ISSN 1392–124X (print), ISSN 2335–884X (online) INFORMATION TECHNOLOGY AND CONTROL, 2014, T. 43, Nr. 4

The Benchmark of Paragraph and Sentence Extraction Summaries using Outlier Document Filtering based Multi -Document Summarizer

Metin Turan

Yıldız Technical University, Department of Computer Engineering
Davutpaşa, İstanbul, Turkey
e-mail: turanmetin@gmail.com

Coşkun Sönmez

İstanbul Technical University, Department of Computer Engineering
Maslak, İstanbul, Turkey
e-mail: sonmez@itu.edu.tr

Murat Can Ganiz

Doğuş University, Department of Computer Engineering
Acıbadem, İstanbul, Turkey
e-mail: mcganiz@dogus.edu.tr

crossref http://dx.doi.org/10.5755/j01.itc.43.4.7010

Abstract. We studied outlier document filtering (ODF) for extractive sentence summarization. Our results are superior compared to the average of the participant systems' using DUC 2006. Furthermore, we add extractive paragraph summarization to the same system. It is surprising that the results are nearly the same for ROUGE metrics. Although extractive paragraph summarization has a better performance for precision, extractive sentence summarization has a slightly better performance on the recall and F-Score which is the harmonic mean of recall and precision. The ODF is successful for both extractive sentence and paragraph summarization. The similarity metric (match percent) suggested in the article prevents the domination of longer sentences/paragraphs on shorter sentences/paragraphs in selection. As a result, the ODF provides the flexibility of paragraph extraction instead of sentence extraction for simplicity and readability and less work load.

Keywords: Document Processing; Extractive Summarization; Outlier Detection; Similarity Measure.

1. Introduction

Document processing and automatic summarization have been attracting growing interest as a result of uncontrollable amount of text data. Extracting an acceptable summary from a large number of documents is not an easy task even for humans. The document summarization is an important topic for scientific research as well as for commercial products due to the sheer size of documents accumulated in the databases of commercial entities. In this study we focus on the multi-document summarization. A survey about multi-document summarization approaches can be found in the article published by Kumar and Salim [1].

Most of the work in the literature is about extractive summarization due to its feasibility. These studies generally focused on a basic unit of a text; the sentence. Scoring is an important part of summarization, deciding which sentence is more informative. One of the first studies using term frequency (TF) is done by Luhn [2]. After one decade, Edmundson [3] suggested three methods (cue, title, location) in addition to the term frequencies to calculate sentence weights. Extractive summarization is still an active research topic with several recent studies [4-8].

Another research area for automatic extractive summarization is about preventing overlapped information at the summary. Researchers have tended to group similar text units by clustering and selecting the most informative text unit from each group. Different clustering techniques have been suggested for this purpose [9-12].

In general, extraction of sentences for summarization requires ranking the sentences. It also requires more work to discover the best ranking technique [13, 14].

Some researchers tend to use paragraph clustering or paragraph partition (relative paragraphs) for summarization [15-17]. Moreover, detecting the text partitions and labeling them with a topic or sub-topic has also been a challenging area [18-20].

There are comparatively a few number of studies that focus on the extraction of paragraphs instead of sentences for summarization [21-24]. By using paragraphs instead of sentences, it is possible that these methods may include meaningless sentences along with the informative ones in the summary. On the other hand, there are several advantages including short processing time, increased readability and simplicity. However, to the best of our knowledge, there are no benchmarks in the literature for comparing the performance of sentence and paragraph extraction fairly. Because of the ever growing size of documents both in the academics and the business, it is important to develop faster and simpler methods such as the paragraph based summarization methods.

The purpose of this study is to compare extractive sentence summarization and extractive paragraph summarization using Outlier Document Filtering (ODF) based multi-document summarizer. The aim is to find concrete differences, advantages and disadvantages and shed light on the future researches.

