ORIGINAL RESEARCH

The role of statistical and semantic features in single-document extractive summarization

Tatiana Vodolazova, Elena Lloret, Rafael Muñoz, Manuel Palomar

Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Spain

Correspondence: Tatiana Vodolazova. Address: Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Spain. Email: naiadhe@gmail.com.

Received: June 30, 2012 **Accepted:** October 22, 2012 **Online Published:** April 10, 2013
DOI: 10.5430/air.v2n3p35 **URL:** http://dx.doi.org/10.5430/air.v2n3p35

Abstract

This paper reports on the further results of the ongoing research analyzing the impact of a range of commonly used statistical and semantic features in the context of extractive text summarization. The features experimented with include word frequency, inverse sentence and term frequencies, stopwords filtering, word senses, resolved anaphora and textual entailment. The obtained results demonstrate the relative importance of each feature and the limitations of the tools available. It has been shown that the inverse sentence frequency combined with the term frequency yields almost the same results as the latter combined with stopwords filtering that in its turn proved to be a highly competitive baseline. To improve the suboptimal results of anaphora resolution, the system was extended with the second anaphora resolution module. The present paper also describes the first attempts of the internal document data representation.

Key words

Extractive text summarization, Semantics, Statistics, Coreference resolution

1 Introduction

The research in extractive Text Summarization (TS) covers a wide range of features that are used to determine the most salient text segments to include them into the final summary. Different approaches select different features and methods, starting from the very basic ones like term frequency [1], position of the sentence within the original document [2, 3], assigning higher weights to the sentences containing terms of the title [2] and inverse sentence frequency [4]; or more complex ones including word sense disambiguation [5], latent semantic analysis and anaphora resolution [6], textual entailment [7]. However, each of the above mentioned system only focuses on a few distinct features, usually two or three. The aim of present research is to assess the relative importance of a set of different features and their impact on the process of extractive summarization generation. The inspected set of features and methods include term frequency, inverse term and sentence frequencies, word sense disambiguation, anaphora resolution, textual entailment recognition and corpus-tailored stopwords.

The initial work [8] was focused on the impact of corpus-tailored stopwords on the process of TS and its integration with the abovementioned features. It was shown that some methods, for example, anaphora resolution implemented using JavaRAP [9], need improvement. The present paper reports on the further results of the ongoing research. The selected features were combined in a slightly different manner with the list of 350 common stopwords of English. The system performance was also tested without the stopwords filtering. BART coreference resolution tool [9] was integrated to compare the results with the Java RAP results.

The final goal of the current research is to identify the features and tools that benefit TS the most with the further objective to use them for abstractive TS. Abstractive TS would involve transforming the text data to an internal semantic data representation. The present paper describes the data representation that was used in the experiments. It was designed to simplify the transition from the term-based data representation used now to the concept representation.

This paper is organized as follows. Section 2 briefly introduces the selected features and methods. Section 3 describes the internal data representation. System settings and architecture are introduced in Section 4 together with the evaluation environment. The results are reported in Section 5. Finally, the conclusions together with the future work can be found in Section 6.

2 Selected features and related work

2.1 Term frequency

Term frequency (TF) is one of the commonly used features in the framework of automatic TS. It is easy to obtain, as it doesn't involve complex data preprocessing, and yields fairly competitive results. The impact of TF isolated from all the other features was analyzed in [11]. In the context of multi-document summarization it is usually combined with the inverse document frequency [12].

For a sentence $S_j = \{t_1, t_2, \dots, t_m\}$ with m tokens, its score based on TF will be calculated using the following formula:

$$Sc_{tf(S_j)} = \frac{\sum_{i=1}^{m} tf_i}{n} \quad (1)$$

where tf_i is the frequency of the i_{th} token (or its stem/lemma) in the text j; n is the number of sentence tokens (stopwords removed).

2.2 Inverse term and sentence frequencies

In the context of single document summarization the inverse document frequency is usually substituted by the inverse sentence frequency (ISF), as a single document summary does not depend on the other documents of the collection. Blake [13] compares different language models and speculates on using ISF for the systems aiming at sentence extraction and inverse term frequency (ITF) for the systems identifying terms as their smallest compositional unit. Thus for the TS systems based on unigrams ITF could be a reliable method to select the sentences for the final summary.

Roughly based on [13] the ISF and ITF measures were calculated the following way:

$$isf_{tf} = log \frac{|S|}{|\{s \in S : t \in s\}|} \quad (2)$$

where $|S|$ is the total number of sentences in the document; $|\{s \in S : t \in s\}|$ number of sentences where the term t appears.

$$itf = log \frac{|V|}{|\{t \in V : t \in d\}|} \quad (3)$$

where $|V|$ is the vocabulary size of the document; $|\{t \in V : t \in d\}|$ number of times where the term t appears in the document, i.e. term frequency.

However, the logical interpretation of ITF is to select the most rare and thus informative terms. Since TF yields rather good summaries, ITF as opposed to it, must result in a poor sentence selection. We integrate the ITF measure to verify this hypothesis.

