The use of Semantic Role Labelling with Triangle-Graph Based Text Summarization

Yazan Alaya AL-Khassawneh
Artificial Intelligence Department, Zarqa University, Jordan, ykhassawneh@zu.edu.jo

ABSTRACT

Text summarization aims to create a condensed version of a text, thus retaining the original ideas. In particular, the textual content on the web is increasing at an exponential pace. The ability to interpret valuable information from such a vast volume of data is an essential undertaking and requires an automated system to assist with the current information repository.

Text summarization systems aim to help reducing content by retaining the relevant material and filtering out non-important sections of the text. There are two basic approaches in the text summarization systems. Single document summarization is outlined in the first method.

In other words, the method uses one document as input and generates a summary version as output. An alternative approach is to take many documents as input and generate a single summary document as output. In terms of performance, the summarization systems are often classified into two groups.

One method will be to remove exact sentences from the original document in order to create a summary production. An alternative would be a more nuanced approach in which the text rendered is rephrased version of the original document.

A formal semantic representation is provided in this paper, which can capture the algorithms and the text meaning to allow mapping the text documents with the meaning representation. This semantic representation, which focuses on establishing relations amongst concepts of text constituents, is considered to enhance the performance of the single-document extractive summarisation processes. To see how text constituents (i.e. words, phrases and morphemes) are related, a graph semantic model (GSM) was used, which is built using a syntactic and semantic analysis. Semantic Role Labelling (SRL) is the fundamental relationship that a contributor has with the main verb in a sentence.

Key words: Extractive Summarization, Triangle Counting, Graph-Based Summarization, Feature Extraction, Semantic Role Labelling, Graph Semantic Model.

1. INTRODUCTION

Since the increasing growth of Internet Technology, the number of electronic documents on the Internet has increased exponentially.

Today, people can acquire information from various sources. Despite the progress achieved in the document summarization field, the problem hasn't been resolved fully. Regarding the automated document summarization, it seeks finding methods for detecting the most significant information existing in a text [1]. In addition, it is important to identify certain features through processing data. That is needed for reducing the amount of time needed for accessing information [2]. In the light of that, people's interests in researching about automated text summarization systems has been increasing [3], [4].

Automated text summarization technologies are therefore rapidly being investigated by scientist in order to accomplish higher efficiency through improved or new methods [5]. Nevertheless, given all the studies conducted in the field of document summarization, the need for change and creativity has not decreased [1].

Computerized document summarization is a prominent subtopic of natural language processing (NLP), with the goal of presenting long text records in a compact and intelligible form [6]. Document summarization methods can usually be categorized into two types: extractive and abstractive.

Extractive document summarization comprises of three stages: text description, sentence score and sentence collection. Abstractive document summarization methods describe the key component of documents utilizing natural language production strategies and then re-express them to construct a summary [7] , [8], [9].

They may also be classified as single and multi-document, based on the quantity of documents to be summarized.

2. RELATED WORK

A. *Graph-Based text summarization*

Scoring statements or phrases and collecting summaries is the most popular approach used in automatic extractive

summarization. Sentence scoring is implemented in the majority of approaches used today. Rating methods are listed as word score, sentence score, and graph score [10], [11].

In the word ranking methods, a ranking is made taking into account the significance of sentences including the occurrence of a word in the text [12] – [13], with terms such as proper nouns, locations and objects that assumed to be a determinant being scored higher [7], [14].

The formal properties of the words (emboldened, italicized, underlined) are taken into account in text scoring methods [15].

In addition, sentences beginning with phrases such as ''Briefly," ''Finally" and ''As a result" in the text are described as sign phrases, and the sentences preceding such statements are known as significant sentences [12]. In the same way, the assessment is based on the title of the text to be summarized. Sentences containing the words mentioned in the title are assumed to be added to the summary, and their degree of significance is expand accordingly [16].

The methods of sentencing often take into consideration the characteristics of sentences, giving greater priority to sentences of greater size [17], [18]. Points are applied to sentences by deciding the location of the sentence and whether or not they have numerical values [13] ,[15] ,[19].

The authors of Ref. [20] explained the implementation and evaluation of extractive summarization method as a means of supporting learners with reading challenges. Graph-based representations are often used in text analysis approaches because they have very efficient solutions.

In Ref. [21], the researchers introduced TextRank, which provides graph-based summary representation using text intersections. Likewise, LexRank was implemented in Ref. [22] using a proprietary centrality-based algorithm, one of the node centrality methods.

