International Journal of Engineering & Technology, 7 (4) (2018) 5160-5166

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.21596
Research paper

A novel hybrid system with hierarchical semantic conceptual dependency parsing for documents text summarization

Dr. Abdulkareem Merhej Radhi *

[1]*Assist. Prof. - Al-Nahrain University- College of Information Engineering- Iraq*
**Corresponding author E-mail: abdulkareemradhi@gmail.com*

Abstract

The rapid development of Internet technologies and the growing number of users who exchange a huge information in a different subject causes the emergence need to extract only the important and useful information from multiple documents. This extraction is text summarization which reduces the time in predicting the purpose of these documents and a mean to access effective and right decisions in important fields. The huge amount of information lead to the problem of memory overload. So, text summarization will minimize the effects of this problem. To overcome the problems of time consuming in extracting a proper information from multiple documents and reduce the redundant information which affect the processing overhead, a novel approach for summarizing information based on Hierarchical Conceptual Dependency of words and sentences representation presented in this paper. Moreover, concepts in text documents with score and token ranks are semantic networks representation of this text achieved for summarization. The proposed technique applied to documents d061j to d070f which are available to compare the results of any proposed approach in text summarization field with the experts output of summarization for the same documents indicated as a (reference summary). The proposed system was evaluated to check the accuracy of results and the tested dataset are Documents Understanding Conference (DUC 2002). Accuracy and efficiency of the output summarization evaluated using ROUGE–N and ROUGE-L metric. Hardware Resources used was Laptop Dell CORE "I3" with RAM 4 GB. Java software was the environment platform for this output. Comparing output results with the related algorithms depicted in this paper for the same target documents reflects that the similarity values between output summary via applying the proposed approach are more appropriate, accurate and minimum cost effective from the related metods.

Keywords: *Tokens; Active Concepts; ROUGE-N; Summary; Rank.*

1. Introduction

Summary is a procedure for creating a short description of a source text [1]. Summarization systems often have additional evidence they can utilized to specify the most important topics of document(s). Different crucial applications depicts this importance, for example when summarizing blogs, there are discussions or posts which are a good sources of information to determine which parts of the blog are critical and interesting. Moreover, in scientific paper summarization, there is a considerable amount of information such as cited papers and conference information which can be leveraged to identify important sentences in the original paper. In the following, we describe some the contexts in more details.

A summary is a short text that is produced from one or more texts contains a significant portion of the information in the original text(s), and that is no longer than half of the original text(s) (Allahyari, P, and July, 2017) [2]. A summary can be defined as a text that is produced from one or more texts ‹that contains a significant portion of the information in the original text(s), and that is no longer than half of the original text(s). According to, text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or user) and task (or tasks) (Blanco R, Matthews M, and Mika P, 2017) [3]. Moreover, the evaluations on multi-document and multilingual datasets prove the effectiveness of the continuous vector representation of Tokens compared to the bag-of-Tokens model (Rossi E, 2017) [9].

2. Related work

Saleh and Kadhim [10] proposes two models for extractive multi-document summarization based on genetic algorithm (GA), described and modeled as a discrete optimization problem with two candidate expressions and a specific fitness function.

(McBurney and Collin McMillan,2015)[8], propose a source code summarization technique that writes English descriptions of Java methods by analyzing how those methods are invoked, then performed two user studies to evaluate this approach.

(Lal, P., 2002)[6] Summarization works by assigning scores to sentences in the document to be summarized, and using the highest scoring sentences in the summary.

(Sankar and Sobha, 2009)[12], offer an efficient text summarization technique that involves two basic operations via finding coherent chunks in the document and ranking the text in the individual coherent chunks and picking the sentences that rank above a given threshold.

(Blanco R, Matthews M, and Mika, P.,2017)[3], Propose a combination of a term- and a concept-based retrieval model that closes the semantic gap between queries and documents expanding both of them with category information.

3. Limitations of related work and conceptual-dependency parsing method novelty

After reviewing literatures in the field of texts summarization, it is detected that till now most used approaches have been achieved text summarization using multi documents. These approaches can be classified as either statistical or linguistic approaches where statistical approaches uses terms frequency vice document frequency (TF-IDF), MMR method, or pattern recognition methods such as (Naives-Bayesian method and others). While, Linguistic approaches such as textual entailment needs requirements of lexicon or lexical database. There are very few methods that have been used the combination of the both approaches. In spite of the statistical methods are faster and less processing time but they produce summarization output which not provide meaning or semantics. These semantics are embedded in the relationship between the text entities (objects). Moreover, since it is not cover all the events and entity relationships in the original text therefore it is not qualified and not coherent.