2.3 Word sense disambiguation (WSD)

The information about word senses can help to capture the cases of synonymy as between the nouns “a house” and “a building”. When all the synonyms in a document are substituted with the selected synset representative, the semantics of the text is brought to the surface and can be captured during the scoring stage by any of the statistical scoring methods employed. The TF is substituted by the concept frequency (CF) and Formulas (1), (2) and (3) are modified the following way:

$$Sc_{cf(s_j)} = \frac{\sum_{i=1}^{m} cf_i}{n}, \tag{4}$$

where S_j is the j_{th} sentence $S_j = \{t_1, t_2, \ldots, t_m\}$ with m tokens; cf_i is the frequency of the WordNet [14] synset id in the document that the sense of the i_{th} term belongs to; n is the number of sentence tokens (stopwords removed).

$$isf_{cf} = log\frac{|S|}{|\{s \in S : c \in s\}|}, \tag{5}$$

where $|S|$ is the total number of sentences in the document; $|\{s \in S : c \in s\}|$ number of sentences s where the WordNet synset id c in the document appears, for all $c \in d$.

$$icf = log\frac{|V|}{|\{t \in V : c \in d\}|}, \tag{6}$$

where $|V|$ is the vocabulary size of the document as measured in the number of different WordNet synset ids appearing in the text; $|\{c \in V : c \in d\}|$ number of times where the concept c appears in the document, i.e. concept frequency.

It was shown in [5] that disambiguating verbs decreases system performance. In our experiments we tested word sense disambiguation applied to both only nouns and nouns, verbs and adjectives.

2.4 Anaphora resolution

Resolving pronominal anaphora can particularly benefit text summarization techniques involving lexical information. The previous work on including anaphora resolution (AR) reported some improvement over the lexical LSA system [6]. However, not all the systems equally benefit from including AR method. Vodolazova et al. [8] integrated JavaRAP [9] to substitute anaphoric expressions by their antecedents and reported that AR per se decreases the quality of final summaries. But applying AR to the original text data prior to redundancy detection carried out using of textual entailment was shown to improve the quality of final summaries as compared to the system using AR or TE only.

2.5 Textual entailment

As opposed to the paraphrasing that aims in capturing the natural language expressions that convey the same information, the Textual Entailment (TE) methods are designed to recognize the semantic inference between those expressions. Together with automatic paraphrasing [15] it offers a promising technique to eliminate redundant information. Lloret et al. [7] reported on significant increase in ROUGE-1 values over the TF baseline when TE was used to identify the redundant information in a text. Similarly, Vodolazova et al. [8] report that the best results are obtained when TE is included into the process of summary generation.

2.6 Stopwords filtering

Stopwords (SW) filtering benefits a wide range of Natural Language Processing (NLP) applications. The very first automatic summarization approach [1] mentions that the noise introduced by the presence of too common words can be eliminated using a stored common words list. However, not all the extractive TS approached equally benefit from the stopwords filtering. Ledeneva et al. [16] have shown that removing the stopwords yields worse results for TS systems based on the multiword descriptions.

3 Data representation

Prior to summary generation the input data is analyzed employing the common preprocessing steps. Those include sentence boundary detection, tokenization, stemming, part-of-speech tagging and lemmatization. All the abovementioned steps are carried out with the help of the Freeling toolkit [17]. The preprocessing step also involves the stopwords identification, WSD, computing term and concept frequencies for each single token of the input document. All the collected information is stored using an internal data representation. The hierarchy of the Representation objects is illustrated in Figure 1. For now the following subtypes of the top Representation object have been implemented: Sentence, Basic, Content Word, Noun, Verb and Adjective Representations. In the context of extractive TS each input document is represented as a list of Sentence Representations. A Sentence Representation stores the original sentence text and the list of its tokens. All the verbs, nouns and adjectives that are not present on the list of stopwords are the instances of the Content Word Representation and the Verb, Noun and Adjective Representations respectively. All the remaining tokens are the instances of the Basic Representation. The Basic Representation includes information about token lemma, stem, part of speech and frequency. The Content Word Representation instances additionally contain information about the WordNet [14] synset id and the concept frequency of this synset id in the input text.

