

Automatic Summarization based on Field Coherent Passages

Mahmoud Rokaya

Taif University College of Computer & Information
Technology Taif, Saudi Arabia

Permanent Address: Tanta, University Faculty of Science, Egypt

ABSTRACT

Today the internet and WWW had caused an information explosion. Readers have no time to explore lengthy documents and passages. Text summarization is the process of getting a shorter version of a document that have the main points and important information contained in the original document. When this is done by means of a computer, i. e. automatically, it is called as Automatic Text Summarization. This work introduces a custom summary based on key words or field of interest that a user determine or select. Despite there are as many as methods of summarization but there are few methods of summarization which considered a custom summary that depends on the user need. Even the methods that tried to rely on the user keywords, they followed these key words blindly without any analysis of the actual relations between these key words or other related field association terms. This work presents improvements of the power link to get a more shorter summary. Also this method reduces the run time of the extraction algorithm.

General Terms

Natural Language Processing, Text Extraction, Text Summarization, Information Retrieval.

Keywords

Field Association Terms, Power Link, Field Coherent Passages, Test summary, Text extraction.

1. INTRODUCTION

The existence of the World Wide Web has caused an information explosion. Readers are overloaded with lengthy text documents where a shorter version would suffice. [31]

Text summarization is the process of distilling the most important information from a source to produce an abridged version for a particular user and task. [33]. When this is done by means of a computer, i. e. automatically, it is called as Automatic Text Summarization. Automatic text summarization is a technique where the text is input to the computer and it returns the clipped and concise extract of the original text and also sustains the overall meaning and main information content. [26]

Summarization can be classified into two approaches: extraction and abstraction. Extraction based summaries are produced by concatenating several sentences taken exactly as they appear in the texts being summarized. Abstraction based summaries are written to convey the main information in the input and may reuse phrases or clauses from it. [8], [10] and [33].

Extractive methods work by selecting a subset of existing words, phrases, or sentences in the original text to form the summary. In contrast, abstractive methods build an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a human might generate. Such a summary might contain words not explicitly present in the original text. The state-of-the-art shows that abstractive methods are still quite weak, so most research had focused on extractive methods. Two particular types of summarization often addressed in the literature are keyphrase extraction, where the goal is to select individual words or phrases to "tag" a document, and document summarization, where the goal is to select whole sentences to create a short passage summary. Recent researches have investigated types of summaries, methods to create them, and methods to evaluate them. [1], [2], [3], [4], [9], [12], [13], [14], [16], [20], [26], [27], [28], [29], [30], [32] and [33] are examples of efforts to develop methods of automatic summarization methods. Despite there are as many as methods of summarization but there are few methods of summarization which considered a custom summary that depends on the user need. Even the methods that tried to rely on the user keywords, they followed these key words blindly without any analysis of the actual relations between these key words or other related key words. [7], [11], [15] and [31] are examples of works that concentrated on evaluation of text extraction methods.

Humans can recognize the field by finding the specific words, these words called Field Association words (FA words). So it is more effective if the search engines could pick these words, FA terms, from the queries and use them as the bases of searching process. [23]. Automatic building of new FA term candidates using search engine was proposed by Atlam. [5]. Uddin et al. present a method for improvement of building FA term dictionary using passage retrieval [34]. Based on co-word analysis, Rokaya & Atlam presented a modification of the rules used to determine the FAts level [23]. Field association terms (FATs) dictionary concept was studied and tested for English. Arabic language has many differences from the English language so it needs special techniques for preprocessing before applying the power link algorithm. Rokaya and Nahla, depending on available FATs dictionary in English, proposed a multilingual FATs dictionary in English and Arabic [22].

Based on the concepts of Field Association Terms and Power Links, Rokaya et., al. presented a custom summary where the user insert the document which he would like to get a summary plus a key words that are used to pave a road that leads to a custom summary. In this method, the user inserts the document which he would like to get a summary. The

system uses these keywords to extract the related field association terms. The system uses these keywords and the extracted field association terms to extract and rank the passages in the document under consideration. [24]

The methods extracts the passages that are highly related to the field association terms. The results show that the length of the extracted text is still long and need refinements to be more shorter. Also the run time of the power link is not practical and the volume of the needed processes during the real time is not suitable for online summarization applications. This work presents improvements to the power link method to get a shorter versions of the summary. For each passage that the system decide to extract and append it to the summary the system ranks the sentences in each passage and decide whether to append the sentence to the passage or to delete the sentence from the passage. Also this work presents a new algorithm for passage scoring. In this work, a custom summary is introduced based on key words or field of interest that a user determine or select.