In order to overcome the above limitations of the statistical approaches, the researchers adopted in recent years on the linguistics approaches. These approaches overcome some limitations of the statistical approaches and provide text meaning but on the other hand they need different requirements such as lexicon, references, and multi documents. These requirements maximize system overhead, response processing time and minimize text summarization qualification. Text summarization is a solution to this problem of overload [13].

The proposed approach is a new model in text summarization. It expresses its entities as a semantic concept without requiring a knowledge base and produces a text summarization from a single document without requiring a multi documents. The results gained via implementing this approach for a single document are depicted in appendices of this paper which reflects its quality in extracting summarization. The utilizing some features as (sentences ranking, sentences score, sentences location and cue tokens …etc.) in the text processing phase before parsing have been affect maximizing text summarization qualification.

4. Problem statement and research questions

In recent years, the growth and continuous increasing of data has been amazing and great in all fields so it is very natural and clear the urgent need to find and provide appropriate methods and techniques to summarize and extract the necessary and useful data via normalizing redundant data that directly affect the efficiency of the system performance and processing the stored data under a certain platform and the costs of additional resources required for storing management in addition to the time required to answer many questions and queries that can be directed to extract specific and useful information and a knowledge in a specific field. Many of questions of the problem statement that can be arises in the field of extracting and summarizing texts, which can be answered by the objective used in this research which can be depicted as follows:

Question 1: What is the appropriate technique for extracting texts that are sufficient to produce summarization through a few texts that do not require a lot of information resources and substantial material costs and required by other trends mentioned in the literature of data extraction. That the trend used in this research provides adequate and appropriate answer to this question by relying on individual texts in the extraction of texts and does not require additional information resources for the production of extracts.

Question 2: Is it possible to provide the summarization of texts at a relatively short response time. The proposed technique should produce the extraction in real time proportional with the importance of the field in which these texts are spoken.

Question 3: Is it possible to obtain an efficient technique capable of determining the accuracy and quality of summarization produced from texts to cover all the important aspects that these texts state about. To answer this question, the new technique used when compared to the rest of the trends used in the extraction of texts noted the extent of efficiency and ability to cover the delicate aspects of the units addressed by those texts by extracting sentences and repetitive words and important texts treated.

5. Text processing

Different approaches adapted for document summarization. Most of these approaches propose identifying semantic and lexical relations between sentences which are the core of target documents. (Codina-Filbà J, Bouayad-Agha J, Burga, Gerard Casamayor A, and Wanner, L, 2017) [4] , use lexical chains in the claims and in the description of the invention and of aligned claim–description segments at the subsentential level to assess the relevance of the individual fragments of the document for the summary. It is necessary to obtain texts for documents containing only the important sentences that reflect the summary of those documents. Therefore, the Tokens mentioned in the list of stop tokens such as "the". "as", "about", "above", "about"…etc., which are provided in a text file were first of all removed by a simple algorithm that eliminates those Tokens in each document. Actually, this list of Tokens does not affect the content of text, so, sentences are easy to determine which can be individual or compound. Therefore, the identification of the document, the number of the sentence in each and the related tokens was constructed. The stem or conversion of tokens into their original roots has been identified. Moreover, tokenization is the second process in identifying tokens. Space is first delimiter used to identify tokens in a text, and there are some widely accepted punctuation delimiters such as Period (.), Comma (,), Semicolon (;), Quotation marks ("), and Colon (:). Actually, Figure [2] presents a certain algorithm has been identified, including the removal of the extra characters so that we get the original Token. The tokens mentioned in list (2) represent the intervals used to separate the Tokens in the different sentences. It is also necessary to identify the frequency of Tokens contained in each document. This repetition of Tokens represented according to Equation (1) which identify the weight of the Token for the extracted text.

5.1. Cue tokens

Extracting semantics from text or sentences can be linearly mining via determining a series of sentences or tokens after a specific connective tokens called a cue tokens like (now, meanwhile, anyway, or on the other hand, therefore…etc.) that links spans of discourse and signals semantic relations in a text.

