ⊆V×V" is an arrangement of edges (conditions). We will audit the earlier reviews the utilized dependency graph in the text representation. Where Dietrich *et al.* (2008) were utilizing the "Object Dependency Exploration Model" (ODEM) to separate the reliance graph. The dependency graph encrypted in "ODEM" incorporates the classes that actress as nodes. These nodes have an explanation that defines their classification. (The interface, class, explanation and so on.), deliberation, vision and final. Of each node likewise contains a list of relationships in one direction (dependencies) with the full name of a class (package name. Class name) indicated to and dependency classification explanation (utilizes, develops, or instruments).

This method enhances perception and consequently shows up the appearance of the graph is far less complicated. Whereas the researchers in Wang and Liu (2008) suggests a new model called "Feature Event Dependency Graph" (FEDG), that is able to provide more efficient knowledge compare with CGs model. Additionally, has been presented another clustering method that combining the dynamic investigation and static dependencies; the dependency graph incorporates the link representation of the basic relationships between the classes. A Graph is a coordinated diagram at the latest with two edges among two categories, all edges have a similar weight. To extract the structural relations is a programmed undertaking bolstered by various tools. Extraction apparatuses vary in their support of different innovations (Patel *et al*., 2009). A few specialists proposed a diagram based approach that utilizes two areas -independent graph representations to cover the content (site pages and email) (Chakravarthy *et al*., 2010). Are chosen graphic representations on the basis of area knowledge for the provision emphasis on the different fields. During the same year by both the of the authors (Mitchell and Mancoridis, 2010) utilized source code examination framework for the creation of a dependency graph that representing framework modules and module-level between relations. After that utilized this graph as an entry point to the bunch tool, which segments of the graph. The outcomes were introduced in a clustering graph utilizing graph visualization. Revealed the empirical results acquired by the Wang *et al.* (2011) the algorithms that rely on the graph shows better perform in a specific document from methods which are based on the BOW Form.

This method is also able to determine the causal relations and enhances the execution of the similarity measure among the texts. In Beck and Diehl (2013) was found on the new approach, which entails integration of dependency graphs before the implementation of the clustering and a connected arrangement of operations, for example "the Union, weighted union and the intersection of a group of the edges". The authors inferred that joining both methodologies enhances the comprehensive the quality of the clustering. Table 1 presents the pros and drawbacks of each kind of text representation schemes.

EXAMPLE OF GRAPH MODELS FOR WEB DOCUMENTS (DEPENDENCY GRAPH).

We have the documents represented as a graph. Graph data structures can without much of a stretch catch the nonlinear connections of the nodes and documents containing different parameters which can be a nonlinear associated consequently a graph can simply represent this data (Wang *et al*., 2011). Balmas (2004) represents documents as a dependency graph approach. Document dependency graph G is meant as G = (V, E).

Where V = {V1, V2, ..., Vn} is the arrangement of vertices in the chart, every word (wi) in the document represented as a vertex (vi) and the arrangement of edges(ej) represent as E = {e1,e2,….,em}, where the relations among words represents the edges between vertices. Consider beneath an illustration having document 1 and 2.

Doc1: "Maker of iphone is "APPLE". Steve Jobs was the CEO at Apple".

Doc2: "The CEO of Apple was Steve jobs."IPHONE" maker is Apple". A dependency graph of document 1 is shown in Fig. 3.

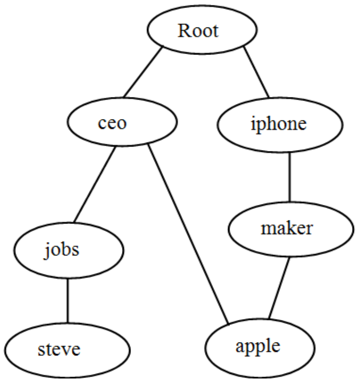

Fig. 3: example of a dependency graph

Table 1: Summary of text representation scheme

Author	The proposed approach	Advantages
Valatkaite and Vasilecas (2004)	Conceptual Graphs Model (CGM)	Conceptual graphs-is flexible and usable at different levels of IS development activities
Wang *et al.* (2011)	Dependency Graph (DG)	Discover causal relationships and improves the performance of a similarity measure between texts
Wang and Liu (2008)	Formal Concept Analysis (FCA)	Able to capture the relation between words
Rajaraman and Tan (2003)	Concept Frame Graph (CFG)	Provides a picture about semantic frames, targets and semantic roles in a text based on Frame Semantics theory and FrameNet linguistic resource.