The remaining parts of this paper are organized as follows. A brief review of the concepts of field association terms and power links is presented in section 2 and section 3 respectively. The details of the proposed method is presented in section 4. Section 5 evaluate, compare and discuss the results of the proposed method. Conclusion is explored in section 6.

2. FIELD ASSOCIATION TERMS

It is natural for people to identify the field of document when they notice specific words. These specific words are referred as Field-Association terms (FA terms); specifically, they are words that allow us to recognize intuitively a field of text or field –coherent passage. Therefore, FA terms can be used to identify the field of a passage, and can be also used to classify different fields among passages. For these reasons FA terms can be used as a clue to identify a passage field. FA terms can be either words or phrases. [34]

Field association terms (FA terms) are the words that indicate each subject matter category in the classification scheme. [19]

A minimum term (or a word) is defined, as one which cannot be further divided without losing its semantic meaning, as a single FA TERM (single FA Term). Compound FA terms are defined to consist of two or more single FA terms. Both terms are expressed by enclosing them within quotation mark. A compound FA TERM is regarded as being single if it loses its field information when divided. Compound FA terms (e.g. nuclear weapon, consumption tax or global warming) are considered to be simple FA terms because document field information is easily lost when those compound terms are divided. So here, proper nouns (e.g. Atlanta Braves, Oakland Athletics and South Africa) are considered to be simple FA terms. Personal names (e.g. Sammy Sosa, Carlos-Delgado) are considered to be simple FA terms, but proper noun containing a title (e.g. Coach T. Lasorda) are divided into two single terms: “Coach” and “T. Lasorda”, which belong to the same document field Baseball but are on different levels. [5]

Field means a basic and common knowledge that can be used in human communication [18] and for convenience; hierarchical fields are categorized as Sub-Fields and Super-Fields. Therefore, Pitcher can relate to sub-field Baseball of super-field SPORTS and Pitcher may be classified SPORTS/Baseball. Selecting useful FA terms requires

consideration of relationships between simple and compound FA terms and field classification. [5].

3. THE POWER LINK

The term power link was proposed by Rokaya and Atlam as a method of building a dynamic field association terms dictionary. Power link algorithm presented a new rules to improve the quality of filed association terms (FATs) dictionary in English [21].

The origin of this concept comes from the co-word analysis researches. Co-word analysis considers the dynamics of science as a result of actor strategies. Changes in the content of a subject area are the combined effect of a large number of individual strategies. This technique should allow us in principle to identify the actors and explain the global dynamics [6].

If any two terms t_1 and t_2 belongs to a document D it is said that there is a link between t_1 and t_2 . The power of this link will be measured by the function $P_D(t_1, t_2)$ where:

$$P_D(t_1, t_2) = \frac{|D| \times cr(t_1, t_2)}{\text{average} \mathcal{L}(t_{1i}, t_{2j})_{i,j}} \quad (1)$$

where $|D|$ is the number of different terms in the document, $cr(t_1, t_2)$ is the co-occurrence frequency of t_1 and t_2 in the document D and $\text{average} \mathcal{L}(t_{1i}, t_{2j})_{i,j}$ represents the average distance between any instants t_{1i} and t_{2i} of the terms t_1 and t_2 in the document D . For more details see Rokaya and Atlam. [21]

To estimate the power link between two terms t_1 and t_2 over a given corps the function $P_C(t_1, t_2)$ is defined. This function can be defined as:

$$P_C(t_1, t_2) = \text{average}_{D \in \text{corps}} P_D(t_1, t_2) \quad (2)$$

This function states that the terms t_1 and t_2 will tend to appear nearer together if the value of this function reasonably high.

Let $P_S(w, D) = \text{average}_{f \in D} P(w, f)$, $f \in S$ be the power link

between the word w and the field $\langle S \rangle$ in the document D , if the set of field association terms that belong to the field $\langle S \rangle$ in the document D is empty then $P_S(w, D) = 0$. The power of link between a word w and the field $\langle S \rangle$ will given by

$$P(w, \langle S \rangle) = \text{average}_{D \in S} P_S(w, D) \quad (3)$$

4. THE METHOD

Fig. 1 illustrates the steps of the proposed method. The colors decides the main parts of the algorithm. The blue part presents the FT extraction process. In this part, the related FA terms to the keywords are extracted. The power link is used to produce those FA terms that have a strong relation to the given key words. This part is added to make the needs of the user to be more precise. The second part extracts the passages from a given document. The algorithm defines the boundaries of

passages and extracts them. Finally the third part extracts the related passages to the produced FA terms in part one. The extraction based on collecting the passages that have a high rank. For each sentence in the extracted passages that belong to any of the extracted passage a ranking procedure is developed. The ranking procedure gives a rank to a sentence based on location and content features. using a given threshold the sentences are extracted. The summary shows the sentences in the same order that they already have in the original document.