5.2. Sentence position and main headers

One of the specific feature contributed in text summary is the location of the processed sentence in a text. The sentence in a first position of discourse or in the end will indicate that it sometimes related to text title or the main subject discussed. So, the proposed research took in a consideration this fact via feeding a weighted numerical score to a ranked sentence as shown in a proposed algorithms presented in Figure [1].

6. Summarization and conceptual dependency

The summarization of a particular document or a set of documents can be classified into two types: in the first class, the summarization is the most frequently repeated sentences in that document (documents), while, the second class is an abstract of these documents. This second type is harder than the first one, it expresses the content of those documents. The summarization of those documents can also be categorized according to the aim of how much the summary may be helpful to cover the reader intent. So, summarization may be indicative, such that via the summarization, the reader can conclude that the document in worthwhile of attention and needs to be read the entire document to see its full intrinsic

facts. The summary may be informative, such that it can contain a wealth of information and facts.
On the other hand, the summarizations in the research and scientific papers are critical. It can include the idea and review of the method used in the research with criticism and discussion of the results founded by the researcher. Persons, through text summarization, recognize the concepts of source text document and produce a summary which expresses the kernel of the document whereas in automatic systems is a complex task [14]
The proposed system adopts the combination of the two above categories which is a new trend in this area, in spite of that this trend needs first complete the identification of the first category, i.e. the most frequently identified sentences in the collection of documents required to be extracted. The proposed system also adopted the use of a new technique not previously used in the field of summarization of those documents, namely the use of (Conceptual Dependency).
Conceptual Dependency is a model text and sentences representation not through the words of those sentences, but through a group of primitive concepts that reflect the exact meaning of those sentences and texts. CD, therefore provides a general structure and a group of primitive concepts that when gathered give the meaning to texts. Concepts such as the relationships between different entities and physical movements like push by a certain power and the movement of objects by the owner and the mental transferring of a certain idea depending on a pretext of the conversions and extrapolations of a certain idea from historical (old) ideas as well as the expression of ideas using sounds, as well as listening to a particular view and the action caused by a certain feeling such as hunger, thirst, and Mentality….etc.
CD can also be classified into four main categories:

ACT	Actions {one of the CD primitives}
PP	Objects {picture producers}
AA	Modifiers of actions {action aiders}
PA	Modifiers of PP's {picture aiders}

CD includes tools for concepts that reflect the time period in which an event occurred, occurred or will occur in the future ... In addition to the fact that the difference in expression of a certain event occurred at a time, the use of CD will provide the same meaning. Conceptual dependency is a model for understanding the contents of the documents in natural language. It includes a set of tools that simulate transfer ideas to persuasion as a concepts of verbal mental, focusing on concepts rather than grammar, focus on understanding content rather than language and grammar structure, to understand content and focus on the main title. It offers a right contents even if the words have been changed and then the conclusion from them. Table [1] shows this concepts, its purpose and specific instance of each.

Table 1: Active Concepts

Active Concept	Purpose	instance
ATRANS	Transfer of an abstract relationship	give
PTRANS	Transfer of the physical location of an object	go
PROPEL	Application of physical force to an object	push
MOVE	Movement of a body part by its owner	kick
GRASP	Grasping of an object by an action	throw
INGEST	Ingesting of an object by an animal	eat
EXPEL	Expulsion of something from the body of an animal	cry
MTRANS	Transfer of mental information	tell
MBUILD	Building new information out of old	decide
SPEAK	Producing of sounds	say
ATTEND	Focusing of a sense organ toward a stimulus	listen

The dependencies among conceptualization correspond to semantic relations among the underlying concepts. Inferences Associated with Primitive Act thus General inferences are stored with each primitive Act therefore reducing the number of inferences that need to be stored explicitly with each concept. For example, from a sentence "John killed Mike", we can infer that "Mike is dead". Let us take another example of primitive Act INGEST.
The following inferences can be associated with it.
The object ingested is no longer available in its original form. If object is eatable, then the actor has less hunger. If object is toxic, then the actor's heath is bad.
The physical position of object has changed. So PTRANS is inferred.
Example: The verbs {give, take, steal, donate} involve a transfer of ownership of an object. If any of them occurs, then inferences about who now has the object and who once had the object may be important. In a CD representation, these possible inferences can be stated once and associated with the primitive ACT "ATRANS". Consider another sentence "Bill threatened John with a broken nose": Sentence interpretation is that Bill informed John that he (Bill) will do something to break john's nose. Bill did (said) so in order that John will believe that if he (john) does some other thing (different from what Bill wanted) then Bill will break John's nose.