We first developed an extractive sentence summary system and proposed a new approach called ODF [25] to improve automatic summary quality. The ODF for sentence extraction (ODF_se) is evaluated on the DUC 2006 data set [26]. The system is compared with other systems which participated to DUC 2006 using the ROUGE metrics. The suggested technique significantly outperforms the DUC 2006 participants systems for each ROUGE metric. Following this, extractive paragraph summarization has been incorporated to this system and it is tested on similar data. Finally, both experiments are compared and discussed. The results show that the ODF for paragraph extraction (ODF_pe) is reasonable.

The article is organized so that Section 2 explains the proposed summarization system. Section 3 explains data set used in our experiments and evaluation methods. Section 4 provides experiment results. Sections 5 and 6 discuss the results and further work.

2. Approach

The proposed method includes simple components for an automatic summarizer and additionally some research modules. The general system flowchart is given in Fig. 1.

The first step is the processing step. In this step, each document is parsed separately by dividing into text units, initially paragraphs, after sentences. Porter stemming is applied to find roots of words. Stop words are filtered. TF's are calculated in the context of each text unit. Sentence term vectors (STV) and paragraph term vectors (PTV) are constructed. Finally, document term vector (DTV) is formed by using all terms in the document.

The second step is about finding important terms and the outlier document filtering. It starts with selecting important terms and using these terms as the representative (t_r) of the all document set. Term dispersion ratio (TDR) is the criteria used for selecting a term. If a term occurs in TDR percent of document set or more, it would be considered as a representative term (t_r) for all document set.

Experiments are divided into three separate TDR's as 25%, 50% and 75%. For example, if a document set contains 25 documents and TDR is 25% , then a term must be referred in at least $\lceil 25 \times 25\% \rceil = 7$ documents to be selected as t_r and it is a property of the document set term vector (DSTV).

Finally, distances between DTV's and DSTV are calculated. The documents that are far away from the 2σ distance on both directions from average distance (μ) are marked as outlier. The documents which are marked as outlier wouldn't be considered for further processing. The algorithm is called MarkOutliers and pseudocode is given at Algorithm 1. It takes two parameters, called documentCount and TDR, and they are used for the total document count in the document set and TDR selection, respectively.

The third step is about selecting the more informative sentences or paragraphs (ranking). It requires calculation of the similarity for each STV and PTV with the DSTV. The further similar sentences/paragraphs are selected for extractive summary under restriction of the summary size (given as parameter).

The following similarity metric, we called match percent (MP), is used to calculate similarity of each STV/PTV with the DSTV (1):

$$match = \sum \begin{Bmatrix} if\ term\ t_i \in STV/PTV\ and\ t_i\ is\ an\ t_r & 1 \\ otherwise & 0 \end{Bmatrix}$$

$$MP = \frac{match}{count\ of\ the\ terms\ in\ STV/PTV}. \qquad (1)$$

MP is structured to eliminate the superiority of long sentences/paragraphs (S/P) (term count is much more than a shorter one) over short S/P. MP is a ratio of similarity success of sentence/paragraph terms.

If two sentences have the same similarity value, document number, paragraph number and sentence number are the ranking criteria. If two paragraphs have the same similarity value, then document number and paragraph number are used for ranking.

MarkOutliers(float TDR, integer documentCount)
{
#minDoc = $\lceil TDR \times documentCount \rceil$
/* construct document set vector */
for each term t_i in the document set
search in $DTV_1, DTV_2, \ldots, DTV_{documentCount}$
if (t_i is a member of at least #minDoc documents)
Set $t_i \in DSTV$;
/* calculate distance of each document */
for each document di in the document set
calculate euclidean distance ED_i of DTV_i with $DSTV$;
/* find outlier boundary using μ and $2 * \sigma$ */

$$distance_{\mu} = \left(\sum_{i=1}^{documentCount} ED_i \right) / documentCount$$

$limit\_upper = distance_{\mu} + 2\sigma;$
$limit\_lower = distance_{\mu} - 2\sigma;$
/* mark outlier documents */
for each di in the document set
if ($ED_i < limit\_lower$ OR $ED_i > limit\_upper$)
mark d_i as outlier;
}

Algorithm 1. MarkOutliers Algorithm

3. Experiment Setup

The DUC 2006 corpus includes 50 document set from Financial Times of London and Los Angeles Times. Each document set includes 25 news. The system summary is limited to 250 words for each document set. Extraction type is sentence based. Four models are used to produce human summaries for each document set. Thirty-five systems were attended to the competition and all generated an output for each document data set. They are evaluated, and scores are published.