This approach is different from any other approach. In the beginning, it is far away from direct retrieval since the passage that contain the key words or even translated key words is not retrieved. Instead, the field association terms that have a strong relation or a high power link connection with respect to the field under interest are retrieved. Also, this approach is not a blind machine translation for the whole text or even to the passages that contain as many as of the key words. Instead, it retrieve the most connected passages to the field of interest through an intelligent guessing of the most related passages to the field. [24]

Rokaya et., al. presented a similar algorithm to extract passages related to the topics of the inserted keywords but the length of the extracted summary in many cases was very long. So, The current work aims to divide the extraction into two phases. The first phase extracts the passages and the second phase deletes the non effective sentences from each extracted passage.

4.1 FA Terms Extraction

It is natural for people to identify the field of document or a passage when they see specific words. These specific words is referred as field-association terms (FA terms). A minimum unit (or a word), is defined as one which cannot be further divided without losing its semantic meaning, as single FA term. For example the term "symmetric key" is a single FA term since if it divided into "symmetric" and "key" it will directly loss its semantic meaning and its relation to the field of cryptography.

For a given keywords, the algorithm computes the power link to each keyword with respect to each field in the corps. The algorithm picks the field with the largest power link for each keyword. In every field the algorithm extracts the FA terms that have a power link greater than α , where α is a given threshold. The union of all extracted FA terms and the keywords forms the terms that will be used in the next steps.

4.2 Total Scoring and Sentence Blending

The first characteristic of a topic flowing is defined as continuity and the second as transition. Passages with different field theme are delimited and the field duplication of passages is prevented. Every sentence is supposed to have one subject or less. The field that a sentence presents is called a theme field, which is denoted by F_{theme} [19].

F_{theme} is maintained by continuity or changed by transition through sentences. For a given sentence S that contains FA terms $(FA_1, FA_2, FA_3, \dots, FA_n)$ the power link between sentence S and all fields is computed according to the formula

$$P(S, F_j) = \sum_{i=1}^n P(FA_i, F_j) \text{ and the } F_{theme} \text{ for the sentence } S \text{ is determined by the field that gives } \max_j P(S, F_j). \text{ If the current}$$

sentence has two or more fields with the following property $|P(S, F_j) - P(S, F_k)| < \lambda$, where λ is small enough, then S is said to have no F_{theme} . Also if the set of FA terms contained in S is empty, S is said to have no field. If the current sentence S has the same F_{theme} as the previous sentence, or has no F_{theme} , or has no field the current sentence is appended to the same passage. And if the current sentence S has a different F_{theme} from the previous sentence, then S is delimited and a new passage begins. According to the previous rules the passages borders are determined and the passage extraction can be done. [24]