7. Representation rules and parsing

Rules are to be carefully designed for each primitive action in order to obtain semantically correct interpretation. The conceptual dependency representation which is adopted in this paper work can be used as a semantic model for reprsenting knowledge, because its theory is based on the representation of events as well as all information related to events.
The rules must be carefully designed for each primitive work in order to obtain a true semantic interpretation. Conceptual parsing is required for generating CD representation from source sentences in natural language. The main steps involved in CD parsing are as follows:
Syntactic processor extracts main verb and noun along with syntactic category of the verb (transitive or intransitive) from the sentence. Conceptual processor then makes use of verb-ACT dictionary. Once the correct entry from dictionary is chosen, CD processor analyses the rest of sentence looking for arguments for empty slots of the verb. Conceptual Dependency examines possible interpretation in a well-defined order. For the purpose of investing and employing conceptual dependency in establishing the dependency of semantic relationships between sentences in different texts. The proposed research presents a set of rules shown in table [2] that were used to construct this dependencies and to obtain an affected summary of these texts. These rules are characterized by the construction of a hierarchical model of the semantic meaning of the sentences after conducting analytical and statistical operations aimed at producing the sequence and the order of each order, that is, the proposed research uses a model to obtain the top-down design summary. The research proposal is also characterized by the use of some important characteristics of texts such as the rank of sentence and the number of words in the text, the impact of each of them and the relationship between titles and sentences affecting the number of sentences in one text. The research proposal is characterized by the ability to find abstract texts for one text rather than several texts. The research proposal is also characterized by the use of some important characteristics of texts such as the rank of sentence, the number of words in the text and the impact of each of them and the relationship between titles and sentences affecting the number of sentences in one text. Moreover, the research proposal is characterized by the ability to produce summarization for one text rather than several texts.

Table 2: Rules and Symbols of Active Concepts

R	symbol	Description
1	PP ↔ACT	Actor & Action
2	ACT↔PP	Action & Object
3	PP↔PP	Object & Object
4	PP←PP	Transpose PP
5	PP↔PA	Object Property
6	PP←PA	Property Object
	PA	ACT Source & ACT Recipient

7	ACT PP	
8	PP PA ↔ zz	Change in State
9	↔ yy ↔ zz	One Concept (action) affect another
10	↔ yy	Concepts with action time

$$r(s) = sf(d) + s(l) + N(wt) \quad (1)$$

Where r(s) is sentence rank and sf(d): sentence frequency in the text. S (L) is the sentence length and N (wt): number of words in the title and the processed sentence.
One of the factors influencing is the relationship between the number of words and the number of sentences in a single text. The use of conceptual dependency and the semantic representation of sentences contributes significantly in the production of the summarization. So, word score can be obtained using Equation [2]:

$$s(t) = \frac{tf}{\sum sf} \quad (2)$$

Where s (t) represent token score in the text, while is token frequency divided by sentences frequency in the text.

8. Methodology

8.1. Pre-processing

After assigning a texts of the target document, segment a processed sentences, the proposed research determine a stem tokens of the processed words via proposed anaphora resolution algorithm and indicate a list of the cue words in a text. Also, stemming which is a procedure of dropping the modulated forms of a token to a root form [16].

