Keeping Maximal Frequent Sequences Facilitates Extractive Summarization

Ledeneva Yulia[1], Alexander Gelbukh[1], René García Hernández[2]

[1]Centro de Investigación en Computación, Instituto Politécnico Nacional
Av. Juan de Dios Bátiz s/n, D.F., 07738, México
yledeneva@yahoo.com
www.Gelbukh.com
[2]Universidad Autónoma del Estado de México
C.U. Atlacamulco, Estado de México, México
renearnulfo@hotmail.com

Abstract. Automatic text summarization helps us to save time looking for information and reading quickly big volumes of documents. The main objective of this paper is to produce a text summary extracting the parts of text which conform the summary of a document. Generally, words, n-grams, phrases or sentences are selected from the original text to make the summary. In this work, we propose to extract maximal frequent sequences of the words for generate the summaries. Maximal frequent sequences can be extracted independently of the language and have the advantage of not defining previously the length of the part to be extracted. The proposed method, the experiments and some results are presented.

1. Introduction

The huge amount of available electronic documents in Internet has motivated the development of very good information retrieval systems. However, the information provided by such systems, like *google*, only show part of the text where the words of the request query appears. Therefore, the user has to decide if a document is interesting only with the extracted part of a text. Moreover, this part does not have any information if the retrieved document is interesting for the user, so it is necessary download and read each retrieved document until the user finds satisfactory information. It was unnecessary and time-consuming routine. A solution for such problems is to achieve an automatic text summarization of the document extracting the essential sentences of the document.

The demand of the automatic generation of text summaries has appeared in other areas, for example, summaries of news articles; summaries of electronic mails and news to send them as SMS; summaries of information (for government officials, businessmen, researches, etc.); summaries of web pages to transmit them through telephone; in searching systems to receive the summaries of found documents and pages.

© G. Sidorov, B. Cruz, M. Martínez, S. Torres. (Eds.)
Advances in Computer Science and Engineering.
Research in Computing Science 34, 2008, pp. 163-174

Received 27/03/08
Accepted 26/04/08
Final version 04/05/08

In this paper, we focus on single text summarization which implies to communicate the principal information of one specific document. There are two options to achieve it: text abstraction and text extraction [1]. Text abstraction examines a given text using linguistic methods which interpret a text and find new concepts to describe it. And then new text is generated which will be shorter with the same content of information. Text extraction means extract parts (words, sequences, sentences, paragraphs, etc.) of a given text based on statistic, linguistic or heuristic methods, and then join them to new text which will be shorter with the same content of information.

The main objective of this paper is to produce a text summary extracting the text, or in other words, we are looking for a selection of the parts of the text for using them to produce the summary of a single document. Generally, in this search candidate parts like words, phrases, sentences or paragraphs are selected for generate the summary. Intuitively, if greater parts are selected a system would tend to produce more understandable summaries. In this sense, extracted Maximal Frequent Sequences (MFSs) can be parts potentially-candidates for the summary, with the benefit to be understandable by the human. Moreover, there is an advantage of not having to define previously the length or type of the part to be extracted. As we try to generate summaries (in other words to obtain a compact representation), an intuitive evidence is to make use of a compact representation of a text with MFSs. Additionally, since MFSs can be extracted language-independently then could be obtained summaries for different languages.

The paper is organized as follows. Section 2 summarizes the state-of-the-art of text summarization methods. In Section 3, MFSs are described. Section 4 proposes a new methods for generate summaries for a single document. The experimental results are presented in Section 5.

2. Related Work

The procedure for obtaining a text summary from a single document can be divided in some steps, such as term selection, term weighting, sentence weighting and sentence selection methods. The state-of-the-art of text summarization methods in each of these steps is as described below.

The *term selection* group consists of the following methods. The first method [2] extracts key words in a text which are determined previously. Liu [3] represent words and vector of words with Basic Elements (BE) considering that BE is a more precise semantic unit than a word. Also, proper names [4], similarity with a title [2], and ontologies [5] are used as selected terms. The most extracted are n-grams [6] which try to find sequences of n consecutive words. Moreover, extracting frequent terms showed good results in [7], and this work is a continuation of [7].

