Int. J. Advance. Soft Comput. Appl., Vol. 2, No. 1, March 2010
ISSN 2074-8523; Copyright © ICSRS Publication, 2010
www.i-csrs.org

Extraction Based Multi Document Summarization using Single Document Summary Cluster

Shanmugasundaram Hariharan

Department of Information Technology,
School of Computer and Information Sciences
B.S.Abdur Rahman University
Vandalur, Chennai-48, Tamilnadu, India.
email:mailtos.hariharan@gmail.com

Abstract

Multi document summarization has very great impact among research community, ever since the growth of online information and availability. Selecting most important sentences from such huge repository of data is quiet tricky and challenging task. While multi document poses some additional overhead in sentence selection, generating summaries for each individual documents and merging the sentences in a coherent order would greater strength. The proposed approach was competitively better as compared to state of MEAD summarizer at focused compression ratios. This paper focus on three different studies namely i. To find the performance of multi document summarizer from single document cluster (using MEAD) ii. Comparison of our approach with MEAD performance for the dataset considered iii. To extract sentences for multi document summarization at 30% compression rate to obtain 100% efficiency using 7-point summary sheet. Investigation carried out from an average of 22 documents shows that our system is promising.

Keywords: *single document summarization, sentence extraction, multi document summarization, MEAD.*

1 Introduction

Summarization is a reductive transformation of source text to summary text through content reduction by selection and/or generation on what is important in source text [9]. Summarizing documents of all kinds of information is continually increasing and it is continued to be a steady subject of research over decades [1]. This process of automatic summarization deals with preprocessing documents, evaluating the importance of sentences, generating summaries, evaluating summarization, and so on.

Multiple documents summarization produces summary from multiple documents instead of a single ones. It can be viewed as either as an extension of single document summarization of a collection of documents covering the same topic, or information extracted from several sources. Multi document summarization differs from single document summarization with the following ways: degree of redundancy, temporal dimension, compression ratio and co-reference problem [10].

A variety of multi-document summarization methods have been developed recently. Generally speaking, those methods can be either extractive summarization or abstractive summarization. Extractive summarization involves assigning saliency scores to some units (e.g. sentences, paragraphs) of the documents and extracting those with highest scores, while abstractive summarization usually needs information fusion, sentence compression and reformulation. Our work focuses on extractive summarization.

The major challenge in multi-document summarization is that a document set may contain diverse information, which is either related or unrelated to the main central topic, and hence we need effective summarization methods to analyze and extract the important information. Additionally these information overlaps with each other, hence we need effective merging techniques to build summary. In order to present the summary readable and inter–related with other sentences, function of cohesion is studied [2]. Cohesion relates part of a text to another part of the same text. Consequently it lends continuity to the text by providing this kind of text continuity. It also enables the reader or listener to ensure continuity in reading the document.

The above issues necessitate the need to investigate multi document summarization. In order that effective summaries are to be built from multi document clusters, there exist two different approaches. The first approach extracts sentences from multi document clusters, while the next approach is to merge sentences extracted by single document approach. Consider an example to illustrate the need or importance of the proposed investigations. If a cluster C1 has 10 documents and each document having 10 sentences. If 10% compression ratio is applied, then the user needs to pick up 10 sentences (out of 100) from the cluster set. On analyzing the performance of such approach, it is found that summarizer tends to select sentences biased towards a document and tends to be repetitive. Hence this paper addresses this issue effectively to form multi document summary set from single document summaries. Also studies were made on the compression rates.

This paper is organized as follows. Section 2 presents related research works carried out in automatic summarization. In the section 3, we discuss the working of popular MEAD multi document summarizer developed by Radev at al 12]. Section 4 and 5 discuss the experimental results and conclusion & future works respectively.