ROUGE metrics [20] are used to evaluate each system generation with human models. ROUGE metrics included in DUC 2006 tests are as follows:

ROUGE-N: N-gram based co-occurence statistic is given. DUC tests are done from 1 to 4 gram.

ROUGE-L: Longest common subsequence (LCS) based statistic is given. It is sentence based similarity and it identifies longest n-gram sequence.

ROUGE-W: Consecutive LCS based statistic is given.

ROUGE-SU: Bigram plus unigram skipped co-occurrence statistic is given.

ROUGE metrics include three properties: precision (2), recall (3), F-Score (4). Their definitions are given as follows:

$$Precision = \frac{\{relevant\ S/P\} \cap \{retrieved\ S/P\}}{\{retrieved\ S/P\}} \qquad (2)$$

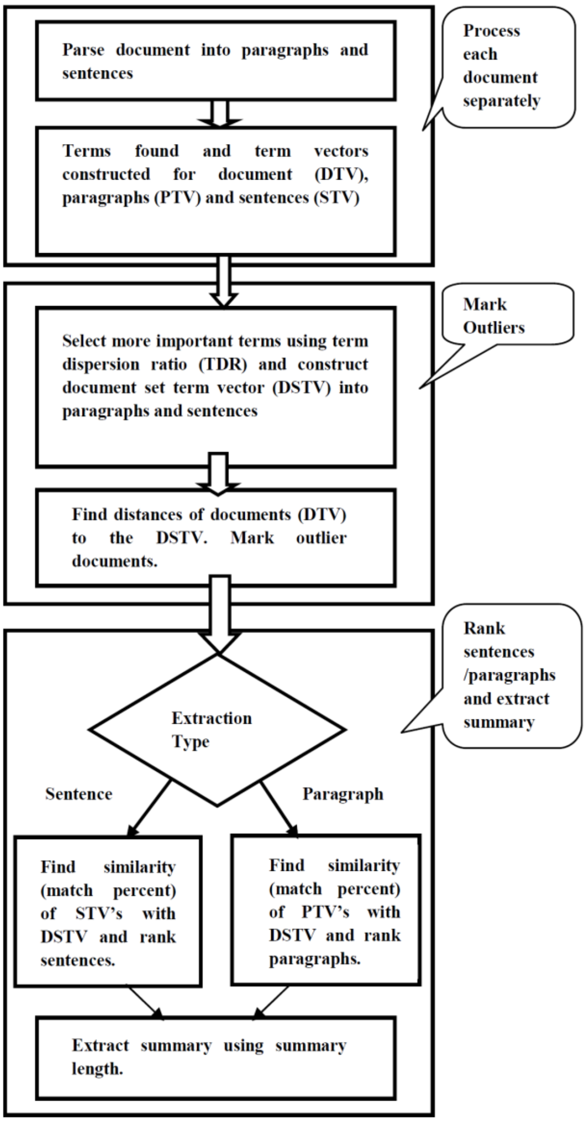

Figure 1. Outlier document filtering applied multi-document summarizer

$$Recall = \frac{(\{relevant\ S/P\} \cap \{retrieved\ S/P\})}{\{relevant\ S/P\}} \qquad (3)$$

$$F - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall}. \qquad (4)$$

Precision gives the percentage of true S/P (match with human summaries) in all retrieved S/P. However recall gives the percentage of true S/P retrieved. On the other hand, F-Score is the harmonic mean of recall and precision of the system.