The *term weighting* group is leaded by calculating a frequency of terms selected by the first group of methods. The weight is assigned for a term if this term is located in a sentence [2]. Also the terms are connected of some form within a document, so the level of relevance of terms can be determined by means of lexical connections, repetitions, co-references, synonymous and by semantic associations that can be

expressed in a thesaurus [8]. The methods of machine learning also can be as an option for weight terms [6].

The *sentence weighting*, or the method determining the importance of the sentence, as we see it, is looking if there is any term in a sentence. The sentences with more terms are chosen to generate a summary. Another criterion is a discourse structure [9] that consists of determining the fundamental structure of the text and of this form to weigh the sentences according to the proximity which they have with the central idea. This type of techniques involves the use of very sophisticated and expensive linguistic resources. Also, the degree of connectivity of the words that conform the sentences is calculated against the rest of the document by means of lexical connections as in [8], to calculate the importance of the sentence.

The methods of *sentence selection* can be based on the position of sentences in a text [2], or using more complicated methods like [10, 11].

In this paper, our hypothesis is that the usage of MFSs for generating summaries gives us better result than usage of n-grams. The main idea of the proposed method is to find MFSs which exceed established threshold. These MFSs allow determine the importance of the sentence in a text. Finally, the most important sentences with higher weight will be included in the text. We propose different schemes to determine the importance of the sentence depending on the MFSs located in this sentence. In the following section the proposed method is described in details.

3. General Scheme of the Proposed Method

We describe the general scheme of the proposed algorithm which consists of the four steps:

Term Selection. When we say the terms, we referred to the features which we use in this step. The terms are words, n-grams, or MFSs extracted from a document. The details of MFSs are described below. Also we extract terms derived from MFSs such as words and n-grams, we call them $n\text{-}grams_{mfs}$.

Term Weighting. We propose a scheme for weighting MFS which take into account T_i frequency of MFS, length of MFS, and frequency of derived terms from MFS. The terms T_i can be weighted in different manners, and have a weight t_i. This general scheme is defined as $p_i(t_j) = X \cdot Y$, where $p_i(t_j)$ - term weighting j in the documents i, X and Y can be determined as frequency of MFS, length of MFS, and frequency of derived terms from MFS. This term weighting scheme permits to detect which of the characteristics of MFS helps better to summarize a text.

Sentence Weighting. Sentence S_i has weight s_j = sum w_{ij}, contribution of T_i in S_j is $w_{ij} = f_{ij} \times t_i$, where f is a presence of T_i in D_j, t is a importance of T_i. Here f is binary.

Sentence Selection. This procedure completes a summary adding the densest sentences or choosing the position of a sentence in a text until the summary is limited by the number of words. As the first option, we chose the sentences which have more weighting score. This type of methods is dominion-independent and can be applied

for a variety of texts. As the second option, the type of methods is position dependent and can be applied only for special topics.

The first contribution of this work is a novel proposal for feature selection: as features we propose to use MFS which is described in Section 4. And the second contribution is an exploration of different methods in each step of generation of a text summary for single document.

4. Maximal Frequent Sequences

The text of a document is expressed by words in a sequential order. Therefore, it could be useful determining the consecutive word sequences that appear frequently in a document. Also, it is possible to determine which of the frequent sequences are not contained in any other frequent sequence, *i.e.*, which of them are maximal. In this work, we extract MFSs taking into account that they are a compact representation of the frequent sequences and can facilitate the generation of text summaries.

García [12] proposed an efficient algorithm to find the maximal consecutive frequent sequences of words from a single document. In his works [12, 13] the maximal frequent sequences are formally defined as follows:

Definition 1. A sequence $P = p_1p_2...p_n$ is a subsequence of a sequence $S = s_1s_2...s_m$, denoted $P \subseteq S$, if there exists an integer $1 \leq i$ such that $p_1 = s_i$, $p_2 = s_{i+1}$, $p_3 = s_{i+2}, ..., p_n = s_{i+(n-1)}$.