2 Related Works

Mihalcea and Tarau [11] have proposed a language independent extractive summarization that relies on iterative graph-based ranking algorithms. The authors have obtained single-document extractive summary, with respect to the importance for the overall understanding of the text. A graph is being constructed by adding a vertex for each sentence in the text, and edges between vertices are established using sentence inter-connections. After the ranking algorithm is run on the graph, sentences are sorted in reversed order of their score, and the top ranked sentences are selected for inclusion in the extractive summary. Multi-document summaries are built using a “meta” summarization procedure. First, for each document in a given cluster of documents, a single document summary is generated using one of the graph-based ranking algorithms. Next, a “summary of summaries” is produced using the same or a different ranking algorithm.

Huang et al. [3] have proposed a method to extract key sentences of a document as its summary by estimating the relevance of sentences through the use of fuzzy-rough sets. By using senses rather than raw words, sentences of the same or similar semantic meaning but written in synonyms are treated differently and to extract key sentences as a summary of a document. After all words in a sentence are disambiguated, sense representations for the sentence in terms of WordNet senses is built to indicate the concept of the sentence.

Liu et al. [4] proposes a strategy for Chinese multi-document summarization based on clustering and sentence extraction. They have adopted term vector to represent the linguistic unit in Chinese document, which obtains higher representation quality than traditional word-based vector space model in a certain extent. The authors have also explained the basic problems involved for summarizing Chinese documents namely representation of sentence in vector space model (VSM), number of clusters appropriate for the sentences in the documents collection, selecting representative sentences from the clusters and evaluating the summary quality.

Chen et al. [5] have performed investigations using lexical chains as a model of multiple documents in Chinese language to generate an indicative, moderately fluent summary based on the HowNet knowledge database Based on an analysis of semanteme, the algorithm removes redundant similarities and retain differences in information content among multiple documents. After pre-processesing, the next step is construction of lexical chains. Then significant sentences are extracted from each document, ordered and redundant information are recognized and removed. Finally, the summary is generated in chronological order.

Yong-dong et al. [6] proposed Multi-document Rhetorical Structure (MRS) for summarization task. This structure simultaneously represents multiple relationships of different text units at different levels including rhetorical relationships, semantic relationships and temporal relationships. MRS is a three-dimensional structure that can be used to simultaneously present all documents in set of multi-document and can represent text simultaneously at different levels of granularity (including sentences, paragraphs, sections and documents).

Qiu et al. [7] have studied the application of document sets categorization and different features. These features for summarization include machine learning to derive weights or a set of rules for combining the features and the other is empirical estimate to determine the weights of different features. The authors have improved extractive summarizer, which analyzes the document sets and decides the categories of the documents using different summarization strategies.

Jun et al. [8] designed a summarization system based on two-step sentence extraction as it combines statistical methods and reducing noisy data through two steps efficiently. In the first step, the system estimates the importance score of bi-gram pseudo sentences by the combination of Title & Location methods and then it removes invaluable bi-gram pseudo sentences which are called as noisy data. In the second step, method separates the bi-gram pseudo sentences into each original single sentence and it performs second sentence extraction by adding Aggregation Similarity method to the linear combination of the first step.

Shanmugasundaram Hariharan [18] have studied the effects of merging on multi document text summaries Issues on merging two or more similar documents or summaries for multi document text summarization were investigated comprehensively. Important sentences extracted from multiple related sources are merged to form a consolidated summary there by producing coherent and non-repetitive summaries.

3 Centroid-Based Summarization (CBS)

The technique used for multi-document summarization by Dragomir et al. [12] is centroid-based summarization (CBS). CBS uses the centroids of the clusters produced by TDT to identify sentences central to the topic of the entire cluster. CBS is implemented in MEAD, which is publicly available multi-document summarizer [15]. A key feature of MEAD is its use of cluster centroids, which consist of words that are central not only to one article in a cluster, but also to all the articles.

3.1 MEAD Extraction algorithm

MEAD compresses a cluster of topically related documents into a summary of the user's desired length. As a first step, three features namely centroid score, position, and overlap with first sentence (which may happen to be the title of a document) [13] is calculated:

- Centroid score- measure of the centrality of a sentence to the overall topic of a cluster (or document in the case of a single-document cluster).
- Position score- decreases linearly as the sentence gets farther from the beginning of a document.
- First sentence overlap score - which is the inner product of the TF*IDF-weighted vector representations of a given sentence and the first sentence (or title, if there is one) of the document).