Author	Disadvantages
Valatkaite and Vasilecas (2004)	A Conceptual Graph (CG) is an unordered list of concepts and supports only data
Wang *et al.* (2011)	Required to improve the visualization
Wang and Liu (2008)	Comparing graphs are computationally complex
Rajaraman and Tan (2003)	The contained semantic information is still shallow

Both of these documents have the same words but in different sequence, yet they ought to have similarity as they are semantically same, discussing the same subject. In the model displayed by Wang *et al.* (2011). In Balmas (2004) Both archives, document1 and 2 relate to a similar dependency graph which effectively shows that they are semantically equivalent to each other. Practically, this is done by building document graph in steps:

- The implementation of cleaning the data by removing all numbers and determinants, eliminating the triple and twofold spaces and changing over every one of the words into lower case.
- The document is isolated into sentences and afterward prevent words are expelled from the sentences.
- Stanford parser in Patel *et al.* (2009) is utilized to get word dependencies from the cleaned sentences.
- Using word conditions, acquired from Standford parser and non-stop of the document, the graph is built in following steps:
    - Vertices and edges are included by handling each sentence in the document.
    - For each sentence, we parse it utilizing the reliance parser, which yields an arrangement of words and the recognized pairwise relations between them.
    - The edges amongst vertices represent the Pairwise relation between words and a set of vertices as a non-stop word.
    - The lengths of the considerable number of edges in the diagram are set to 1.
    - Getting dependency graph is converted to weighted dependencies by computing the weight of each vertex utilizing tf-idf measure. Every vertex vi in diagram G relates to word wi of document graphs.

The graph is conversion to feature weight matrix. Utilized the similarity measure of Wang *et al.* (2011). Balmas (2004) to compute the compare among the two graphs. Table 1 displays the pros and cons of each type of graphic used to represent texts.

CONCLUSION

Sparsity all too familiar in the field of statistical modelling problem. The significance of the subject has increased in recent. One of the basic qualities of real-world data is the sparsity problem. Clustering is a system used to reduce of the sparsity issue. The paper provides kinds of graphs used in the text representation and its impact on the clustering procedure which helps to decrease the problem of sparsity of huge document sets. Where to display the four types of graph. The main advantage of these techniques originates from: Find causal relations and enhances the execution of a similarity measure between texts. Exploratory outcomes demonstrate that our technique can accomplish critical execution change over the conventional clustering strategies, for example, tf-idf and Bow.

REFERENCES

Abdulsahib, A.K. and S.S. Kamaruddin, 2015. Graph based text representation for document clustering. J. Theor. Appl. Inform. Technol., 76(1): 1-13.

Balmas, F., 2004. Displaying dependence graphs: A hierarchical approach. J. Softw-Evol. Proc., 16(3): 151-185.

Beck, F. and S. Diehl, 2013. On the impact of software evolution on software clustering. Empir. Softw. Eng., 18(5): 970-1004.

Bloehdorn, S., P. Cimiano, A. Hotho and S. Staab, 2005. An ontology-based framework for text mining. J. Comput. Linguist. Lang. Technol., 20(1): 87-112.

Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells *et al*., 2005. The transcriptional landscape of the mammalian genome. Science, 309(5740): 1559-1563.

Chakravarthy, S., A. Venkatachalam and A. Telang, 2010. A graph-based approach for multi-folder email classification. Proceeding of the IEEE 10th International Conference on the Data Mining (ICDM).

Chu, S. and B. Cesnik, 2001. Knowledge representation and retrieval using conceptual graphs and free text document self-organisation techniques. Int. J. Med. Inform., 62(2-3): 121-133.

Cimiaon, P., A. Hotho and S. Staab, 2005. Learning concept hierarchies from text corpora using formal concept analysis. J. Artif. Intell. Res., 24: 305-339.

Dietrich, J., V. Yakovlev, C. McCartin, G. Jenson and M. Duchrow, 2008. Cluster analysis of Java dependency graphs. Proceeding of the 4th ACM Symposium on Software Visualization, pp: 91-94.

Harish, B.S., D.S. Guru and S. Manjunath, 2010. Representation and classification of text documents: A brief review. IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition", 2: 110-119.

Hassan, G.S., S.K. Mohammad and F.M. Alwan, 2015. Categorization of 'Holy Quran-Tafseer' using K-nearest neighbor algorithm. Int. J. Comput. Appl., 129(12).