Both the TexRank and LexRank strategies were motivated by the PageRank [23] technique, a text summarization method that was introduced to achieve central sentences in a text using shared knowledge between the word and sentence sets [24].

The researchers of Ref. [25] identified the graphic documents in their study using link generation for automated document summarization. They established the structure by unveiling the relationships of the text in the documents and analyzing the summaries by contrasting them with those produced by human hand.

In Ref. [26], a graph-based approach was proposed to ensure semantic consistency, with nodes referring to the terms of the documents and the edges representing the semantic relationships between the nodes. In general, the graph diameter calculation is implemented for all the nodes in the graph, and the shortest and longest paths are defined as the poorest and best bonds. Although graph structures and documents have been described in Ref. [27], nodes and edges

have been generatedwere based on local similarities. Random Walk was used to provide a list of the relevant documents.

A summary method for the biomedical sector was presented in Ref [28]. A graph based on concepts and relationships with a semi-dictionary- based framework was generated using a method called Unified Medical Language, and then a PageRank algorithm was used. In Ref. [29], the authors suggested a new graph based on improved random walking.

In Ref. [30], the researchers adopted a graph-based approach for carrying out an extractive summarization. The latter researchers proposed a new method for summarization. This method is based on hybrid modelling graph. They recommended delivering an innovative hybrid similarity function (H). The latter function hybridises 4 distinct measures of similarity. These measures are cosine, Jaccard, word alignment and window-based similarity. The method employed a trainable summarizer and takes into consideration several features. The impact of those features on the task of summarization has been examined.

B. *Semantic Role Labelling (SRL)*

Semantic Role Labelling (SRL) is the fundamental relationship that a contributor has with the main verb in a sentence which is also known as theta role (generative grammar), semantic role, semantic case, thematic role and deep case (case grammar) [31]. The objective of SRL is to determine the argument structure of each predicate in a certain input sentence [31]. SRL is greater in producing arguments for every sentence semantically.

The SRL technique is a very shallow semantic processing method which is quite popularly used in the Natural Language Processing (NLP) studies [32]. Several tasks adapt SRL techniques to their own approaches e.g. Biomedical domain [32] , [33]. Conventionally, several SRL techniques are used for the documents in the newswire domain like question-answering.

In Ref. [34], researchers suggested the semantic frames or structures and also used similar frames for ordinary roles or themes like the FrameNet which was proposed by [35] and the PropBank introduced by [36]. In Ref. [37] researchers suggested several applications of the neural networks which are based on the SRL techniques for extracting relations from the biomedical documents automatically. This technique mainly uses the SENNA toolkit which was developed by the authors of Ref. [38] and has been used in several applications that deal with documents. The SENNA toolkit uses the neural network algorithms for extracting the arguments and the semantic roles from all the statements in a text. The authors, who introduced this toolkit, used it for extraction of the semantic relations between the various terms in a biomedical text document.

SRL is commonly used for the text content analytical tasks like text retrieval [39], information extracting [40], text

classification [41] and sentiment analysis [42]. In Ref. [43], the researchers introduced a new technique which combined the SRL approach with the General Statistic Method (GSM) for determination of the key sentences present in the text for the purpose of generating single-document extractive summaries. Initially, they used the SRL technique and the semantic similarity measure for computing the statement similarity score. Next, the GSM technique was used for computing the sentence scores based on the features, without considering their weights. Thereafter, the sentence scores, which were obtained by both these methods, were combined for assigning an overall score to every sentence and the high-ranked sentences were then extracted depending upon the 20% compression rate.

C.Triangle Counting Approach

Many studies were conducted for discovering fresh algorithms in order to count triangles in Data Graph Sets. Ref [44] aimed to explore a new sampling algorithm in order to count triangles. They aimed to implement the technique on networks that are considered big. They proved speed-ups which are seventy thousand times faster in the triangles of counting. Regarding the performance level of the concerned algorithm, it shall be deemed accurate when the densities of the triangles are considered mild. Ref [45] showed that the parallel development shows a high level. They displayed a fresh algorithm that is indiscriminate to estimate how many triangulations there are in a graph that is undirected. Regarding the algorithm, it employs the simulation method of Monte-Carlo for counting how many triangles are there.

Based on Ref [46], the researchers adopted an approach which can be used for adjusting a graph for turning it into an alternative graph along with showing a lower number of edges and nodes. The latter study aimed at utilizing the technique of counting triangles for mining the graph-based association rules. A triangle counting technique for the ARM that is graph-based was proposed for having the graph reduced for searching for item sets that are common. Regarding the triangle counting, it was merged with the ARM method that is graph-based. It involves 4 significant steps. These steps are: Demonstration of Data, Production of triangle, Demonstration of bit vector and Triangle combination with the ARM technique that is graph-based.