As next step, the sentences are ranked according to their combined score which is a linear combination of all the sentence features used. MEAD uses a cosine similarity metric to compare each candidate sentence (for inclusion in the summary) to each higher-ranking sentence. If the candidate sentence is too similar to the specified threshold [14], it is penalized and is not included in the summary. Finally, the top remaining n-percent of the sentences (with the compression rate 'n' being determined by the user), are returned to the user as the summary. The main contribution of MEAD is explained in section 3.2 and 3.3 shortly.

3.2 Cluster-Based Relative Utility (CBRU)

Cluster-based relative utility (CBRU, or relative utility, RU in short) refers to the degree of relevance (from 0 to 10) of a particular sentence to the general topic of the entire. A utility of 0 means that the sentence is not relevant to the cluster and a 10 marks an essential sentence.[12]

3.3 Cross-Sentence Informational Subsumption (CSIS)

A related notion to RU is cross-sentence informational subsumption (CSIS, or subsumption). CSIS reflects that certain sentences repeat some of the information present in other sentences and therefore omitted during summarization. If the information content of sentence a (denoted as i(a)) is contained within sentence b, then a becomes informationally redundant and the content of b is said to subsume that of a:

$$i(a)<i(b)$$

3.4 Utility-based evaluation of both single and multiple document summaries

The interjudge agreement measures to what extent each judge satisfies the utility of the other judges by picking the right sentences. The author has also calculated mean cross-judge agreement (J), Random performance (R), System performance (S), Normalized system performance (D) to incorporate CSIS.

4. Experimental Setup

4.1 Corpus description:

Our data corpus consists of news documents collected from commercially available news service providers like Google News, Hindu, Indian Express, Deccan Herald and other news services [16]. Each document includes the title, timestamp. In order that target set to be achieved, we have created an ideal summary (explained in next section), for evaluating our results. We present the study results corresponding to a 22 document clusters. Each cluster consists of two documents.

The main features of the document corpus are:

a. The document has minimum of 8 sentences and maximum of 17 sentences.
b. The total number of sentences in the corpus is 399.

However we have not focused on clustering of document through automated techniques like k-means or other algorithms, instead clusters were formed manually. Such automated clustering approach is left for future work.

4.2 Ideal summary preparation:

Evaluation is a crucial step for both single and multi-document summarizations. Evaluation is generally categorized into two major categories as intrinsic and extrinsic modes of evaluation (Mani and Maybury, 1999). In intrinsic evaluation humans judge the quality of summary by directly analyzing it in terms of fluency, coverage or resemblance to manually constructed ideal summary. The second type of evaluation method is extrinsic, where the quality of summary is judged based on how it affects the completion of some other task. We stick on to the former method of evaluation by evaluating the automated summary with the human generated reference summary based on ranking of sentences by judges. Ranking involves assigning weights in terms of numerical scores based on the level of importance, by which the experts feel that the sentences should appear in summary and ordering the sentences in the descending order of weights.

As summary evaluation is a crucial task in evaluating summaries, we have focused in depth on some of the issues pertaining to target golden standard summary generation. Target set involves forming a cluster of sentences that matches the interest of majority of judges considered. For the experiment we have took 3 judges and we eliminate the set, for which no two judges agree. Agreement among judges summary or ranking is carried out using Kendall's and Spearman's rank correlation coefficients. We have also dealt some of the issues in the preparation of gold standard summaries like the number of judges, cut-off percentage for measuring the correlation [17]. Preparing such gold standard summary set is very crucial and indeed affects the summarizer performance.

To obtain a target set of ideal results, we distributed document sets to three judges and asked them to rank the sentences according to their importance. Their age group varies from 25 to 40 and all of them are postgraduates. We found that majority of the disagreement cases were pertaining to ranking lower order sentences. Table 1 shows the agreement among judges for 3, 2 and 1 sentence agreement respectively for 30% and 100% agreement. It is clearly inferred from Table 1 that there is close agreement at 30% compression ration, while the agreement decreases as the compression increases. Moreover it is seen that the agreement concerning all three judges is poor followed by 2 judge and single judge.