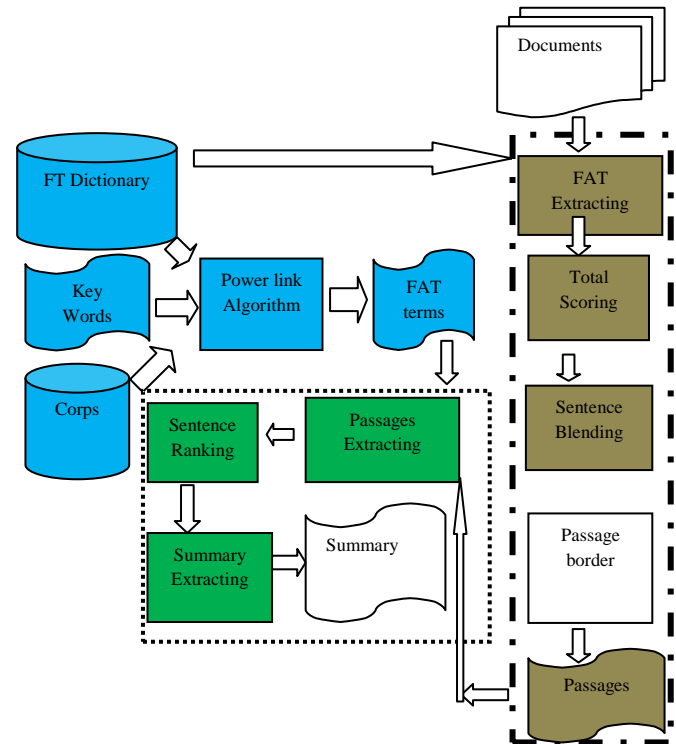


Fig.1. System configuration

4.3 Passages Extraction

Rokaya et., al. presented a method to decide passage delimitation based on power links calculations. [24] Despite this method was proved to give a high degree of accuracy the complexity of the method make a barrier against of applying it practically. Lee et al. presented an algorithm for passage retrieval based on field association terms levels. This algorithm can summarized as follows: [19]

Passage determination can be done through three processes, appearance, continuity and transition. The values of F_{theme} , α (continuity) and β (transition). F_{theme} is either maintained by continuity or changed by transition through sentences. If a specific FT appears in a document, its field is set to F_{theme} . Continuity is denoted in a sentence S_j with a current field F_{theme} as $\alpha_j(F_{theme})$. If $\alpha_j(F_{theme})$ is larger than threshold (α_{th}) or is greater than $\beta_j(F_k)$, a sentence is appended to the current passage and transition value is prepared with all possible fields except for F_{theme} by FTs that indicates a specific field. If a transition in k th appeared field at S_j which is denoted by $\beta_j(F_k)$, is larger than a continuity of

$\alpha_j(F_{theme})$, then the current field F_{theme} is replaced by F_k . If a topic transition, it is needed to introduce a delimiter at the sentence where the topic transition occurs. For example, if a transition happens in sentence S_j (i.e., F_{theme} has been continued and a new topic (F_k) appears in S_j , it is reasonable to decide to locate passage border between S_{j-1} and S_j . Even though text prior to S_{j-1} might have been on the same topic, some of the prior text may not have FTs for the topic. For this reason, a backtrack is made to set a location for the passage not have FTs for the passage border, namely, S'_j . The S'_j should be a sentence satisfying $\beta_j(F_{theme})=0$ because it can be considered that the field of F_{theme} continuous as long as the continuity of F_{theme} increases. The S'_j is a following sentence of the sentence whose continuity of F_{theme} starts to decline. In that case, sentences up to S'_j are grouped as a passage for the current field, and sentences from S'_{j+1} to S_j are grouped in a new passage to be transited as a field k. For more details see [19].

In this work, Lee method will be adopted Lee method to determine the passages. To prove the practicality of the method the time of execution for the proposed method and the method proposed by Rokaya et. al. [24] will be compared.

4.4 Passage Refinement

After passage extraction, the length of each passage can be reduced. If the sentence has no F_{theme} . The sentence has two or more fields with the property $|P(S, F_j) - P(S, F_k)| < \lambda$, where λ is small enough, or FA terms contained in S is empty. Then these sentences are dropped from the passages.

5. EVALUATION AND DISCUSSION

There are three sets of experiments should be done here

- (1) Experiments to evaluate the efficiency of delimiting the passages.
- (2) Experiments to evaluate the efficiency of the extractor
- (3) Experiments to evaluate the run time for the proposed method

5.1 Experiments to Evaluate the Efficiency of Delimiting Passages.

To estimate the quality of the presented passage delimitation method, The new method is compared to the method of Rokaya et. al. fifty articles composed artificially of some fields. each articles length were between 1 KB and 3 KB. Now the real field that each passage should belong is known. Two measures are used here, namely, precision and recall. Precision and recall are defined as follows. [24]

$$Precision = \frac{P_{accord}}{P_{output}} \text{ and } Recall = \frac{P_{accord}}{P_{answer}}$$

where P_{output} is the number of characters in passages that the system produces, P_{answer} is the number of characters in correct answer passage human decides, and P_{accord} are those occurred in both in the system output and in the correct answer passages. Fig. 2 and Fig. 3 presents the results of Rokaya method and Fig. 4 and 5 presents the results of the

proposed method. Namely Fig. 2 and 4 present results for Paccord, Poutput and Panswer. and Fig. 3 and 5 presents the results of recall and precision for Rokaya et., al. and the proposed method. The results for Rokaya et., al. show that the average value of precision is 0.78% and the average value of recall is 0.74%. These results proof that the this system gave a high accuracy. While the results of the proposed method show that the average value of precision is 0.82% and the average value of recall is 0.68%. These results proof that the proposed system gives an accuracy greater than Rokaya method for precision and lower for recall.