The performance level of the proposed method has been compared with the main ARM method that is graph-based. Based on the outcomes of the experiment, it was found that the developed method is capable of reducing the time of execution needed for the production of rule. It was found that the latter method creates a lower amount of rules along with showing an assurance level that is higher.

In Ref [47], a graph elimination technique called the Triangle Counting Method is introduced to pick the most important sentences of the text. The first step is to represent a text as a graph, in which the nodes are the phrases and the edges are the similarity between the phrases. The second step is to create the triangles, after the depiction of the bit vector,

and the final stage is to acquire sentences based on the values of the bit vector.

This study has shown that one graph can be transformed to another graph, with slightly reduced number triangles. The utilization of Adjacency Matrix Representation is straightforward and this feature has led to faster implementation times.

3. OVERVIEW OF APPROACH

In this section, we describe the steps we took to generate the summary based on the semantic role labelling method. The main steps that we undergo are as follows:

1. Documents preparation
2. Identify Sentence Boundary
3. SRL Approach
    1. Graph Semantic Model Construction
        i. Tokenization
        ii. Part of Speech Tagging
        iii. Syntactic Parsing
        iv. Semantic and Discourse Analysis
    2. Calculate Semantic Similarity based on WordNet
4. Calculate Sentence Score based on Semantic Similarity
5. Construct Semantic Sub-Graph
6. Summary Generation based on SRL
7. Evaluation and Results

A. Data Pre-processing

Regarding the initial stage in the summarisation process of texts, it involves data pre-processing. In the present study, the latter step consists from three sub-steps. These sub-steps are listed below:

- The segmentation of a text
- Removing the stop-words
- Stemming the word

Regarding step No. 1 in the segmentation process of text, it seeks dividing the text document into several sentences. The researcher employed the method of the stop words removal in order to remove the words that are meaningless. The researcher employed the stemming algorithm in order to delete the affixes (i.e. suffixes and prefixes) which is connected to a word. That is done to produce the root of the word. Through that step, the researcher extracted the significant words that exist in the concerned document. He also disregarded the other words. That shall significantly influence the similarity between the concerned documents.

B. Feature Extraction

Regarding the document that is textual, it's represented by the following set $D = (S_1, S_2, \ldots, S_k)$. S_i refers to the sentence which is found within a specific document D. Then, the extraction of features is implemented to the content that is textual. The primary useful sentence along with the structure of words shall be determined. Each document shall feature structures. Such structures include: proper-noun instances, thematic-word, sentence similarities, term weights, numerical data, sentence positions, sentence lengths, and title words. Information about such structures are listed below:

1. Title words: Regarding the higher scores, they shall be assigned to sentences which involve words that have been obtained from titles. The meaning of the contents shall get conveyed in the title words. This is determined as it's presented below:

$$F1 = \frac{Words_{Title} \, Count \, in \, sentence \, S}{Word_{Count} \, in \, Title} \quad (1)$$

2. Sentence lengths: Sentences that are shortened –e.g. author lines or date- shall be removed. Regarding the normalised length of each sentence, it is assessed. Regarding the normalised length of each sentence, it is determined as it's displayed below:

$$F2 = \frac{Words_{Count} \, in \, Sentence \, S}{Words_{count} \, in \, longest \, sentence} \quad (2)$$

3. Sentence positions: Regarding the higher scores, they are assigned to the sentences that occur further ahead in their paragraph. Regarding each paragraph that has n sentences, the score of each sentence is determined as displayed:

$$F3 = \begin{cases} 1 & for \, first \, sentence \\ \frac{n-i}{n}, where \, i = 1,2,3,\ldots,n & for \, other \, sentences \end{cases} \quad (3)$$

4. Numerical data: Regarding each sentence that shows numerical terms that replicates main statistical figures within the text, they shall go through the process of summarizing. As for the numerical score of each sentence, it is determined as it's listed below:

$$F4 = \frac{Numerical\_dat_{Count} \, in \, sentence \, S}{Sentence_{Length} \, (S)} \quad (4)$$

5. Thematic words: The quantity of domain-specific or thematic words showing maximum-possible relativeness and exist in a sentence shall be divided by the maximum quantity that exist in the sentences

$$F5 = \frac{Thematic_{word_{Count}} \, in \, sentence \, S}{Max(Thematic\_word_{Count})} \quad (5)$$