Table 1: Agreement among evaluators

Number of Judges agree on a sentence	Agreement at 30%	Agreement at 100%
3	25.89	8.59
2	40.47	44.36
1	33.55	47.04

4.3 Experimental Results and Analysis:

Multi document summary can be obtained from cluster of documents or from summary generated by single document cluster. While the former approach is quiet complex, we adopt the later approach. This section explains the study carried out in three different ways namely.

Study 1: Comparison of MEAD multi document summary from single document summary

Study 2: Comparison of MEAD and our approach based on summary generated from single document cluster.

4.3.1 Study 1: MEAD multi document generated from single document cluster:

The first study reported is based on generation of summary generation from multi document summary generation from single document cluster. For the data set used, experiments were carried out by generating multi document cluster from single document summary. For instance, if 20% compression ratio is focused, then we try to obtain single document summary cluster at 20% compression ration and then we try to merge both the single document summaries. Table 2 presents the results for the study1. It is inferred from the average of 22 document clusters that the system was able to pick up summaries effectively at 10% (yielding 100% accuracy). By accuracy, we mean that the number of sentences retrieved by the summarizer as picked up the judges. Hence from study1, we conclude that it is enough to generate summaries from single document clusters at an effective accuracy.

Table 2: Comparison of Multi document summaries with Single document summary produced by MEAD

Doc ID	**Compression Rate**		
	10%	**20%**	**30%**
MDS/C1	100	100	77.7
MDS/C2	100	100	60
MDS/C6	100	100	61.1
MDS/C8	100	100	100
MDS/C9	100	100	100
MDS/C10	100	100	100
MDS/C11	100	100	83
MDS/C12	100	100	100
MDS/C14	100	100	85.7
MDS/C15	100	80	57.1
MDS/C16	100	100	88
MDS/C19	100	100	100
MDS/C20	100	100	88
MDS/C21	100	100	100
MDS/C22	100	100	100
MDS/C39	100	100	100
MDS/C45	100	100	66.6
MDS/C46	100	100	100
MDS/C47	100	75	60
MDS/C48	100	100	75
MDS/C49	100	100	100
MDS/C50	100	100	71.4
Average	**100**	**98**	**85**

4.3.2 Study 2: Comparison of MEAD and our approach based on summary generated from single document cluster.

The summary generation process of MEAD is explained in detail in section 3.1, where in MEAD works on Centroid based approach. We proceed to explain the sequence of steps carried out during the summary generation process of our system. The results were then compared using the data set chosen for the study (discussed in section 4.1). The steps involved in the proposed approach are given below:

a. Pre-processing the documents:

Preprocessing of documents involves several steps, each of which is explained in subsections that follow.

i. Removal of stop words:

Stop words are frequently occurring, insignificant words that appear in a database record, article or web page. Stop words are an application dependent. They apply to the particular database or application (e.g.: searching, summarization). It is commonly assumed that words, which are not members of the noun-verb-adjective classes, should be on stop words lists. When a document is summarized by sentence extraction method we assign weights to all the keywords or tokens in the input document. The process of doing such stop word elimination results in better summary generation. Since we eliminate these stop words unwanted sentences would never climb higher up the order [9]. Single characters, common two-character and three-character words, frequently repeated words are typically included in the stop word list to maximize performance of summarization process.

ii. Applying Porter Stemming algorithm:

Truncation, also called stemming, is a technique that allows us to search for various word endings and spellings simultaneously. Stemming algorithms [8] are used in many types of language processing and text analysis systems, and are also widely used in summarization, information retrieval and database search systems. A stemmer is a program determines a stem form of a given word. In other words, generates the morphological root of the word. Terms with a common stem will usually have similar meanings. For the example shown below the common root word is 'IMPROV'.