We have selected DUC 2006 data set for the evaluation because the data set contains much more errors in the documents intently. As a result, the average values of ROUGE metrics produced by participants systems are really lower than the values obtained from other data sets. Moreover, this data set is produced for extractive summarization originally.

4. Experiment Results and Discussions

ODF_se was run for every TDR on each data set (50 data sets) separately. The average ROUGE score was calculated for 3 TDR values. In other words, one data set requires 3 runs, totally (all data set) 150 runs.

Table 1 gives the average of the participants systems and ODF_se ROUGE metrics using DUC 2006 data set for different TDR's.

Average_R, *Average_P* and *Average_F* abbreviations on Table 1 and Table 2 are used for average recall, average precision and average F-Score, respectively. On the other hand, the maximum value for each row on the table is marked bold to be realized easily.

It is clear that ODF has a performance acceptable for automatic summarization. The attractive point of ODF is that it is successful (acceptable) even for 4-gram values. It doesn't drop down sharply for ngrams because only the more informative terms selected and outlier documents are filtered.

Later, ODF for paragraph extraction (ODF_pe) was run in the same way. The property score for ROUGE metrics was recorded for each data set in DUC 2006 corpus. Averages were calculated and finally Table 2 was constructed. It includes the comparison of ROUGE metrics obtained for different TDR runs side by side.

There are lots of experiments done on DUC [27] and TAC (Text Analysis Conference) [28] data sets. The best ROUGE metrics obtained are known by the Kumar and his colleagues work [29]. They also shared and compared the values of ROUGE metrics which are previously obtained from well-known automatic multi-document summarization systems. Their system had been experimented on DUC 2002 and DUC 2004 data sets. The higher score for ROUGE-1 is 0.51746 (on DUC 2002) which is still below our system result. Another important work which ROUGE metrics are over the DUC participants average is done by Wan and Xiao [30]. They devised a topic-focused graph-based

Table 1. Comparison of DUC 2006 participants and ODF_se ROUGE metrics for different TDR's

	DUC-participants	**Outlier-TDR 25%**	**Outlier-TDR 50%**	**Outlier-TDR 75%**
ROUGE-1				
Average-R	0.371	**0.641**	0.635	0.637
Average-P	0.386	**0.540**	**0.540**	0.535
Average-F	0.377	**0.583**	0.581	0.579
ROUGE-2				
Average-R	0.073	0.413	0.413	**0.417**
Average-P	0.076	0.347	**0.351**	0.349
Average-F	0.074	0.375	0.377	**0.379**
ROUGE-3				
Average-R	0.020	0.344	0.344	**0.348**
Average-P	0.021	0.289	**0.292**	0.291
Average-F	0.021	0.312	0.315	**0.316**
ROUGE-4				
Average-R	0.008	0.301	0.302	**0.305**
Average-P	0.009	0.253	**0.256**	0.255
Average-F	0.008	0.273	0.276	**0.277**
ROUGE-L				
Average-R	0.340	0.568	0.571	**0.574**
Average-P	0.353	0.478	**0.485**	0.481
Average-F	0.346	0.516	**0.522**	**0.522**
ROUGE-W				
Average-R	0.099	0.156	0.157	**0.158**
Average-P	0.188	0.256	**0.260**	0.258
Average-F	0.129	0.193	**0.195**	**0.195**
ROUGE-SU4				
Average-R	0.128	**0.408**	0.399	0.401
Average-P	0.133	**0.293**	**0.293**	0.285
Average-F	0.130	**0.334**	0.332	0.330

Table 2. Comparison of ODF_se and ODF_pe summaries ROUGE metrics for different TDR's