Definition 2. Let $X \subseteq S$ and $Y \subseteq S$ then X and Y are exclusive if X and Y do not share items *i.e.* if ($x_n = s_i$ and $y_1 = s_j$) or ($y_n = s_i$ and $x_1 = s_j$) then $i < j$.

Definition 3. Let T be a text expressed as a sequence, a sequence S is *frequent* in T if it is contained at least β times in T in an exclusive way, where β is the user-specified threshold.

Definition 4. A frequent sequence is *maximal* if it is not a subsequence of any other frequent sequence.

See Table 1 for some examples of MFSs. Here, MFSs are shown for a collection of 4 documents with a threshold $\beta = 3$, where resulted MFS is "***is the most beautiful***".

Table 1. MFSs for a collection of 4 documents.

...In Cholula you can find 365 churches and *the most beautiful is* of the pyramid...
...The church of Tonantzintla **is the most beautiful** ...
...Xcaret **is the most beautiful** natural park of Cancun ...
...Cantona **is the most beautiful** city of the ancient cities ...

The MFSs present some important characteristics. First, they keep the sequential order of the words; it means the MFSs do not lose the sequential order of the text. Second, the length of the MFSs is not previously determined; it is determined by the document content. And third, the MFSs can be obtained independently of the language of the documents. See more applications of MFSs in natural language processing tasks [14, 15].

5. Experimental Results

5.1. Experimental Configuration

Test data set. The proposed methods were tested with a collection Document Understanding Conference (DUC) provided in [16]. In particular, we use the data set of 567 news articles of different length and with different topics. For each document different summaries are given. Each summary which is used for evaluation was generated by two different experts manually. The experts were asked to generate summaries expressing the content of an original text by concatenating representative sentences. Each expert was asked to generate summary of different length. We use summaries which have 100 words.

Evaluation. ROUGE evaluation toolkit [17] was used for evaluation. ROUGE is a method that based on n-gram statistics, found to be highly correlated with human evaluations. This method compares the summaries made manually and generated by proposed methods. The following measures are used for evaluation. The evaluation is done using n-gram (1, 1) setting of ROUGE, which was found to have the highest correlation with human judgments, at a confidence level of 95%.

Baseline. The baseline configuration selects the first sentences of the text until the desired size is reached [16]. This configuration gives very good results on the kind of the texts (news reports) that we experimented with, but would not give so good results on other types of texts. Thus we use another baseline configuration which is called *baseline-random*.

Baseline-*random*. This configuration selects random sentences. The results presented below are averaged by 10 runs are taken from [7].

5.2. Experimental Procedures

Experimental Procedure 1. For this experiment, we use three term selection methods and two term weighting schemes; specifically, we try a proposed hypothesis for a term selection method with MFSs.

Term Selection. In the first term selection method only words are extracted; in the second method the n-grams are extracted; and in the third method MFSs are extracted.

Term Weighting. First scheme: term frequency of corresponding selected term is calculated as $p_i(t_j) = Z$, where Z is a frequency of selected term T_i in a text. Second scheme: is similar to the first scheme but stop-words are eliminated previously.

Sentence Weighting. For each sentence, the sum of all term weights is calculated.

Sentence Selection. The sentences with more weight are composed a summary.
Table 3 shows the results obtained for described procedure. The results proved that the proposed term selection method exceeds the performance of words and n-gram methods.

Table 2. Results for experimental procedure 1.

Term Selection	With stop-words	Without stop-words
Words	0.39421	0.41371
n-grams	0.40810	0.42173
Proposed (MFS)	0.43066	0.44085

Experimental Procedure 2. In this procedure, we try different term selection, term weighting and sentence selection schemes; precisely, all of these schemes are proposed in this paper.