IMPROVE, IMPROVED, IMPROVEMENTS

The suffix stripping process will reduce the total number of terms in the IR system, and hence reduce the size and complexity of the data in the system, which is advantageous.

b. Generation of Single document summary:

Consider an example to illustrate the summary generation process. If there are two documents in a cluster say document A & B, each having 10 sentences each correspondingly. For 10%, 20% and 30% compression rates we have to pick up 2, 4 and 6 sentences respectively. Picking up these sentences from multi document cluster is challenging (since it has 20 sentences).

The sentence extraction algorithm generally applies statistical techniques to generate summary. In our extraction process, sentences are scored based on the term frequency. If the term matches the title words then special weight is given to those terms. We adopted the above approach of giving importance to title terms []. Note that sentences are scored after removal and stop words and stemming the samples. We have not focused features like bold, Italics, Uppercase letters features for special weights. An important aspect that is to be discussed is whether a document should have a title or not. Once each sentence is scored those sentences are ranked based on the descending order of weights.

Each sentence is scored based on the frequency of ‘n’ terms occurring in the document (i.e TF). If the term matches the title of the document, then each term that matches the title is multiplied by title factor of 2. This special weight is not equal to first sentence overlap [20]. The single document summarizer process single document at a time and generated summary. Term Frequency is calculated using expression (1):

$$TF_i = \sum_{j=1}^{m} \partial_{i,j} W_j \qquad \rightarrow (1)$$

where $\partial_{i,j} = 1$ if j^{th} term exists in i, otherwise $\partial_{i,j} = 0$. W_j = Number of occurrences of j^{th} term in the document. If W_j is among the terms in title its weight is multiplied by title factor of 2. Sentence score has been obtained by adding obtaining the cumulative sum of Term Frequency as given by expression 2.

$$Sentence_{score_i} = TF_i \qquad \rightarrow (2)$$

Based on the scores generated, sentences are chosen for summary depending on the compression ratio (for each document).

c. Merging Single document summaries:

Merging single document summaries involves identifying the similarity between each sentence in the document. Let us illustrate the calculation of similarity using a pseudo document having 4 sentences. Fig. 1 presents a graphical representation of the document while Table 3 gives the adjacency matrix. The entries in the matrix correspond to the similarities between sentences. Thus sentences 1 and 2 have a similarity of 0.5.

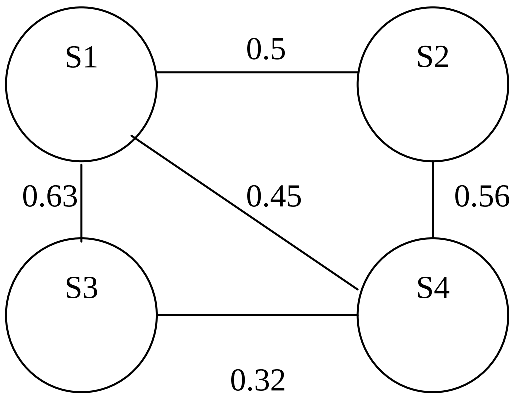

Fig.1 Representation of pseudo document by a graph

Table 3: Adjacency Matrix Corresponding to Figure 1

1.00	0.50	0.63	0.45
0.50	1.00	0.00	0.56
0.63	0.00	1.00	0.32
0.45	0.56	0.32	1.00

The interconnection weights are obtained using the cosine similarity across each sentences as shown by expression (3).

$$Cosine(t_i, t_j) = \sum_{h=1}^{k} t_{ih} t_{jh} \Big/ \sqrt{\sum_{h=1}^{k} t_{ih}^2 \sum_{h=1}^{k} t_{jh}^2} \quad \rightarrow (3)$$

Here i,j refers to the i^{th} and j^{th} sentences of the document. The above expression is without incorporation of IDF. Algorithm for merging is given in Figure 2. Depending on the target ratio, sentences were chosen.