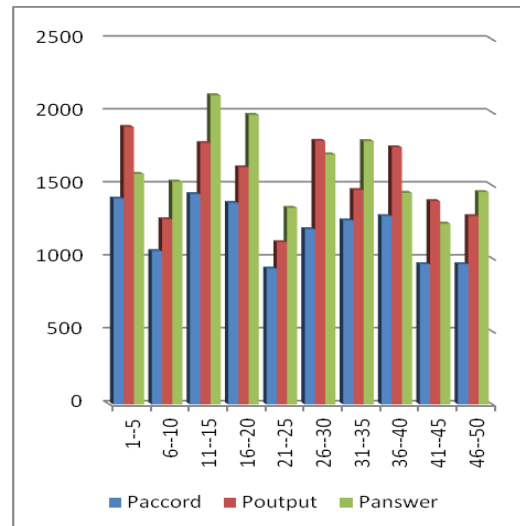


Fig.2 Results for Paccord, Poutput and Panswer for Rokaya, 2013 method

To ensure the strong of the results, F is also calculated using the formula.

$$F = \frac{2 * Precision * Recall}{Precision + Recall}$$

The average value of F measure for both methods is 75% which reflects a high performance of the algorithms and proves that there is no significant difference between the two methods.

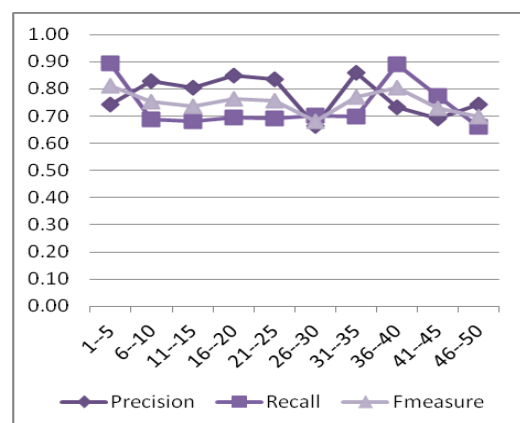


Fig.3: Precision and Recall results for Rokaya 2013

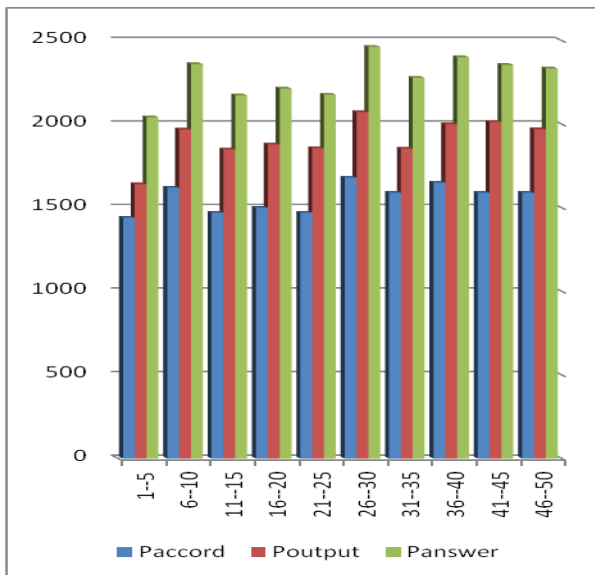


Fig.4 Results for Paccord, Poutput and Panswer for the proposed methods

5.2 Experiments to Evaluate the Efficiency of the Extractor

Since the goal of summarization schemes is to automate a process that has traditionally been done manually, a comparison of automatically generated extracts with those produced by humans would provide a reasonable evaluation of these methods. [31]

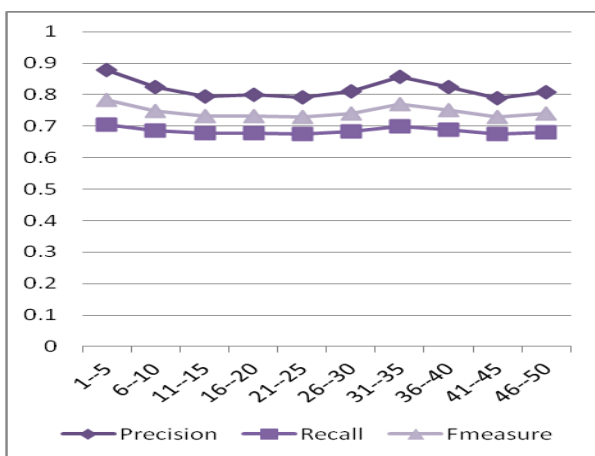


Fig. 5. Precision and Recall results for the proposed method

Assuming that a human would be able to identify the passages beginnings and ends and also would be able to identify the most important passages effectively. If the set of passages selected by an automatic extraction has a high overlap with the human generated extract, the automatic method should be regarded as effective. In all experiments, two manually generated extracts are generated. In both cases the user is asked to extract a specific number of passages, delimited using our method, based on a given keywords for each document.