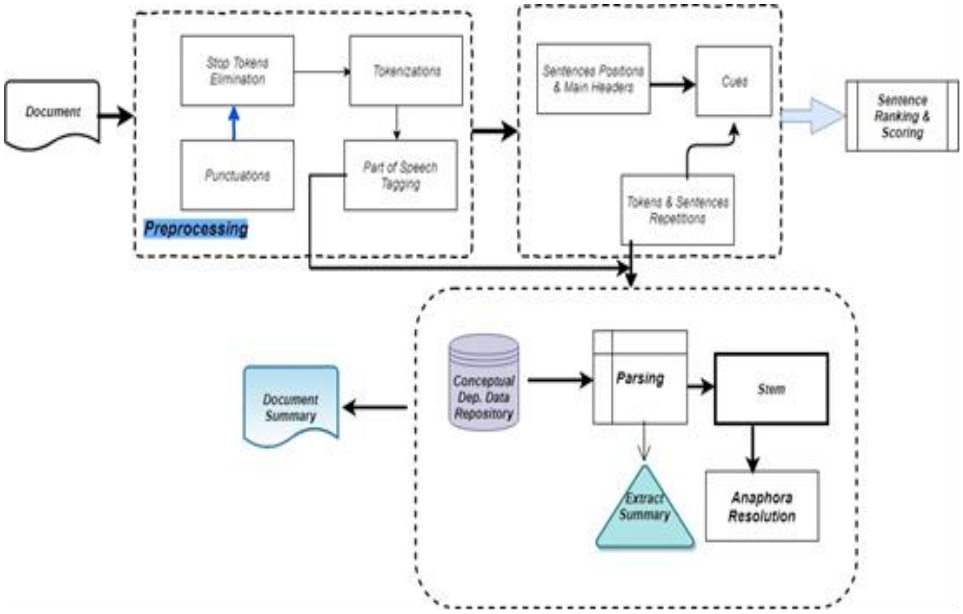


Fig. 1: Frame Work Architecture of the Proposed System.

8.2. Sentences scoring and ranking

The proposed research adopts a series of algorithms that start by determining a score for each sentence according to a series of features such as the length of the sentence, the number of each token, and the relationship between the title of the document and those tokens. Scoring sentences after identifying and processing each target document achieving tokens probabilities is the first algorithm of research cascading algorithms. Ranking sentences is the second algorithm. Exploiting the weight of each token in the sentence. Equation [3] depict the weight of each token:

$$\omega(t) = f_d(t) * log\frac{|D|}{f_D(t)} \quad (3)$$

Since the basic parameter used to determine the weight of any token is the frequency in the document that is identified by Equation [2], which depends basically on the repetition of the specific token in the document and that this measurement of the weight depends on the length of the text. It is noted that the increasing the size of the text affects that weight. So for the purpose of simplifying and adjusting these weights we use the following equation to adjusting this weight:

$$w_i = \frac{TF(t_j,d)}{\sqrt{\sum_j TF(t_j,d)^2}} \quad (4)$$

$$w_i = \frac{TF(t_j,d)\log(\frac{|D|}{DF(t_j)})}{\sqrt{\sum_j[TF(t_j,d)\log(\frac{|D|}{DF(t_j)})]^2}} \quad (5)$$

To setup sentence score and sentence rank after preprocessing texts of documents, specific algorithms for scoring and ranking was used. The seeds of sentences are tokens and above equations was used to adjust tokens weights.
The document length play an important role in affecting score and rank, therefore summarization of set of documents affecting scoring and ranking metrics. Hierarchical semantic conceptual dependency and parsing rules were be the second process to obtaining summarization after the following statistics algorithms.

Algorithm Sentence _Score (Sentence [id, m], Score [id, n])

$Pr[tokens] = \frac{f(tokens)}{N}$ ..... // Probability of tokens
L ← Sent (length)
k := 0
for i = 1 to L do
begin
k := k+1 ;
T[k] ← token[i]
for j := 1 to m do
begin
If T[k] ← (Ti[m] and (tf[i] ≥doc [L]/pr[tokens]))
score[id,i] ← scor[id,i]+1
end;
end.

Algorithm Sentence Ranking (Sentence_ score [id, n], rank [id, r])

$\omega(t) = f_d(t) * log \frac{|D|}{f_D(t)}$ ..... // weights tokens
s(t) ← $\omega(t)$ ..... // sentence score
L ← sent (length)
D_lenght ← tf(t) ..... // document Length
P[sent] ← 1 ..... // sentence position
i := 0 ;
While (i ≠ L) do begin
i := i+1 ;
$\Delta s(t) := \omega(t)[i]$; ..... // score increment
$s(t)_{new} := s(t)_{old} + \Delta s(t)$;
for j := 1 to m do
... // No. of sentences in target document
begin
If p[sent] = (id ∨ (D_length/2 ∨ d_lenght))
score[id,i] ← score[id,i]+1
end;
$score[id,i][id,i] + s(t)_{new}$
ank[id,r] ← score[id,i]
end.