	TDR Senten.	25% Parag.	TDR Senten.	50% Parag.	TDR Senten.	75% Parag.
ROUGE-1						
Average-R	**0.641**	0.608	0.635	0.604	0.637	0.606
Average-P	0.540	**0.575**	0.540	0.574	0.535	0.566
Average-F	0.583	**0.589**	0.581	0.587	0.579	0.584
ROUGE-2						
Average-R	0.413	0.383	0.413	0.382	**0.417**	0.386
Average-P	0.347	0.362	0.351	**0.363**	0.349	0.359
Average-F	0.375	0.371	0.377	0.371	**0.379**	0.371
ROUGE-3						
Average-R	0.344	0.316	0.344	0.300	**0.348**	0.320
Average-P	0.289	0.298	0.292	**0.307**	0.291	0.298
Average-F	0.312	0.306	0.315	0.307	**0.316**	0.308
ROUGE-4						
Average-R	0.301	0.276	0.302	0.277	**0.305**	0.280
Average-P	0.253	0.260	0.256	**0.262**	0.255	0.260
Average-F	0.273	0.267	0.276	0.268	**0.277**	0.269
ROUGE-L						
Average-R	0.568	0.534	0.571	0.534	**0.574**	0.540
Average-P	0.478	0.505	0.485	**0.508**	0.481	0.503
Average-F	0.516	0.518	**0.522**	0.519	**0.522**	0.520
ROUGE-W						
Average-R	0.156	0.147	0.157	0.147	**0.158**	0.148
Average-P	0.256	0.270	0.260	**0.273**	0.258	0.270
Average-F	0.193	0.190	**0.195**	0.191	**0.195**	0.191
ROUGE-SU4						
Average-R	**0.408**	0.362	0.399	0.359	0.401	0.362
Average-P	0.293	0.324	0.293	**0.325**	0.285	0.316
Average-F	0.334	**0.339**	0.332	0.337	0.330	0.334

system and evaluated ROUGE metrics F-Score property for DUC 2005, DUC 2006 and DUC 2007 data sets. The obtained values for F-Score property are higher than the values produced by DUC participants systems. However the success is limited to a small percent, such that ROUGE-1 metric F-Score property is 0.40306 for the best system suggested in the article. Finally, Galanis [31] produced a system which uses integer linear programming and support vector regression together. The authors compare their work with other important works in the literature. ROUGE-2 and ROUGESU4 metrics were calculated. DUC 2006 data set is used for training, where DUC 2007 and DUC 2005 are used for evaluation. The results of this system are also much lower than the results produced by our system. As a result, to the best of our knowledge, we can state that our system has better ROUGE metrics results than the studies in the literature for the automatic multi-document summarization systems.

5. Conclusion

Even a perfect summary is produced, the reader may want to read the referred documents in the summary for the sake of details. If the summary consists of bigger text units such as paragraphs instead of sentences, it would be helpful for the readers. This insight supports the idea of extracting informative paragraphs from text documents. This approach also increases the applicability of summarization in real world business problems. Based on this notion, our research efforts are focused on devising a successful (at an acceptable degree) paragraph extraction system.

Our experiment results show that ODF_pe performs as well as ODF_se in terms of several ROUGE metrics. Moreover, ODF_pe is more consistent as indicated by the higher precision results.

In general, ROUGE metrics show that ODF_pe increases the precision while decreasing the recall. It is well known that there is an inverse relationship between precision and recall. Precision is higher because the ratio of true sentences to the all sentences in the ODF_pe is much larger. On the other hand, in ODF_se recall is higher because the ratio of true sentences to all true sentences is much larger as in summaries generated by humans.

As a result, we conclude that the number of the total sentences in the ODF_pe should be less than the number of total sentences in the ODF_se because denominator should decrease in order to maximize Eq. (2). On the other hand, the number of true sentences found in the ODF_se is equal or much more than the number of true sentences found in the ODF_pe because numerator should increase in order to maximize Eq. (3).

Considering the TDR values, we could easily conclude that generally better ROUGE metric values for ODF_pe are obtained for 50% TDR. On the other hand, generally better ROUGE metric values for ODF_se are obtained for 75% TDR. These results point out that paragraph extraction is more tolerable to the outlier documents in the document set.