The evaluation takes the following form,

- A user walks to the system and picks a document for a custom summary based on a given keywords

- the summary based on the same keywords and presents it to the user. The user compares this summary to his own one.
- In the second case, the system generates the summary using our proposed method to get the automatic one and the user compares this summary to his own.

In both cases, the user satisfaction is measured by the number of common passage between his own summary and the target summary (manually or automatically extracted).

For more strict comparison, two other automatic summarization are considered:

- A random summarization is done by randomly extracting the required number of passages from all passages that contains at least one of the given keywords. Any extraction algorithm must do better than the random method.
- Another method depends on extracting the first passage that contains at least one of the given keywords, then the program goes through the passages to pick the second passage that contains at least one of the given keywords. This process is repeated until the required number of passages are extracted.
- Fifty articles were chosen for the evaluation process. It is known that the overlapping between two manual free extracts, free extracts are those that do not depend on any given keywords, is lower than 50 %. [17] and [25].

This can be understood as follows, each human has interests that may differ from the other human. So the lower overlapping between the two manual extracted summaries reflects the lower overlapping between their needs. But if the needs were unified, it is expected that percentage of overlapping will be increased. The overlapping between the two manual extracts were 85.7 % which were expected. Table 1, summarizes the results of the experiments. The overlapping between the automatic extracts and the manual extracts. From this table the automatic summary do better than the random and the initial methods. But the results of the initial summary were extremely unexpected. The overlapping between the initial method and the manual extracts were 80.42 % and 74.67%. This high result were very strange. But the nature of the human is to not go deeply through the documents to get the required summary. In most cases the user will add the passage whenever he found some of the keywords. He will repeat this process to get the required number of passage more than interesting to get most of the important passage related to the keywords. Also since he was advised to complete the reading to the end of each documents, he might found some more important passage and will replace them in the place of earlier added passages. Finally the results for the power link method were 87.92% and 75.26%. The average is 81.59%. The results for the proposed method were 88.88% and 76.17%. The average is 82.53%. These results proof that the proposed method is fine and gave a considered performance than the initial method. Despite the difference between the proposed method the power link method is not high, the proposed method does as much the power link can do. The proposed method presents a more shorter summary through deleting non important sentences. Also, the next part presents evidence that the proposed method has a run time lower than the power link method in all cases.

5.3 Experiments to Evaluate the Run Time for the Proposed Method

In this experiment, the 50 articles are grouped in groups randomly, each of 5 articles, and run the two algorithms, power link algorithm and the proposed algorithm, on the same machine. Fig. 6 gives the run time results of both algorithms in milliseconds the average time for the power link = 4206 milliseconds and the average run time for the proposed method =2036 milliseconds, the average difference = 2171 milliseconds. This means that the proposed method improved the run time by 48%. This proof the expected assumption that the power link will not be practically as much as needed and insures that the proposed method put a stone in the way of presenting a more practical method with a high performance.

Table1: Evaluation measures for automatic extraction

Overlap between manual extracts: 85.7 %

Algorithm	S1	S2	Average
Power link	87.92609	75.26087	81.59348
Random	57.74783	48.36087	53.05435
Intial	89.42283	74.67717	82.05
Passage Coherent	88.88609	76.17087	82.52848

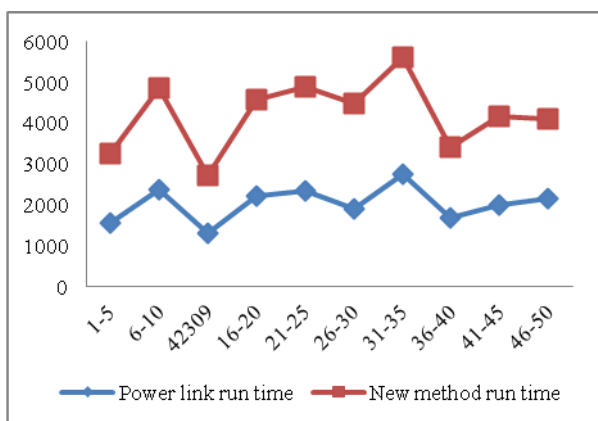


Fig.6. Run Time Results

6. CONCLUSION

In this work a new method of text summarization is presented. This method is an improvement of the power link summarization algorithm. The results showed that the automatic extraction based on the proposed method gave the same results as the power links which is better than the random and the initial lead methods. The proposed method presented an algorithm which produced a more shorter summary than the power links. Also the proposed method presented a new procedure for scoring passages. This procedure reduced the run time of the whole algorithm.