6. Sentence-to-sentence similarities: The methods of token-matching are employed for measuring the extent of similarity between each sentence and with other ones. A matrix [N][N] is set up. N refers to the overall amount of the sentences that have been found. It also refers to the components that are diagonal and fixed at 0. The assessment of sentences shouldn't be conducted in comparison to themselves. The similarity score of each sentence shall be assessed as below:

$$F6 = \frac{\Sigma[(S_i, S_j)]}{MAX[(S_i, S_j)]}, where \, i = 1 \, to \, N \, and \, j = 1 \, to \, N \quad (6)$$

C. Computation of semantic similarity between pair of predicate arguments

In this stage, the similarity level of the predicates argument shall be measured pair wise in accordance with accepted comparisons for the time-time, verb-verb, noun-noun and location-location. Regarding the semantic similarity of any predicate, it shall be measured in accordance with the similarity level of the pair words through the use of WordNet thesaurus in order for capturing the relationships that are found between sentences. The similarity in the semantic aspect is deemed as a score of confidence. It aims at indicating the semantic relationships existing between the meaning derived from a pair of sentences. If the pair words that exist in a predicate label are deemed identical or semantic relations are existent, like: hypernym, synonym, holonym and merony, the words shall be considered related to one another.

Based on the experimental results reached in Ref. [48], the measure of Jiang & Conrath shows the closest correlation with the human assessment in comparison to all the other measures of semantic similarity. Thus, this study aimed to employ the measure of semantic similarity that is developed by Jiang [49]. It aimed to employ this measure for computing the level of semantic similarity that exist between each pair of predicate arguments. The measure of Jiang is based on information content. It indicates that each concept within the WordNet has specific information [50]. Based on that measure, the similarity level of 2 concepts is determined based on the information that is shared by a pair of concepts.

WordNet was designed at the Princeton University [51]. It was designed for establishing a connection between 4 kinds of parts of speech. The latter types include: verb, noun, and adverb. They also include adjective. In WordNet, each part of speech words shall be organized into taxonomies. As for each node, it refers to a set of synonyms or a synset (synonyms set: group of words which have the same meaning or shall be seen in the same place within the thesaurus), and a gloss or definition which convey a specific meaning of a word. In case a certain word has several senses, it shall appear in several synsets at several locations in the taxonomy. The WordNet provides a definition for the relations existing between synsets. It provides that for the relations existing between word senses. Regarding each synset, it has a single parent and forms a tree structure. The latter structure permits the proximity of any pair of nodes to be deemed as one of the sub-sumer's function.

D.The Graph Semantic Model (GSM) Creation

For document summarisation processes, the graph semantic model (GSM) is considered highly suitable for scheme representation. The GSM employs sentence by sentence approach for representing information in a document. In this method, a directed acyclic graph, with relations as links and concepts as nodes, is created by converting each sentence. Moreover, the expression of knowledge within a document in concepts is based on predicate relations between concepts, their senses in the current context, and other aspects such as semantic variations and semantic roles, which are all captured as attributes of concepts. The document representation model is the sum of all representations of sentences.

In the present study, the researcher represents the nodes of the semantic graph through sentences. The edge $e_{s_i s_j}$ of the semantic graph is associated with similarity weight $Sim(s_i, s_j)$ that is between sentences s_i and $s_j (i \neq j)$. The weight of similarity between sentences is measured based on the measure of semantic similarity that was developed by Jiang..

In the present study, the researcher is concerned in significant similarities in semantic aspects. He aimed to define a similarity threshold which is set to 0.5 empirically [52]. Thus, a link was added between the sentences (nodes) that have a semantic similarity level within the range of $(0 < \beta \leq 0.5)$. Otherwise, there won't be link established. We let $Sim(s_i, s_i) = 0$ to avoid self-transitions. The sentences with a similarity that is higher than 0.5 shall be presumed equivalent in semantic aspects. It shall be presumed as free from the semantic graph in order. That's presumed to avoid having redundant sentences within the stage of generating summary. The weight of similarity $Sim(s_i, s_j)$ that is between sentences s_i and $s_j (i \neq j)$ shall be identified based on the measure of semantic similarity that was developed by Jiang.