Term Selection. First, MFSs are selected for each document. Second, the terms derived from MFSs are extracted. The resulted terms derived from MFSs are *n*-grams or $n\text{-}grams_{mfs}$, where n is a length of MFS and denoted as n_{len}. We are talking about a word selected from MFS, when $n = 1$, and we are talking about *n*-grams, when $n = 2, 3, ..., n_{len} - 1$, and we are talking about MFSs when $n = n_{len}$ (see Table 3).

Table 3. Terms for selection.

Size of n	Terms from MFS
1	words
$2, 3, \ldots, n_{len} - 1$	*n*-grams
n	MFSs

Term Weighting. First, the frequency of terms derived from MFSs are calculated for each document. Second, three schemes are tried as weights for selected terms. First scheme: $p_i(t_j) = 1$, where $p_i(t_j)$ is a term weighting j in the documents i.
Second scheme: $p_i(t_j) = X \cdot Y$, where X is a length of MFS, Y is a frequency of MFS. In Table 4, you can find the results for different combinations of X and Y.
Third scheme: $p_i(t_j) = Z$, where Z is a frequency of $n\text{-}grams_{mfs}$ in a text.

Sentence Weighting. For each experiment, the sum of all term weights is calculated for each sentence.

Sentence Selection. First scheme: sentences N_{sent} with more weight are composed a summary. Second scheme: the combination of sentences from different methods is used. We start composing a summary with the first sentence $N_{com(s1)}$ and the second sentence $N_{com(s1+s2)}$ selected from the best scheme of this paper, and then each summary is completed with first sentences from baseline configuration. In Table 4 the number of sentences is indicated.

Table 4 presents the results of the proposed schemes for experimental procedure 2. The best results are highlighted with bold type of letter. We detect that the scheme weighted with frequency of words derived from MFSs gives the best sentence for a summary, and together with sentences obtained with baseline configuration, the best summary for this scheme was obtained.

Table 4. Results for experimental procedure 2 with $\beta = 2, 3, 4$.

Term Selection	Term Weighting		Sentence Selection	Results		
n-grams$_{mfs}$	stop-words	$p_i(t_j)$		Recall	Precision	Fmeasure
n_{len}	yes	$X \cdot Y$	N_{sent}	0.43353	0.44737	0.44022
n_{len}	no	$X \cdot Y$	N_{sent}	0.43812	0.45411	0.44581
n_{len}	no	1	N_{sent}	**0.44128**	**0.45609**	**0.44840**
n_{len}	no	X	N_{sent}	0.43977	0.45587	0.44752
n_{len}	no	$X \cdot X$	N_{sent}	0.42995	0.44766	0.43847
n_{len}	no	Y	N_{sent}	0.44085	0.45564	0.44796
1	yes	Z	N_{sent}	**0.44582**	**0.45820**	**0.45181**
1	no	1	N_{sent}	0.38364	0.40277	0.39284
1	no	Z	N_{sent}	0.44609	0.45953	0.45259
1	no	$Z \cdot Z$	N_{sent}	0.43892	0.45265	0.44556
2... n_{len}-1	no	Z	N_{sent}	0.43711	0.45099	0.44383
1	no	Z	$N_{com(s1)}$	**0.46576**	**0.48278**	**0.47399**
1	no	Z	$N_{com(s1+s2)}$	0.46158	0.47682	0.46895
n_{len}	no	X	$N_{com(s1)}$	0.46381	0.48124	0.47223
n_{len}	no	1	$N_{com(s1)}$	0.46354	0.48072	0.47185
n_{len}	no	X	$N_{com(s1+s2)}$	0.45790	0.47430	0.46583

We tested the proposed schemes with $\beta = 2, 3, 4$ (Table 4). Then, we tested the proposed schemes with $\beta = 2$ (see Table 5), $\beta = 3$ (see Table 6), and $\beta = 4$ (see Table 7). The comparison results are shown below in Tables 8 – 10.

Table 5. Results for experimental procedure 2 with $\beta = 2$.