7. REFERENCES

[1] Alkesh Patel, Tanveer Siddiqui, And U. S. Tiwary, a Language Independent Approach to Multilingual Text Summarization. in Large Scale Semantic Access to Content (Text, Image, Video, and Sound), Riao '07, Pp 123–132. 2007

[2] Aparna Ladekar, Archana Mujumdar, Prajakta Nipane, Sonam Titar, Guide: Mrs. Kavitha S. / International Journal of Engineering Research and Applications (Ijera)

Issn: 2248-9622 Wwww.Ijera.Com Vol. 2, Issue 2, Pp.1551-1555, Mar-Apr 2012.

[3] Arman Kiani -B, M. R. Akbarzadeh -T, Automatic Text Summarization Using: Hybrid Fuzzy Ga-Gp, 2006 Ieee International Conference on Fuzzy Systems Sheraton Vancouver Wall Centre Hotel, Vancouver, Bc, Canada July 16-21, 977-983, 2006.

[4] Atefeh Farzindar And Guy Lapalme. Letsum, a Text Summarization System in Law Field. the Face of Text Conference (Computer Assisted Text Analysis in The Humanities), P. 27-36, McMaster University, Hamilton, Ontario, Canada, Nov 2004.

[5] Atlam E., Morita, K., Fuketa, M., & Aoe, J., a New Method For Selecting English Field Association Terms of Compound Words and its Knowledge Representation. Information Processing and Management, Vol. 38, Pp. 807-821, 2002.

[6] Callon, M., Courtid J., & Ladle, F, Co-Word Analysis As A Tool For Describing The Network Of Interactions Between Basic And Technological Research: The Case Of Polymer Chemistry. Science Metrics, Vol. 22(1), Pp. 155-205, 1991.

[7] C.-Y. Lin, E. Hovy, "From Single to Multi-Document Summarization: a Prototype System and its Evaluation", in "Proceedings Of The Acl", Pp. 457–464, 2002

[8] Chin-Yew Lin. Automated Text Summarization. in Tutorial of The 2nd International Joint Conference on Natural Language Processing (Ijcnlp-05), Jeju Island, Korea, October 11 - 13, 2005.

[9] Constantin Or`Asan, an Evolutionary Approach for Improving the Quality of Automatic Summaries, Multisumqa '03 Proceedings of the Acl 2003 Workshop on Multilingual Summarization and Question Answering - Volume 12, Pages 37-45, 2003

[10] D. Marcu, L. Carlson, And M. Watanabe. the Automatic Translation of Discourse Structures. in Proceedings Of Naacl, 2000.

[11] Dipanjan Das, André F.T. Martins. "a Survey on Automatic Text Summarization." Literature Survey for The Language and Statistics Course at Cmu, November, 2007.

[12] Dragomir R. Radev, Eduard H. Hovy, Kathleen Mckeown: Introduction to the Special Issue on Summarization. Computational Linguistics 28(4): 399-408, 2002.

[13] E. Hovy, C. Y. Lin, Automated Text Summarization in Summarist. in Proceeding of Acl'1997/Eacl'1997 Workshop on Intelligent Scalable Text Summarization. Pp 18-24, 1997

[14] Elena Lloret, Topic Detection and Segmentation in Automatic Text Summarization December 13, this Work is Protected under the Creative Commons License"Attribution, Wwww.Dlsi.Ua.Es, 2009.

[15] Eugene Santos Jr., Ahmed A. Mohamed And Qunhua Zhao, Automatic Evaluation of Summaries Using Document Graphs, Workshop at Acl 2004 (42nd Annual Meeting of the Association for Computational Linguistics), 2004.