Fig. 2 Algorithm to merge summaries

```
Input:   Set of single documents
         Set of files              ::  Filei
         compression ratio         ::   r
Output : Merged list of sentences depending on multi document clusters;
begin
         list ← empty;              /* initially merged list is empty
extract the sentences from each file depending on  'r'
similarity ( );                     /* measures the similarity of each sentence
sort( );                            / * sort the sentences based on score
repeat( )                           / * repeat for all files;
merged_list ← merged_list + nn ;      /* merge the sentences
end;
```

Table 4 presents the study results performed for generating multi document summaries from single document cluster. It is inferred from the values shown in Table 4 that the system was able to choose summaries effectively at 10% compression. While the system performance degrades as the percentage of compression increases. Figure 3 agrees with the above conclusion.

Table 4: Comparison of our single document summary with ideal summary generated by experts

MEAD			**Our approach**		
10%	**20%**	**30%**	**10%**	**20%**	**30%**
72	70	65	78	74	69

We have also carried out a study, to obtain 100% accuracy by varying the compression rates for both documents in each cluster. We have represented the Seven-point summary sheet of the documents using Minimum (Min), Maximum (Max), Median, Quartile1 (Q1), Quartile3 (Q3), Standard deviation (SD), Mean in Table 5.

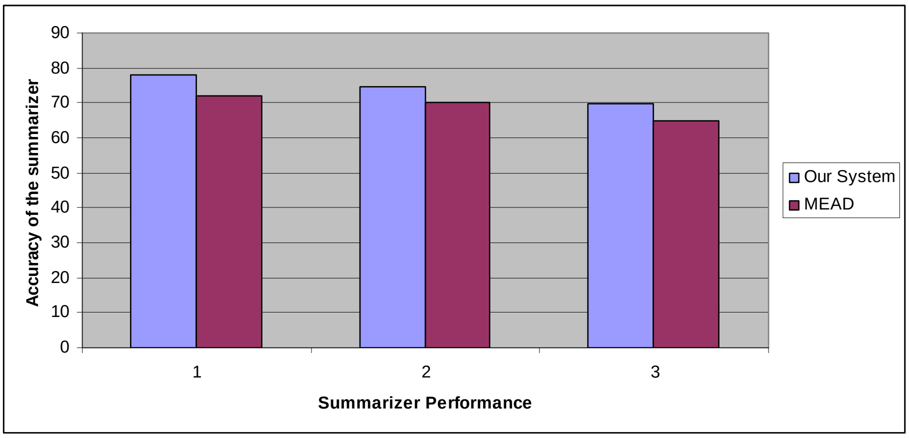

Fig. 3 Comparison of MEAD and our Summarizer at various compression rates

Table 5: Seven-point summary sheet

File No.	Compression Ratio	Min	Max	Median	Quartile1	Quartile3	SD	Mean
File1	**10%**	30	30	30	30	30	0.000	30.000
	20%	30	40	30	30	30	2.132	30.455
	30%	30	60	30	30	40	11.291	36.818
File2	**10%**	30	30	30	30	30	0.000	30.000
	20%	30	40	30	30	30	2.132	30.455
	30%	30	70	30	30	47.5	13.200	38.636

From the study, we found that summary generation at specified compression ratio is proportional to the single document summary generated at the same compression. The results would be enhanced further using linguistic processing tools to achieve 100% accuracy for the system with minimal compression ratio.

5 Conclusion & Future Work:

An attempt to choose sentences effectively from single document summary cluster is attempted rather from multi document source. It is shown from the study that generating summary from multi document cluster set poses some additional overhead misleading to generate in- effective target summaries. It is also investigated that the summary generation for multi document attempts at specified user compression ratio is proportional to the summaries corresponding to single document summaries at appropriate compression ratios.

Attempts focusing on document dimensions are not focused here. This work is performed only for generic summarization, while we focus our work on summary generation beyond generic summaries like task based approaches. We focus on to reduce the pick to sentences from reduced compression rates using some graphical techniques.

ACKNOWLEDGEMENTS

The author would wish to thank the Management, Vice Chancellor, Registrar, Head of the Department (Information Technology) of B.S.Abdur Rahman University, Vandalur, Chennai-48, Tamilnadu, INDIA for providing facility and guidance to carry out this work.