If we consider the run time performance issues, ODF processing step is shorter due to parsing into big text units. This means we only have a small number of paragraphs instead of lots of sentences. This leads to a simpler algorithm. Perhaps the more important result is that paragraphs are not partial text units. However, sentence is a portion of a paragraph. Consequently, the summary is more readable.

6. FutureWork

The term filtering used to construct DSTV is a part of ODF. It is a successful technique to select more representative terms. It could be valuable applying term filtering during processing step of automatic summaryzer additionally. We also consider using counting and max-min methods. On both methods, all terms in document set are sorted by frequencies in decreasing order initially.

Counting method: The terms are selected based on total term count to construct STV's and PTV's. For example, if 100 terms are found in a document set and percentage selected is 50%, then the most frequent 50 terms would be selected for STV's and PTV's.

Max-Min method: A frequency limit determined using the minimum and maximum frequencies (minFreq and maxFreq) of the terms. The formula for Max-Min method is defined in (5):

$$\geq (maxFreq - minFreq) \times percent + minFreq. \tag{5}$$

For example, if maximum and minimum frequent terms are 100 and 10, respectively, and percentage selected is 50%, then the terms equal or above the frequency 55 would be selected for STV's and PTV's:

$$\geq ((100 - 10) \times 0.5 + 10) = 55.$$

ODF can be improved additionally. ODF uses 2σ distance on both sides and covers nearly 95 percent of all data for normally distributed data set. In other words, it only filters 5% of the documents by using outlier filtering. Using $1.5 \times \sigma$ distance instead of 2σ distance in the MarkOutliers() algorithm could be suggested. It would filter nearly 13% of the document set. It may produce considerably better result on the unbalanced document set intuitively.

References

[1] **Y. J. Kumar, N. Salim.** Automatic multi document summarization approaches. *Journal of Computer Science*, 2012, Vol. 8, No. 1, 133-140.

[2] **H. P. Luhn.** The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 1958, Vol. 2, No. 2, 159-165.

[3] **H. P. Edmundson.** New methods in automatic extracting. *Journal of the ACM*, 1969, Vol. 16, No. 2, 264-285.

[4] **R. M. Aliguliyev.** A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems With Applications*, 2009, Vol. 36, No. 4, 7764-7772.

[5] **A. Bossard, C. Rodrigues.** Combining a multi-document update summarization system-CBSEASwith a genetic algorithm. In: *Proceedings of the 2nd International WorkShop on Combinations of Intelligent Methods and Applications*, 2011, Vol. 8, pp. 71-87.

[6] **V. Gupta, G. S. Lehal.** A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence*, 2010, Vol. 2, No. 3, 258-268.

[7] **S. Hariharan.** Multi document summarization by combinational approach. *International Journal of Computational Cognition*, 2010, Vol. 8, No. 4, 68-74.

[8] **L. Suanmali, N. Salim, M. S. Binwahlan.** Fuzzy logic based method for improving text summarization. *International Journal of Computer Science and Information Security*, 2009, Vol. 2, No. 1.

[9] **M. Kirubakaran, M. Kirubakaran, S. Hariharan.** Experiments on clustering and multi-document summarization. *ICFAI Journal of Computer Sciences*, 2009, Vol. 3, No. 2, 64-73.

[10] **Y. Nie, D. Ji, L. Yang, Z. Niu, T. He.** Multi-document summarization using a clustering-based hybrid strategy. In: *Asia Information Retrieval Symposium*, 2006, pp. 608-614.

[11] **S. Park, B. Cha, D. U. An.** Automatic multi-document summarization based on clustering and nonnegative matrix factorization. *IETE Technical Review*, 2010, Vol. 27, No. 2, 167-178.

[12] **D. Wang, S. Zhu, T. Li, Y. Chi, Y. Gong.** Integrating document clustering and multi-document summarization. *ACM Transactions on Knowledge Discovery from Data*, 2011, Vol. 5, article 14.

[13] **D. Bollegala, N. Okazaki, M. Ishizuka.** A bottom-up approach to sentence ordering for multi-document summarization. *Information Processing and Management*, 2010, Vol. 46, No. 1, 89-109.