8.3. Hierarchical semantics concepts

Using Conceptual Dependency for text summarization is an important and efficient tool in getting satisfactory and convincing results in this field due to its ability to draw the semantic network of the processed text via connecting the underlying concepts and giving its weight and rank at the connected edges. Describing the text as a hierarchical model give a comprehensive relations between tokens (Nouns, Verbs, adjectives, and adverbs). The hierarchical structure of tokens and concepts in terms of use of equations [1 to 5] gives a clear indication and real possibility of rapid access to summaries the general hierarchy of texts using Conceptual Dependency and the rules described earlier in Tables [1] and [2].

9. Results and discussion

One of the important features should be presented in summarization text is the consistent ratio of compression (τ) [15]. Assessment and evaluation of the proposed technique performed and applied with the following environment [1]:

9.1. Dataset

Summarized text not only can enhance text processing but also can help to recover and process text more efficiently [18]. The proposed technique applied to a specific Document Understanding Technique (DUC 2002). There are a number of different topics in this dataset, such that there are exists a number of documents of each topic. The accuracy of summarization can be detected via compared with the similarity of reference human similarity for the same dataset. Moreover, accuracy of output results compared with other previous techniques. In order to ensure the consistency and sobriety of the proposed technique in obtaining text summaries, we compare those results with the output results obtained using other previous techniques to the same data.
[1] Results depicted and appendix A, while appendix B presents the second target document which is tested by the system.
Assessment the results is a crucial phase in evaluating obtained results. Different approaches was used for this task. In this field, (Abdi, A. Idris, N, Alguliev, R, and Aliguliyev, R, July 2015) [1] propose a method to integrate the semantic relations between words and their syntactic composition then evaluate the results. ROUGE toolkit was used to test accuracy of the proposed technique ROUGE is the abbreviation of Recall Oriented Understanding for Gisting Evaluation. This metric is accuracy between the summarization obtained by applying proposed algorithm technique and reference summary generated by human for the source data. There are different ROUGE formulas dependent on frequency of the overlapping units as follows:
(1) ROUGE-N: For N-grams.
(2) ROUGE-L: For L-grams
(*) ROUGE-L: It is metric for measuring accuracy of machine generated summary. It is reference summary between two summaries A and B with different lengths (Saleh, H., & Kadhim, 2015):
N-gram counts the number of N-grams of the above two summaries which is depicted by equation [9] (Saleh, H., & Kadhim, 2015):
ROUGE- N =

$$\frac{\sum S_E(\text{refernce summaries}) \sum N\text{-gram} \epsilon_s Count_{match}(N\text{-Gram})}{\sum S_E(\text{refernce summaries}) \sum N\text{-gram} \epsilon_s Count(N\text{-Gram})} \quad (6)$$

$$R_{lcs} = \frac{lcs(A,B)}{N} \quad (7)$$

Where N: is the length of reference summary.

$$\rho_{lcs} = \frac{lcs(A,B)}{M} \quad (8)$$

Where M: is the length of candidate summary.

$$F_{lcs} = \frac{(1+\beta^2) R_{lcs} \rho_{lcs}}{R_{lcs} + \beta^2 \rho_{lcs}} \text{ [1]} \quad (9)$$

[1]Scores of each sentences are produced via multiplying the value of each feature for a sentence as shown in equation (10) [17]:

$Score[s] = \Pi f_i$ where $f_i \in F$ (10)

9.2. Results

Table [3] shows the accuracy of applying the proposed technique for ROUGE-2 and ROUGE-L to documents d061j to d070f. The output result obtained using specific hardware resources. The results reflects that similarity values between output summary via applying proposed Hybrid Technique and Human reference summary for the target data are more suitable and accurate to be adapted for documents summarization. Hardware Resources used to get this output was Laptop Dell CORE "I3" with RAM 4 GB. Java software was the environment platform for this output. Output results compared with previous related algorithms shown in section [2] ref [10] in this paper for the same target documents.