Term Selection	Term Weighting		Sentence Selection	Results		
				Recall	Precision	Fmeasure
$n\text{-}grams_{mfs}$	stop-words	$p_i(t_j)$				
n_{len}	No	$X \cdot Y$	N_{sent}	0.43731	0.45347	0.44508
n_{len}	No	1	N_{sent}	**0.43749**	**0.45182**	**0.44438**
n_{len}	No	X	N_{sent}	0.43731	0.45347	0.44508
n_{len}	No	$X \cdot X$	N_{sent}	0.42781	0.44566	0.43640
1	No	1	N_{sent}	0.38367	0.40290	0.39291
1	No	Z	N_{sent}	**0.44659**	**0.45968**	**0.45293**
1	No	$Z \cdot Z$	N_{sent}	0.44114	0.45512	0.44790
1	No	Z	$N_{com(s1)}$	**0.46536**	**0.48230**	**0.47355**
1	No	Z	$N_{com(s1+s2)}$	0.46296	0.47769	0.47009
n_{len}	No	X	$N_{com(s1)}$	0.46342	0.48069	0.47177
n_{len}	No	1	$N_{com(s1)}$	0.45674	0.47551	0.46582
n_{len}	No	X	$N_{com(s1+s2)}$	0.45701	0.47320	0.46484

Table 6. Results for experimental procedure 2 with $\beta = 3$.

Term Selection	Term Weighting		Sentence Selection	Results		
				Recall	Precision	Fmeasure
$n\text{-}grams_{mfs}$	stop-words	$p_i(t_j)$				
n_{len}	no	$X \cdot Y$	N_{sent}	0.43470	0.45120	0.44247
n_{len}	no	1	N_{sent}	**0.43701**	**0.45310**	**0.44459**
n_{len}	no	X	N_{sent}	0.43470	0.45120	0.44247
n_{len}	no	$X \cdot X$	N_{sent}	0.42686	0.44463	0.43525
1	no	1	N_{sent}	0.38367	0.40290	0.39291
1	no	Z	N_{sent}	**0.44397**	**0.45773**	**0.45062**
1	no	$Z \cdot Z$	N_{sent}	0.43797	0.45220	0.44485
1	no	Z	$N_{com(s1)}$	**0.46622**	**0.48407**	**0.47486**
1	no	Z	$N_{com(s1+s2)}$	0.46223	0.47806	0.46989
n_{len}	no	X	$N_{com(s1)}$	0.46631	0.48392	0.47483
n_{len}	no	1	$N_{com(s1)}$	0.45674	0.47551	0.46582
n_{len}	no	X	$N_{com(s1+s2)}$	0.46007	0.47638	0.46796

Table 7. Results for experimental procedure 2 with $\beta = 4$.

Term Selection	Term Weighting		Sentence Selection	Results		
$n\text{-}grams_{mfs}$	stop-words	$p_i(t_j)$		Recall	Precision	Fmeasure
n_{len}	no	$X \cdot Y$	N_{sent}	0.43013	0.44680	0.43812
n_{len}	no	1	N_{sent}	**0.43266**	**0.44861**	**0.44025**
n_{len}	no	X	N_{sent}	0.43013	0.44680	0.43812
n_{len}	no	$X \cdot X$	N_{sent}	0.42354	0.44084	0.43183
1	no	1	N_{sent}	0.44631	0.46505	0.45536
1	no	Z	N_{sent}	**0.44090**	**0.45509**	**0.44776**
1	no	$Z \cdot Z$	N_{sent}	0.43712	0.45138	0.44402
1	no	Z	$N_{com(s1)}$	**0.46788**	**0.48537**	**0.47634**
1	no	Z	$N_{com(s1+s2)}$	0.46397	0.47985	0.47165
n_{len}	no	X	$N_{com(s1)}$	0.46568	0.48373	0.47441
n_{len}	no	1	$N_{com(s1)}$	0.45674	0.47551	0.46582
n_{len}	no	X	$N_{com(s1+s2)}$	0.45977	0.47604	0.46734

Experimental Procedure 3. In previous experimental procedure, we obtained better results with proposed schemes using different thresholds. Here, we compare best results of the previous experimental procedures (see Tables 8 – 10). We detect that the best configuration for MFSs as selected terms was obtained with combination of threshold ($\beta = 2, 3, 4$). Also, we detect that for the terms derived from MFSs, the best threshold is $\beta = 2$. The results of the configuration with combination of sentences with $\beta = 4$ is the best obtained result.