- [16] Feng Jin , Minlie Huang, Zhiyong Lu, Xiaoyan Zhu, Towards Automatic Generation of Gene Summary, Proceedings of the Workshop on Bionlp, Pages 97–105, Boulder, Colorado, June 2009. Association for Computational Linguistics, 2009
- [17] Gerard Salton and Christopher Buckley, Term-Weighting Approaches in Automatic Text Retrieval, Information Processing & Management Vol. 24, No. 5, Pp. 513-523, 1988
- [18] Kawabe, K., & Matsumoto, Y., Acquisition of Normal Lexical Knowledge Based on Basic Level Category. Information Processing Society of Japan, Sig Note, Vol. 125(9), Pp. 87–92, 1998.
- [19] Lee, S., Shishibori, M., Sumitomo, S., Aoe, J., Extraction of Field-Coherent Passages. Information Processing And Management, Vol. 38, Pp. 173-207, 2002.
- [20] Lei Yu; Mengge Liu; Fuji Ren; Shingo Kuroiwa, a Chinese Automatic Text Summarization System for Mobile Devices, Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation, Pp 426-429, 2009.
- [21] Mahmoud Rokaya, Elsayed Atlam, Masao Fuketa, Tshering C. Dorji And Jun-Ichi Aoe, Ranking of Field Association Terms Using Co-Word Analysis, Information Processing & Management, Volume 44, Issue 2, March, Pages 738-755, 2008.
- [22] Mahmoud Rokaya And Abdallah Nahla. Building A Multi-Lingual Field Association Terms Dictionary, International Journal of Computer Science and Network Security, Vol. 11 No. 3 Pp. 208-213, 2011.
- [23] Mahmoud Rokaya and Atlam, E-S., Building of Field Association Terms Based on Links', Int. J. Computer Applications in Technology, Vol. 38, No. 4, Pp.298–305, 2010
- [24] Mahmoud Rokaya, Sultan Aljahdali, Dalia Hemdan, Automatic Text Extraction Based on Field Association Terms and Power Links, 26th International Conference on Computer Applications in Industry and Engineering (CAINE-2013), September 25{27, Los Angeles, 2013.
- [25] Mandar Mitra, Amit Singhal, Chris Buckley , Automatic Text Summarization by Passage Extraction, Workshop on Intelligent Scalable Text Summarization, 1997.
- [26] Manisha Prabhakar*, Nidhi Chandra, Automatic Text Summarization Based on Pragmatic Analysis, International Journal of Scientific and Research Publications, Volume 2, Issue 5, May 2012
- [27] Martin Hassel, Resource Lean and Portable Automatic Text Summarization, Doctoral Thesis, Kth School of Computer Science and Communication, Stockholm, Sweden 2007.
- [28] Massih-Reza Amini, Patrick Gallinari, Self-Supervised Learning for Automatic Text, Summarization by Text-Span Extraction 23rd Bcs European Annual Colloquium on Information Retrieval, Pp 1-9, 2001.
- [29] Mohsen Pourvali And Mohammad Saniee Abadeh ,Automated Text Summarization Base on Lexicales Chain and Graph Using of Wordnet And Wikipedia Knowledge Base, Ijcsi International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012.
- [30] Nitin Madnani, David Zajic, Bonnie Dorr, Necip Fazil Ayan & Jimmy Lin, Multiple Alternative Sentence Compressions for Automatic Text Summarization, in Proceedings of the 2007 Document Understanding Conference (Duc-2007) at Nlt/NaacL 2007, April 2007, Rochester, New York, 2007.
- [31] Oi Mean Foong1, Alan Oxley1 And Suziah Sulaiman1, Challenges and Trends of Automatic Text Summarization, International Journal of Information and Telecommunication Technology, Vol. 1, Issue 1, Pp 34-39, 2010.
- [32] Seonggi Ryang; Takeshi Abekawa Framework of Automatic Text Summarization Using Reinforcement Learning, Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Pages 256–265, Jeju Island, Korea, 12–14 July 2012. C 2012 Association for Computational Linguistics, Pp 256-265, 2012.
- [33] Vishal Gupta And Gurpreet Singh Lehal, "a Survey of Text Summarization Extractive Techniques", Journal of Emerging Technologies in Web Intelligence, Vol. 2, No. 3, August 2010.
- [34] Uddin, S. Elmarhomy G., Atlam, E., Fuketa, M., Morita K., Aoe J., Improvement of Automatic Building Field Association Term Dictionary Using Passage Retrieval. Information Processing & Management Journal, Vol. 43, Pp. 1793–1807, 2007.

ACKNOWLEDGMENTS

The author gratefully acknowledges the Taif University for the financial support.