E. Calculation of Sentence Score based on Semantic Similarity

The measures of sentence similarity play a significant role in carrying out the text-related research. They a significant role in carrying out the process of text summarization. In simple words, a measure represents how alike several items are when after comparing them with one another. An accurate measure for measuring sentence similarity shall improve the computer-human interaction. In the present study, we aimed to measure the sentence similarity. We aimed to measure it based on the semantic roles labelling. We aimed to measure it based on the semantic similarity as it it's presented through equation number 7. The sentences scores that are within the range of (0-1) are defined as shown below based on the semantic similarity combined with the features chosen for each sentence.

$$Sentence_{Score} = Sim(s_i, s_j) * BV(s_i) * \sum_{K=1}^{N} Feature\_Score_K(s_i)$$

(7)

Where $Sim(s_i, s_j)$ is the semantic similarity between s_i and $s_j (i \neq j)$, $BV(s_i)$ is the bit vector of sentence s_i and $Feature\_Score_K(s_i)$ is the score of features selected for each sentence.

F. Sub graph construction

The following step is represented in constructing the triangle sub-graph. Triangles adopt the fact that friends of friends tend- as a base line- to be friends. At first, the researcher created an Adjacency Matrix. The following below represents an algorithm for the creation of the adjacency matrix.

Next, build list of triangles representing the text. To find the triangles in the graph, De-Morgan's laws algorithm is used. The following is the algorithm of De-Morgan's lows.

G.Summary Generation

After measuring the scores of sentences, each one of the sentences located in the document got assigned based on its score value. Regarding the sentences that have triangle sub-graph structures, they got chosen for consideration. The latter sentences got associated two other ones at least. A rate of compression that is deemed as close to 20% of the core textual content -was deemed to be as instructive of the contents as the document's complete text [53]. In the last phase, the summarizing sentences got organized in accordance with the order of their conceptual occurrences as they were existing within the initial text.

The algorithm represents the method used to produce of a summary and is based on the Maximal Marginal Relevance (MMR) concept [54]. Sentences are measured on the basis of hybrid similarity values (threshold value). If a comparison, is made, if it is noted that the sentences have a similarity value greater than the threshold value, then the sentence is eliminated because the sentences already imply some correlation between them. Whether the value is greater than the threshold value, it shall be included in the summary. This allows all the potential ideas mentioned in the text to be included. In this way, any person reading the summary would understand the idea expressed by the text.

4. EXPERIMENTAL SETTINGS

To conduct an experiment on the implementation of the suggested semantic-graph based approach for the processes of the single-document extractive summarization, the researcher employed DUC 2002 document sets (DUC, 2002). In the beginning, the researcher employed perform pre-processing on the documents. The latter step includes word stemming and elimination of stop words. It also

includes tokenization, and sentence splitting. When the documents were pre-processed, the researcher applied SRL for extracting the predicate argument structure from the document's sentences. The researcher ranked the sentences based on 6 features with their probabilities combined with the Bit-Vector for each sentence and the semantic similarity value. At the end, the top ranked sentences got selected as being a summary for the primary text. They got selected based on the rate of compression. The researcher employed 3 measures for assessment. These measures are: Mean coverage score (Recall), Precision and F-measure for assessing the approach of this study.

5. EXPERIMENTAL RESULTS

The proposed semantic approach is assessed in the context of single-document extractive summarization task, using 103 articles/data sets delivered by the Document Understanding Evaluations (DUC, 2002). For each set of data, the approach of this study provides a summary with 20% rate of compression. For comparing the performance level of the proposed approach, the researcher conducted a comparison between the results of this study and several benchmark summarizers. These benchmark summarizers are: GSM-SRL [44], FazzyGA-SRL [55], Microsoft Word 2007 summarizer, Copernic summarizer, Best automatic summarization system in DUC 2002, Worst automatic systems in DUC 2002, and the average of human model summaries (Models). It was found that researchers who employed the GSM-SRL and FuzzyGA-SRL based methods have extracted only the ROUGE-1 results used to represent the methods performance.

For comparative evaluation, Table 1 shows the mean coverage score (recall), average precision and average F-measure obtained on DUC 2002 dataset for the semantic sub-graph approach using ROUGE-1, while Table 2 compares the semantic sub-graph approach using ROUGE-L. All these values are generalized at a 95% confidence interval. The present study focused on finding the influence of features. It focused on the employment of several measurements of similarity for creating the Triangles Sub-Graph. It sheds a light on using SRL for creating the Semantic Sub-Graph for the production of the summary. Based on the generalization of the results and the performance level of the developed model, employing the sub-graph obtained shall show the best value. That applies when employing single feature or combined ones.