Table 8. Comparison of results between proposed schemes (terms are MFS).

Method	Recall	Precision	F-measure
Proposed with $\beta = 2, 3, 4$	**0.44128**	**0.45609**	**0.44840**
Proposed with $\beta = 2$	0.43749	0.45182	0.44438
Proposed with $\beta = 3$	0.43701	0.45310	0.44459
Proposed with $\beta = 4$	0.43266	0.44861	0.44025

Table 9. Comparison of results between proposed schemes (terms derived from MFS).

Method	Recall	Precision	F-measure
Proposed with $\beta = 2, 3, 4$	0.44582	0.45820	0.45181
Proposed with $\beta = 2$	**0.44659**	**0.45968**	**0.45293**
Proposed with $\beta = 3$	0.44397	0.45773	0.45062
Proposed with $\beta = 4$	0.44090	0.45509	0.44776

Table 10. Comparison of results between proposed schemes (combination of sentences).

Method	Recall	Precision	F-measure
Proposed with $\beta = 2, 3, 4$	0.46576	0.48278	0.47399
Proposed with $\beta = 2$	0.46536	0.48230	0.47355
Proposed with $\beta = 3$	0.46622	0.48407	0.47486
Proposed with $\beta = 4$	**0.46788**	**0.48537**	**0.47634**

Experimental Procedure 4. There are only five better systems [16] than *baseline* with little differences of the results. In previous experimental procedures, we obtained better results than *baseline*. For the collection of DUC2002 the results of baseline configuration is very high because the majority of texts is consisted of news descriptions and in such type of texts is common that the first sentences describe briefly the given news. In other words, some of the first sentences are abstract or summary of a given file. In other types of texts the configuration of *baseline* will not work at all. So, it is fair to compare with the state-of-the-art methods like Random Walks [10]. The author of this work provided the data of its summaries which were evaluated in the same conditions as proposed methods. Specifically, *DirectedBackward* version of *TextRank* was evaluated (see Table 11, *TextRank*). Finally, the best of the proposed methods is included.

Table 11. Results with other methods.

Additional info	Method	Recall	Precision	F-measure
None	Baseline: *random*	0.37892	0.39816	0.38817
	TextRank: [10]	**0.45220**	0.43487	0.44320
	Proposed: *Z, best*	0.44659	**0.45968**	**0.45293**
Order of sentences	Baseline: *first*	0.46407	0.48240	0.47294
	Proposed: *Z, 1best+first*	**0.46788**	**0.48537**	**0.47634**

We can see that the proposed method is better than *baseline*, which is difficult task because of special baseline configuration. Comparing with *TextRank* we can deduce that we locate in the-state-of-the art methods for single text summarization.

6. Conclusions

We proposed new method for the automatic generation of text summaries for a single document based on the discovery of MFSs. New method permits to generate text summaries in language and dominion-independent ways. We tested different combinations of term selection, term weighting, sentence weighting and sentence selection schemes with different thresholds. With first experiment, we observed that MFSs are good terms and help us to obtain good results comparing with words and n-grams. In the second experiment, we tested the proposed schemes with different

thresholds. With the third experiment, we compare the best results obtained in the second procedure. We conclude that words derived from MFSs are the best terms with $\beta = 2$ and MFSs are good terms with $\beta = 2, 3, 4$. Also we observed that term weighting scheme using terms derived from MFSs gives us betters results.

Finally, combining the best configuration of term selection, term weighting, and sentence selection schemes, we obtained the results superior to the-state-of-the-art methods: n-grams, *Textrank* and *baseline*, in spite of special baseline configuration for news articles. In our future work, we explore new multiword descriptions in order to improve the-state-of-the-art methods.