References

[1] Luhn, H. P, "The automatic creation of literature abstracts", *IBM Journal of Research and Development 2(2)*, Reprinted in Mani and Maybury ,1999,pp. 159–165.

[2] M. Hasan and R. Halliday. Cohesion in English. *Longman,London*, 1976.

[3] Hsun-Hui Huang, Yau-Hwang Kuo and Horng-Chang Yang ,"Fuzzy-Rough Set Aided Sentence Extraction Summarization",*IEEE, Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)*,2006.

[4] De-Xi Liu, Yan-Xiang He, Dong-Hong Ji and Hua Yang, "A Novel Chinese Multi-Document Summarization Using Clustering Based Sentence Extraction", *IEEE, Proceedings Of The Fifth International Conference On Machine Learning And Cybernetics*, Dalian, 2006, pp. 2592-2597.

[5] Yan-Min Chen, Xiao-Long Wang and Bing-Quan Liu, "Multi-Document Summarization Based On Lexical Chains", IEEE, Proceedings of the Fourth *International Conference on Machine Learning and Cybernetics*, Guangzhou, 2005,pp.1937-1972.

[6] Xu Yong-dong, Wang Xiao-long, Liu Tao, Xu Zhi-ming, "Multi-document Summarization Based on Rhetorical Structure: Sentence Extraction and Evaluation",*IEEE International Conference on Systems, Man and Cybernetics*, 2007,pp.3034 –3039.

[7] Li-Qing Qiu, Bin Pang, Sai-Qun Lin and Peng Chen, "A Novel Approach to Multi-document Summarization",*IEEE, 18th International Workshop on Database and Expert Systems Applications*,2007, pp.187-191.

[8] Wooncheol Jung, Youngjoong Ko and Jungyun Seo, "Automatic Text Summarization Using Two-Step Sentence Extraction", *LNCS, Springer-Verlag Berlin Heidelberg*, 2005, pp. 71 – 81.

[9] K. S. Jones. Automatic summarizing: Factors and directions. In I. Mani and M. T. Maybury, editors, Advances in Automatic Text Summarization, chapter 1, *MIT Press*, Cambridge, Massachusetts, 1999, pages 1–12.

[10] Jade Goldstein, Vibhu Mittal,Jaime Carbonell and Mark Kantrowitzt, “Multi-Document Summarization By Sentence Extraction”, *NAACL-ANLP 2000 Workshop on Automatic summarization*, Vol 4 ,2000, pp. 40 - 48.

[11] Rada Mihalcea and Paul Tarau, “An Algorithm for Language Independent Single and Multiple Document Summarization”, *In Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, Korea, 2005.

[12] Dragomir R. Radev, Hongyan Jing, Malgorzata Stys, and Daniel Tam, Centroid-based summarization of multiple documents”, *Information Processing and Management*,Vol. 40,2004, pp.919–938.

[13] Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *ANLP/NAACLWorkshop on Summarization*, Seattle,WA, 2000,pp.21-30.

[14] Radev, D. R., Blair-Goldensohn, S., & Zhang, Z, “Experiments in Single and Multi-document summarisation using MEAD”, *In Proceedings of the Document Understanding Conferences- 2001*, 2001.

[15] www.summarization.com/mead.

[16] www.google.com/news,www.rediffnews.com,www.yahoonews.com,www.hindu.com,www.indianexpress.co.in,www.cnn.com

[17] Wilbur J., “Human Subjectivity and Performance Limits in Document Retrieval”, *Information Processing and Management*, Vol. 32, No. 5, pp. 515-527, 1996.

[18] Shanmugasundaram Hariharan, "Merging Multi-Document Text Summaries- A Case Study", *Journal of Science and Technology*, Vol.5, No.4, pp.63-74, December 2009.

[19] Mani.I., and M.T Maybury. (1999) ‘Advances in Automatic Summarization’, *MIT Press* , Cambridege, MA.

[20] Nedunchelian Ramanujam, "Centroid Based Summarization of Multiple Documents Implemented Using Timestamps", *Proceedings of First International Conference on Emerging Trends in Engineering and Technology*,pp.480-485, 2008.