Using single feature, in Triangles sub-graph, has the best value is related to Sentence to Sentence (S2S) feature, which is 51.213% similar to human generated summaries using ROUGE-1; and 51.315% similar to human performance using ROUGE-L. The same for combined features, the best sub-graph representation is the Triangles Sub-Graph, which is 54.387% similar to human generated summaries using ROUGE-1; and 51.414% similar to human performance using ROUGE-L. In addition, it is very clear from the results that by using combined features, the results are better than using single feature.

Table 1: Comparison of Single Extractive Document Summarization using ROUGE-1 result at the 95% confidence interval for Triangles SRL model using single feature

Method	Precision	Recall	F-Measure
H2:H1	0.51656	0.51642	0.51627
MS-Word	0.47705	0.40325	0.42888
Copernic	0.46144	0.41969	0.43611
Best-System	0.50244	0.40259	0.43642
Worst-System	0.06705	0.68331	0.1209
GSM-SRL	0.48886	0.44795	0.46336
FuzzyGA-SRL	0.49296	0.45590	0.47033
Triangles SRL	0.51213	0.5134	**0.51269**

Table 2: Comparison of Single Extractive Document Summarization using ROUGE-1 result at the 95% confidence interval for Triangles SRL model using combined features

Method	Precision	Recall	F-Measure
H2:H1	0.51656	0.51642	0.51627
MS-Word	0.47705	0.40325	0.42888
Copernic	0.46144	0.41969	0.43611
Best-System	0.50244	0.40259	0.43642
Worst-System	0.06705	0.68331	0.1209
Triangles SRL	0.51315	0.51445	**0.51414**

6. CONCLUSION

The aim of this study was to use Semantic Role Labelling (SRL) to create a Graph Semantic Model (GSM) that would be used to find a summary of a single document. This model was trained and validated using a series of one hundred and three documents obtained from the DUC 2002 dataset.

The system takes into consideration many features: Title words, Lengths of Sentence, Positions of Sentence, Numerical data, thematic words and similarities of sentences. Also, it focuses on the use of SRL, which focuses on establishing relations amongst concepts of text constituents, is considered to enhance the performance of the single-document extractive summarisation processes.

The findings of the recommended summarizer in this research were compared with specific summarizers, such as such as Microsoft Word 2007 summarizer, Copernic summarizer, Best system, Worst system, GSM-SRL and FazzyGA-SRL. The ROUGE toolkit has been used to test system summaries at 95% confidence interval and to obtain results using average recall, precision and F-measurement. The F-measurement was chosen as a validation criterion because it incorporates both the recall and precision of the results of the test.

The experimental findings based on the proposed system indicate that the Triangle-Subgraph using SRL (Triangles SRL) is the best illustration for a comprehensive summary. The results show that the best average precision, recall and F-Measure are produced by our proposed method.

REFERENCES

1. Ermakova L, Cossu JV, Mothe J. **A survey on evaluation of summarization methods.** *Inf Process Manage 2019;56(5):1794–814.* https://doi.org/10.1016/j.ipm.2019.04.001

2. Hark C, Seyyarer A, Uçkan T, Karci A. **Doğal dil isleme yaklasimlari ile yapisal olmayan dökümanlarin benzerliği.** *In: IDAP 2017 – international artificial intelligence and data processing symposium. p. 1–6.*

3. Ge Yao J, Wan X, Xiao J. **Recent advances in document summarization.** *Knowl Inf Syst 2017;53(2):297–336.*

4. Chitturi, Arun Krshna. **Survey on Abstractive Text Summarization using various approaches.** *International Journal of Advanced Trends in Computer Science and Engineering. 8. 2956-2964. 10.30534/ijatcse/2019/45862019.*

5. Hark C, Uçkan T, Seyyarer A, Karci Abubekir. Metin Özetleme _ **Için Çizge Tabanlı Bir Öneri**. *IDAP 2018 – international artificial intelligence and data processing symposium, 2018.*

6. Joshi A, Fidalgo E, Alegre E, Fernández-Robles L. **SummCoder: an unsupervised framework for extractive text summarization based on deep auto-encoders.** *Expert Syst Appl 2019;129:200–15.* https://doi.org/10.1016/j.eswa.2019.03.045

7. Gambhir M, Gupta V. **Recent automatic text summarization techniques: a survey.** *Artif Intell Rev 2017;47(1):1–66.*

8. R. B. Saroo Raj, Gurpartap Singh, Balaji S & K. H. Ajit Baskar. **A Model to Predict Loan Defaulters using Machine Learning.** *International Journal of Emerging Technologies in Engineering Research, Volume 6, Issue 10, October (2018)*