[14] **A. Kogilavani, P. Balasubramani.** Clustering and feature specific sentence extraction based summarization of multiple documents. *International Journal of Computer Science Information Technology*, 2010, Vol. 2, No. 4, 99-111.

[15] **J. Chang-Jin, P. Hong, M. Qian-Li, C. Jian-Chao.** Automatic summarization for Chinese text based on combined words recognition and paragraph clustering.

In: *Third International Symposium on Intelligent Information Technology and Security Informatics*, 2010, pp. 591-594.

[16] **P. Hu, T. He, D. Ji, M. Wang.** A study of Chinese text summarization using adaptive clustering of paragraphs. In: *Proceedings of the 4th International Conference on Computer and Information Technology*, 2004, pp. 1159-1164.

[17] **W. Min, L. Zhensheng, G. Yuqing.** Study on semantic paragraph partition in automatic abstracting system. *IEEE International Conference on Systems, Man, and Cybernetics*, 2001, Vol. 2, pp. 892-897.

[18] **R. Mihalcea, C. Corley, C. Strapparava.** Corpus based and knowledge-based measures of text semantic similarity. In: *Proceed. of American Association for Artificial Intelligence*, 2006, pp. 775-780.

[19] **C. W. Wu, C. L. Liu.** Ontology-based text summarization for business news articles. In: *Proceedings of the ISCA 18th International Conference on Computers and Their Applications*, 2003, pp. 389-392.

[20] **Z. Zhenfang, L. Peiyu, L. Ran, C. Xuezhi.** A logical paragraph division based on semantic characteristics and its application. *The Nineth International Symposium on IT in Medicine & Education*, 2009, pp. 966-973.

[21] **C. Jaruskulchai, C. Kruengkrai.** A practical text summarizer by paragraph extraction for Thai. *The Sixth International Workshop on Information Retrieval with Asian Languages*, 2003, pp. 9-16.

[22] **M. Mitra, A. Singhal, C. Buckley.** Automatic text summarization by paragraph extraction. *Workshop on Intelligent Scalable Text Summarization*, 1997, pp. 1-11.

[23] **G. Salton, A. Singhal, C. Buckley, M. Mitra.** Automatic text decomposition using text segments and text themes. *The Seventh ACM Conference on Hypertext*, 1996, pp. 53-65.

[24] **G. Salton, J. Allan, C. Buckley, A. Singhal.** Automatic analysis. Theme generation and summarization of machine readable texts. *Science*, 1994, Vol. 264, No. 5164, 1421- 1426.

[25] **M. Turan, C. Sönmez.** Outlier document filtering applied to the extractive summarization. *International Journal of Mathematical and Computational Science*, 2014, Vol. 1, 26-29.

[26] **NIST.** Document Understanding Conference Past Data (DUC 2006). *http://www.nlpir.nist.gov/projects/duc/duc2006*, [Online: accessed 11-12-2013].

[27] **NIST.** Publications of Document Understanding Conferences. *http://www.nlpir.nist.gov/projects/duc/pubs.html*, [Online: accessed 03-03-2014].

[28] **NIST.** Publications of Text Analysis Conferences. *http://www.nist.gov/tac/publications/index.html*, [Online: accessed 11-08-2014].

[29] **N. Kumar, K. Srinathan, V. Varma.** Using Wikipedia anchor text and weighted clustering coefficient to enhance the traditional multi-document summarization. In: *Proceedings of CICLing 2012*, 2012, pp. 390- 401.

[30] **X. Wan, J. Xiao.** Graph-based multi-modality learning for topic-focused multi-document summarization. *IJCAI'09 Proceedings of the 21st international joint Conference on Artifical intelligence*, 2009, 1586-1591.

[31] **D. Galanis, G. Lampouras, I. Androutsopoulos.** Extractive multi-document summarization with integer linear programming and support vector regression. *The 24th International Conference on Computational Linguistics*, 2012, pp. 911-926.

Received April 2014.