Table 3: ROUGE-2 and ROUGE-L of the Conceptual Dependency Technique

Topic #	Model in [10]	Proposed Model ROUGE-2	Model in [10]	Proposed Model ROUGE-L
d061j	0.306	0.325	0.554	0.412
d062j	0.200	0.201	0.481	0.441
d063j	0.275	0.324	0.528	0.544
d064j	0.233	0.304	0.488	0.525
d065j	0.182	0.292	0.457	0.423
d066j	0.181	0.382	0.441	0.538
d067j	0.260	0.324	0.529	0.544
d068j	0.496	0.553	0.626	0.697
d069j	0.232	0.272	0.476	0.563
d070j	0.262	0.433	0.513	0.586

10. Conclusion

Different features was exploited to summarize text documents. New technique was used in this field. A Hierarchical conceptual dependency and rules with parsing shows that the proposed technique gives efficient results with a comprehensive summary. Results evaluation shows the accuracy of the resulted summary of this target documents. Future work can modify this technique to summarize Arabic Texts documents with enhancing and modification of preprocessing model depicted in section [3] to cover the different features of this language.

References

[1] Abdi, A. et al., Information Processing and Management, Automatic Summarization assessment through a combination of semantic and syntactic information of intelligent educational systems, July 2015, Volume 51, Issue 4, P340-358.

[2] Allahyari, P., ACM, Text Summarization Techniques: A Brief Survey, arXiv 28 July 2017.

[3] Blanco, R., Matthews M, and Mika P, Information Processing and Management, Ranking of Daily

[4] Codina-Filbà J, Bouayad-Agha J, Burga, Gerard Casamayor A, and Wanner L, Information Processing and Management, Using genre – specific features for patent summaries, January 2017, Volume 53 Issue Issue 1, pages 151- 174.

[5] Hovy, E., "Text Summarization", Book: chapter 32, p584-595, 2000.

[6] Lal, P., Project, Text Summarization, June 13, 2002.

[7] Lloret, E., Text Summarization: An Overview, Alicante, the Spanish Government under the project TEXT-MESSS elloret@dlsi.ua.e,2008.

[8] McBurney, W., "Automatic Source Code Summarization of Context for Java Methods", IEEE Transactions on Software Engineering, 2015.

[9] Rossi, E., Centroid-based Text Summarization through Compositionality of Word Embedding" , Department of Computer Science University of Bari, 70125 Bari, Italy, Proceedings of the Multilingual 2017, Workshop on Summarization and Summary Evaluation Across Source Types and Genres, pages 12–21, Valencia, Spain, April 3, 2017.

[10] Saleh, H., & Kadhim, Genetic Based Optimization Models for Enhancing Multi Document Text Summarization, Computer Science Department: University of Technology-Baghdad, 2015.

[11] Sarraf, P., "Summarization of Document using Java|", International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 2, February – 2014.

[12] Sankar, k., "An Approach to Text Summarization", AU-KBC Research Centre MIT Campus, Anna University MIT Campus, Anna University, Chennai- 44. Chennai- 44,sankar@au-kbc.org, Proceedings of CLIAWS3, Third International Cross Lingual Information Access Workshop, pages 53–60, Boulder, Colorado, June 2009.

[13] Sood, A., "Towards Summarization of Written Text Conversations ", International Institute of Information Technology, June 2013.

[14] Satyan, S., "Automatic Text Summarization", International Journal of IT, Engineering and Applied Sciences Research (IJIEASR) ISSN: 2319-4413 Volume 4, No. 4, April 2015.

[15] Stefan., C. et. al, " On Definition of Automatic Text Summarization ", Proceedings of Second International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC2015), Dubai, UAE, 2015.

[16] Sizov, G., " Extraction-Based Automatic Summarization" Norwegian University of Science and Technology" June 2010.

[17] Shubham, A." Automatic Text Summarization", INDIAN INSTITUTE OF TECHNOLOGY MANDI JUNE, 2015.

[18] YANG, G." Contextual Text Summarization for Content Processing in Mobile Learning ", Publications of the University of Eastern Finland Dissertations in Forestry and Natural Sciences No 150, 2014.