9. Tan J, Wan X, Xiao J. **Abstractive document summarization with a graph-based attentional neural model.** *In: Proceedings of the 55th annual meeting of the association for computational linguistics. p. 1171–81.*

10. Dr.Balika.J.Chelliah, Darshan Lathia, Sandeep Yadav, Meet Trivedi & Shubham Sagar Soni. **Sentiment Analysis of Twitter Data using CNN.** *International Journal of Emerging Technologies in Engineering Research (IJETER) Volume 6, Issue 4, April (2018)*

11. Ferreira R et al. **Assessing sentence scoring techniques for extractive text summarization**, *2013.*

12. Luhn HP. **The automatic creation of literature abstracts**. *IBM J Res Dev 1958;2 (2):159–65.*

13. Shardan R, Kulkarni U. **Implementation and evaluation of evolutionary connectionist approaches to automated text summarization**, *2010.*

14. Nasr Azadani M, Ghadiri N, Davoodijam E. **Graph-based biomedical text summarization: an itemset mining and sentence clustering approach.** *J Biomed Inform 2018;84:42–58.* https://doi.org/10.1016/j.jbi.2018.06.005

15. Student PG, Coe DM. **A comparative study of Hindi text summarization techniques: genetic algorithm and neural network**, *2015.*

16. Gupta V. **Hybrid algorithm for multilingual summarization of Hindi and Punjabi documents.** *In: Mining intelligence and knowledge exploration. Springer; 2013. p. 717–27.*

17. Gupta V, Lehal GS. **A survey of text summarization extractive techniques**. *J Emerg Technol Web Intell 2010;2(3):258–68.*

18. Abuobieda A, Salim N, Albaham AT, Osman AH, Kumar YJ. **Text summarization features selection method using pseudo genetic-based model**. *In: Proc. – 2012 int. conf. inf. retr. knowl. manag. CAMP'12. p. 193–7.*

19. Fattah MA, Ren F. **GA, MR, FFNN, PNN and GMM based models for automatic text summarization**. *Comput Speech Lang 2009;23(1):126–44.*

20. Nandhini K, Balasundaram SR. **Improving readability through extractive summarization for learners with reading difficulties**. *Egypt Inform J 2013;14 (3):195–204.*

21. Mihalcea R, Tarau P. **A language independent algorithm for single and multiple document summarization**. *In: Proc. IJCNLP 2005, 2nd int. join conf. nat. lang. process.. p. 19–24.*

22. Erkan G, Radev DR. **Lexrank: graph-based lexical centrality as salience in text summarization**. *J Artif Intell Res 2004;22:457–79.* https://doi.org/10.1613/jair.1523

23. Brin S, Page L. **The anatomy of a large-scale hypertextual web search engine**. *Comput Networks ISDN Syst 1998;30(1–7):107–17.*

24. Parveen D, Ramsl H-M, Strube M. **Topical coherence for graph-based extractive summarization**. *In: Proceedings of the 2015 conference on empirical methods in natural language processing. p. 1949–54.*

25. Salton G, Singhal A, Mitra M, Buckley C. **Automatic text structuring and summarization.** *Inf Process Manage 1997;33(2):193–207.*

26. Medelyan O. **Computing Lexical Chains with Graph Clustering**. *In: Proceedings of the ACL 2007 student research workshop. p. 85–90.*

27. Chen Y-N, Huang Y, Yeh C-F, Lee L-S. **Spoken lecture summarization by random walk over a graph constructed with automatically extracted key terms**. *In twelfth annual conference of the international speech communication association, 2011.*

28. Plaza L, Stevenson M, Díaz A. **Resolving ambiguity in biomedical text to improve summarization**. *Inf Process Manage 2012;48(4):755–66.*

29. Xiong S, Ji D. **Query-focused multi-document summarization using hypergraph-based ranking**. *Inf Process Manage 2016;52(4):670–81.* https://doi.org/10.1016/j.ipm.2015.12.012

30. AL-Khassawneh, Y. A., Salim, N., & Jarrah, M. **Improving triangle-graph based text summarization**

using hybrid similarity function. *Indian Journal of Science & Technolog, 10(8),2017.*

31. Ondas, S., Hladek, D., and Juhar, J. **Semantic roles labeling system for Slovak sentences**. *Paper presented at the Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on, 161-166.*

32. He, L., Lewis, M., and Zettlemoyer, L. **Question-Answer Driven Semantic Role Labeling: Using Natural Language to Annotate Natural Language**. *Paper presented at the EMNLP, 643-653,2015.*