Hensman, S. and J. Dunnion, 2004. Automatically building conceptual graphs using VerbNet and WordNet. Proceeding of the International Symposium on Information and Communication Technologies, pp: 115-120.

Hensman, S. and J. Dunnion, 2005. Constructing conceptual graphs using linguistic resources. Proceeding of the 4th WSEAS International Conference on Telecommunications and Informatics (TELE-INFO'05). Stevens Point, Wisconsin, USA, Article No. 34.

Holder, L.B., 2009. Graph-based Data Mining. In: Encyclopedia of Data Warehousing and Mining. 2nd Edn., IGI Global, pp: 943-949.

Hulpus, I., C. Hayes, M. Karnstedt and D. Greene, 2013. Unsupervised graph-based topic labelling using dbpedia. Proceeding of the 6th ACM International Conference on Web Search and Data Mining, pp: 465-474.

Mitchell, B.S. and S. Mancoridis, 2010. Clustering module dependency graphs of software systems using the bunch tool. Department of Mathematics and Computer Science, Drexel University, Philadelphia, PA, USA.

Montes-y-Gómez, M., A. López-López and A. Gelbukh, 2000. Information retrieval with conceptual graph matching. Proceeding of the International Conference on Database and Expert Systems Applications, LNCS, 1873: 312-321.

Ordoñez-Salinas, S. and A. Gelbukh, 2010. Information retrieval with a simplified conceptual graph-like representation. Proceeding of the 9th Mexican International Conference on Artificial Intelligence (MICAI'10), Part I. Springer-Verlag, Berlin, Heidelberg, pp: 92-104.

Patel, C., A. Hamou-Lhadj and J. Rilling, 2009. Software clustering using dynamic analysis and static dependencies. Proceeding of the 13th European Conference on IEEE Software Maintenance and Reengineering (CSMR'09).

Priss, U., 2006. Formal concept analysis in information science. Annu. Rev. Inform. Sci., 40(1): 521-543.

Qadi, A.E., D. Aboutajedine and Y. Ennouary, 2010. Formal concept analysis for information retrieval. Int. J. Comput. Sci. Inform. Secur., 7(2).

Quynh, T.N. and A. Napoli, 2012. A graph model for text analysis and text mining. M.Sc. Thesis, Université de Lorraine.

Rajani, N., K. McArdle and I.S. Dhillon, 2015. Parallel k nearest neighbor graph construction using tree-based data structures. Proceeding of 1st High Performance Graph Mining Workshop. Sydney, August 10, 2015.

Rajaraman, K. and A.H. Tan, 2002. Knowledge discovery from texts: A concept frame graph approach. Proceeding of the 11th International Conference on Information and Knowledge Management, pp: 669-671.

Rajaraman, K. and A.H. Tan, 2003. Mining semantic networks for knowledge discovery. Proceeding of the 3rd IEEE International Conference on Data Mining.

Schenker, A., M. Last, H. Bunke and A. Kandel, 2003. Classification of web documents using a graph model. Proceeding of the 7th International Conference on IEEE Document Analysis and Recognition.

Siti, S.K., 2011. Frame work for deviation detection in text. Thesis, Universit Kebangsaan Malaysia, Bangi.

Sowa, J.F. and E.C. Way, 1986. Implementing a semantic interpreter using conceptual graphs. IBM J. Res. Dev., 30(1): 57-69.

Stumme, G., 2002. Formal concept analysis on its way from mathematics to computer science. In: Priss, U., D. Corbett and G. Angelova (Eds.), Conceptual Structures: Integration and Interfaces. ICCS-ConceptStruct, 2002. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2393: 2-19.

Valatkaite, I. and O. Vasilecas, 2004. Automatic enforcement of business rules as ADBMS triggers from Conceptual Graphs model. Inform. Technol. Control, 31(2).

Wang, L. and X. Liu, 2008. A new model of evaluating concept similarity. Knowl-Based Syst., 21(8): 842-846.

Wang, Y., X. Ni, J.T. Sun, Y. Tong and Z. Chen, 2011. Representing document as dependency graph for document clustering. Proceeding of the 20th ACM International Conference on Information and Knowledge Management, pp: 2177-2180.

Wille, R., 1982. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. In: Rival, I. (Eds.), Ordered Sets. Springer, Netherlands, pp: 445-470.

Zimmermann, T. and N. Nagappan, 2008. Predicting defects using network analysis on dependency graphs. Proceeding of the ACM/IEEE 30th International Conference on Software Engineering (ICSE'08).