Appendix A

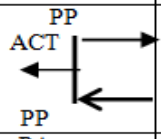
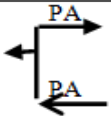
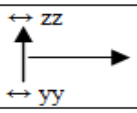
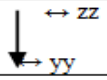
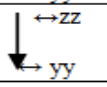
Reference _Summary _d085da_200

Killer Storm Hits South Carolina Coast

Step 21 (midnight) _ Hugo crashes into Charleston, S.C, with winds of 135 mph.
Leveling buildings and flooding streets.
Ten people in the Carolinas are killed.
Before the hurricane warning system was established, a storm of Hugo's power would have coast thousands of lives.
Forecasters at the National Hurricane Center used computer models to track Hugo's path into Charleston, S.C.
To determine the track of the storm, the Forecasters analyze supercomputer predictions, satellite data, the history of similar storms and the current path of the hurricane.
Several thousands of people who were in the shelters and the tens of thousands of people who evacuated inland were potential victims of injury and death."
Although claims adjusters are still tallying the losses, there is now little doubt that overall damage to insured property will far exceed the record $752.5 million in claims paid by the industry after Hurricane Fredric struck Mississippi and Alabama in 1979, Parsons said.
But if the seas warm up, say by 3 to 6 degrees Fahrenheit, the wind speeds might be 25 percent higher which means the destructive force would be more like 50 percent higher, he said.

Extracted Summary of Reference Summary _d085da_200

Implementing proposed algorithm (Proposed preprocessing, weighing, scoring, and ranking sentences) with conceptual dependency, extracting summary will be obtained and evaluated.
The following is a result of CD and extracted summary for Reference Summary _ d085_200.

R	symbol	Description
1	PP ↔ACT	Killer Storm & Hits
2	ACT↔PP	Hits & South Carolina
3	PP↔PP	People & Coast
4	PP←PP	Transpose killing
5	PP↔PA	Killing & Victims
6	PP←PA	Injured People
7		Forecasters Analyze & Prediction current path of the hurricane
8		Injury, Death, and Losses
9		Killing & Victims
10		Wind Speed & Destructive Force
11		Seas warm up by 3 to 6 degree

Extracted Summary of Reference _Summary _d085da_200

The killing Storm hits South Carolina Coast.

Thousands of People are victims: injured, dead, and losses.

Forecasters Analyze & Prediction current path of the hurricane.

Sea warm up by 3 to 6 degrees Fahrenheit.

97 percent of the buildings were damaged or destroyed

Appendix B

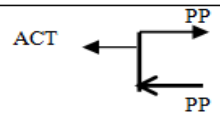
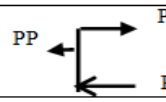
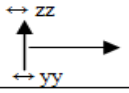
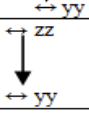
Reference _Summary _d085dh_200

Guadeloupe is devasted by winds measuring up to 150 mph. Early reports say 97 percent of the buildings were damaged or destroyed on St Croix, population 53000.

Sept 18 (daybreak) _Hugo crosses the northeast corner of Puerto Rico. Hurricane Hugo struck South Carolina with renewed fury Thursday after carry and fled inland on jammed highway. The storm, which caused billions in damage, claimed 17 lives in South Carolina, and only two were in the Charleston area, which bore the brunt of Hugo's 135 mph winds. Using the information from the satellite, supercomputers at the National Metrological Center in Suitland, Md, send information to the hurricane center where a tracking model constantly changes to account for current weather conditions and the position of the hurricane. The hurricane specialists were surprised by the last minute increase in wind speed, which was reported to them by Air Force reconnaissance.

The greenhouse effect my breed bigger and deadlier hurricanes in the future, storms up to 50 percent stronger than Hugo and last year's record setting Gilbert, some meteorologist say.

The two dominant home and automobile insures in the Carolinas said Tuesday that they expect to be hit with about $600 million in claims from Hurricane Hugo.

R	symbol	Description
1	PP ↔ACT	Winds 150 & Damaged
2	ACT↔PP	devasted & Guadeloupe
3	PP↔PP	Building & Dollars
4	PP←PP	Claimed 17 lives
5	PP↔PA	Damage & Cost
6	PP←PA	Damage Building
7		Information changes for current weather condition
8		$600 millions
9		Winds hit
10		Storm up to 50 percent than Hugo
11		Breed bigger

Extracted Summary of Reference _Summary d085dh_200

97 percent of the buildings were damaged or destroyed

breed bigger and deadlier hurricanes storms up to 50 percent stronger than Hugo and last year's

Hit with about $600 million in claims from Hurricane Hugo.

Changes to account for current weather conditions and the position of the hurricane.

Guadeloupe is devasted by winds measuring up to 150 mph.