33. Dahlmeier, D., and Ng, H. T. **Domain adaptation for semantic role labeling in the biomedical domain**. *Bioinformatics, 26(8), 1098-1104, 2010.*

34. Zhang, Y., Tang, B., Jiang, M., Wang, J., and Xu, H. **Domain adaptation for semantic role labeling of clinical text**. *Journal of the American Medical Informatics Association, ocu048, 2015.*

35. Fillmore, C. J. **The case for case reopened. Syntax and semantic**s, *8(1977), 59-82*

36. Baker, C. F. **FrameNet: a knowledge base for natural language processing**. *Paper presented at the proceedings of frame semantics in NLP: a workshop in honor of Chuck Fillmore, 1-5, 2014.*

37. Bonial, C., Bonn, J., Conger, K., Hwang, J. D., and Palmer, M. **PropBank: Semantics of New Predicate Types**. *Paper presented at the LREC, 3013-3019, 2014.*

38. Collobert, R., and Weston, J. **Fast semantic extraction using a novel neural network architecture.** *Annual meeting-association for computational linguistics, 560, 2007.*

39. Shehata, S., Karray, F., and Kamel, M. S. **An efficient concept-based retrieval model for enhancing text retrieval quality.** *Knowledge and information systems, 35(2), 411-434, 2013.*
https://doi.org/10.1007/s10115-012-0504-y

40. Del Corro, L., and Gemulla, R. **Clausie: clause-based open information extraction.** *22nd international conference on World Wide Web, 355-366, 2013.*

41. Sinoara, R. A., Rossi, R. G., and Rezende, S. O. **Semantic role-based representations in text classification. Pattern Recognition (ICPR)**, *2016 23rd International Conference on, 2313-2318.*

42. Jadhav, N., and Bhattacharyya, P. (2014). **Dive deeper: deep semantics for sentiment analysis**. *ACL 2014, 113.*

43. Salim, N., Suanmali, L., and Binwahlan, M. **SRL-GSM: a hybrid approach based on semantic role labeling and general statistic method for text summarization**. *(2010).*

44. Tsourakakis CE, Kang U, Miller GL, Faloutsos C. **Doulion: counting triangles in massive graphs with a coin.** *The Proceedings of the 15th ACM SIGKDD international con- ference on Knowledge discovery and data mining. 2009; 837-46. Crossref*

45. Avron H. **Counting triangles in large graphs using random- ized matrix trace estimation**. *The Workshop on Large-scale Data Mining: Theory and Applications. 2010; p. 1-10.*

46. Al-Khassawneh YAJ, Bakar AA, Zainudin S**. Triangle counting approach for graph-based association rules mining**. *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). 2012; p. 661-65.*
https://doi.org/10.1109/FSKD.2012.6233981

47. AL-Khassawneh YA, Salim N, Isiaka OA. **Extractive text summarisation using graph triangle counting approach: Proposed method**. *In1 st International Conference of Recent Trends in Information and Communication Technologies in Universiti Teknologi Malaysia, Johor, Malaysia 2014 (pp. 300-311).*

48. Li, Y., Bandar, Z. A., and McLean, D. **An approach for measuring semantic similarity between words using multiple information sources.** *IEEE Transactions on knowledge and data engineering, 15(4), 871-882, 2003.*

49. Jiang, J. J., and Conrath, D. W. **Semantic similarity based on corpus statistics and lexical taxonomy**. *arXiv preprint cmp-lg/9709008, 1997.*

50. Miller, G. **WordNet®: A Lexical Basis data for English**. *Communications of the ACM, 38(11), 40-41, 2014.*

51. Fellbaum, C. **WordNet: An On-Line Lexical Database**: *MIT Press, 1998.*

52. Mihalcea, R., and Tarau, P. **A language independent algorithm for single and multiple document summarization.** *(2005).*

53. Morris, A. H., Kasper, G. M., and Adams, D. A. **The effects and limitations of automated text condensing on reading comprehension performance.** *Information Systems Research, 3(1), 17-35, 1992.*
https://doi.org/10.1287/isre.3.1.17

54. Carbonell, J., and Goldstein, J. **The use of MMR, diversity-based reranking for reordering documents and producing summaries.** *Paper presented at the Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, 335-336, 1998.*

55. Suanmali, L., Salim, N., and Binwahlan, M. S. **Fuzzy genetic semantic based text summarization**. *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on, 1184-1191, 2011.*
https://doi.org/10.1109/DASC.2011.192