Acknowledgments
This work was done under the partial support of Mexican Government (CONACyT, SNI, SIP-IPN, COTEPABE-IPN, COFAA-IPN, PROMEP). We also thank CIC-IPN and UAEM for their assistance.

References

1. Lin, C.Y., Hovy, E.: Automated Text Summarization in SUMMARIST. In: Proceedings of ACL Workshop on Intelligent, Scalable Text Summarization, Madrid, Spain, 1997.
2. Kupiec, J., Pedersen, J.O., Chen, F.: A Trainable Document Summarizer. In: Proceedings of the 18th ACM-SIGIR Conference on Research and Development in Information Retrieval, Seattle, pp. 68–73, 1995.
3. Liu, D., *et al.*: Multi-Document Summarization Based on BE-Vector Clustering. In: Gelbukh, A. (ed.) CICLing 2006, LNCS, vol. 3878, pp. 470-479, Springer-Verlag 2006.
4. Chuang, T.W., Yang, J.: Text Summarization by Sentence Segment Extraction Using Machine Learning Algorithms. In: Proceedings of the ACL 2004 Workshop, España, 2004.
5. Xu, W., Li, W., *et al.* Deriving Event Relevance from the Ontology Constructed with Formal Concept Analysis. CICLing 2006, LNCS, vol. 3878, pp. 480-489, Springer-Verlag 2006.
6. Villatoro-Tello, E., Villaseñor-Pineda, L., Montes-y-Gómez, M.: Using Word Sequences for Text Summarization. In: Sojka, P., Kopecek, I., Pala, K. (eds.) TSD 2006. LNAI, vol. 4188, pp. 293-300, Springer-Verlag 2006.
7. Ledeneva, Y., Gelbukh, A., García Hernández, R.. Terms Derived from Frequent Sequences for Extractive Text Summarization. In: Gelbukh, A. (Ed.): CICLing 2008, LNCS 4919, pp. 593-604, Springer-Verlag 2008.
8. Song, Y., *et al.*: A Term Weighting Method based on Lexical Chain for Automatic Summarization. In: Gelbukh, A. (ed.) CICLing 2004, LNCS, vol. 2945, pp. 636-639, Springer-Verlag 2004.
9. Cristea D., *et al.*: Summarization through Discourse Structure. In: Gelbukh, A. (ed.) CICLing 2005, LNCS, vol. 3406, pp. 621-632, Springer-Verlag 2005.
10. Mihalcea, R.: Random Walks on Text Structures. In: Gelbukh, A. (ed.) CICLing 2006, LNCS, vol. 3878, pp. 249–262, Springer-Verlag 2006.
11. Mihalcea, R., Tarau, P. TextRank: Bringing Order into Texts. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Spain, 2004.
12. García-Hernández, R.A., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A.: A Fast Algorithm to Find All the Maximal Frequent Sequences in a Text, In: Sanfeliu, A.,

Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A. (eds.) CIARP 2004. LNCS vol. 3287, pp. 478–486, Springer-Verlag 2004.
13. García-Hernández, R.A., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A.: A New Algorithm for Fast Discovery of Maximal Sequential Patterns in a Document Collection. In: Gelbukh, A. (ed.) CICLing 2006, LNCS vol. 3878, pp. 514–523, Springer-Verlag 2006.
14. Hernández-Reyes E., García-Hernández R. A., Martínez-Trinidad J. F., and Carrasco-Ochoa J. A. Document Clustering based on Maximal Frequent Sequences, 5th International Conference on Natural Language Processing, LNCS (FinTal 2006), Springer-Verlag 2006.
15. Denicia-Carral C., Montes-y-Gómez M., Villaseñor-Pineda L., García-Hernández. R. Text Mining Approach for Definition Question Answering, 5th International Conference on Natural Language Processing (FinTal 2006), LNCS, Springer-Verlag 2006.
16. DUC. Document understanding conference 2002; www-nlpir.nist.gov/projects/duc.
17. Lin, C.Y.: ROUGE: A Package for Automatic Evaluation of Summaries. In: Proceedings of Workshop on Text Summarization of ACL, Spain, 2004.