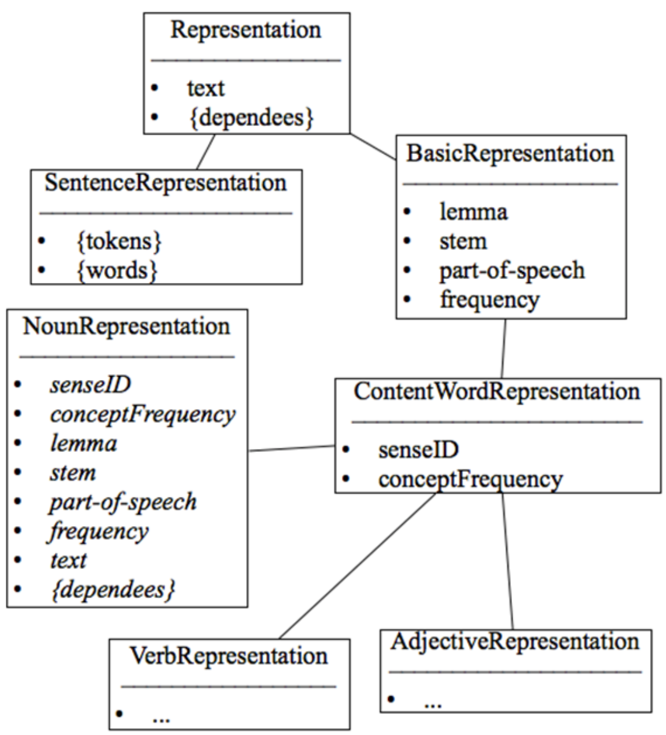

Figure 1. Internal Data Representation

4 Evaluation environment

4.1 System architecture and settings

The designed system consists of a set of modules that interact with each other in a complex way. They include the Scoring, the TE, the AR and the WSD modules. The allowed modules combinations are illustrated in Figure 2. The two obligatory steps include transforming the input text into the internal data representation as described in the Section 3 and the sentence scoring carried out by the Scoring module.

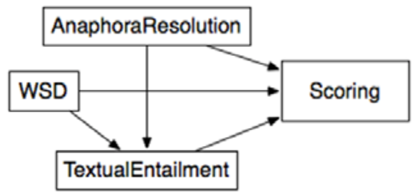

Figure 2. Modules interaction

4.1.1 The scoring module

The Scoring module in its basic setting computes the score of each sentence according to its term frequency (see Formula 1). The TFs are extracted using the Word Vector Tool (Note 2). The Scoring module arranges the original text sentences in the descending order. The top N sentences are further selected for the final summary, where N depends on the desired final summary size. For the present research all the summaries were set not to extend the 100 words size. The Scoring module can be set to include the stopwords filtering or not. The other statistical scoring strategies include the ISF and ITF as described in Formulas 2 and 3. Since the Scoring module rates sentences based on terms, in the case of ISF the terms weights are computed using the TF/ISF of CF/ISF.

4.1.2 The WSD module

The more sophisticated scoring strategy involves WSD. The present system employs the WSD modules integrated into Freeling. Freeling provides a number of WSD algorithms. We integrated the most frequent sense (MFS) and the PageRank-based (UKB) algorithms [18] in our TS system. When the information about the word senses is included the Scoring module rates the sentences based on the CF. For both MFS and UKB based algorithms the scores are computed as in the Formula 4. Formulas 5 and 6 represent the variations of the standard ISF and ITF based on the concept frequency and are also used for scoring when the Scoring module is supplied with the word senses.

The final aim of any TS system is to convey the main topics of the analyzed input text in a concise manner. The topic of a document is usually associated with the nouns phrases and to less extend with the verb phrases. Thus is can be hypothesized that applying WSD to only nouns should yield better results than doing so for verbs and adjectives. The previous research on the impact of WSD in the framework of TS confirms this hypothesis [5]. The implemented Scoring module allows to rate both adjectives, nouns and verbs and nouns only based on the CF.

4.1.3 The AR module

The anaphoric expressions can be resolved prior to scoring. Vodolazova et al. [8] report on the poor results when the anaphora resolution is included. To verify the negative impact of AR on TS we integrated an additional coreference resolution system BART [10]. Previously used JavaRAP produces a list of pairs of anaphor with its antecedent. As opposed to JavaRAP, the BART system produces a set of coreference chains. The heuristics applied to the BART output was to substitute each member of such a chain by the longest chain representative. Here is an example chain generated by BART:

$$[Caroline\ Smedvig, \qquad she, \qquad the\ composer]$$

The proper name “Caroline Smedvig” is the longest character sequence, thus all the occurrences of the chain members in the text are substituted by “Caroline Smedvig”.

4.1.4 The TE module

TE is used to eliminate the sentences with repeated information and thus it is applied prior to scoring. TE can be combined with WSD. The TE system used for the present research is described in [7]. It was extended with the preprocessing stage where all the representatives of the same WordNet synset are substituted by one and the same synset member. Only after the substitution is carried out the sentences are processed by the TE recognizer. The recognizer takes each of the sentences of the original text in sequential order and compares it to the previously seen sentences available on the sentence stack. If the current sentence can be inferred it is discarded, otherwise pushed on to the sentence stack.

The TE module can be combined with the AR module. In this case the TE recognizer works with the text where anaphoric expressions have already been replaced by their antecedents.

4.1.5 System Settings

The following system settings were tested in this research:

- NOSW: not using the stopwords filtering
- NOSW ITF: NOSW combined with the ITF
- NOSW ISF: NOSW combined with the ISF
- NOSW AR: NOSW combined with the AR
- NOSW TE: NOSW combined with the TE
- NOSW WSD TE: do not use the stopwords filtering, replace the words of the selected parts of speech with the same member of the WordNet synset, and then process the result using the TE module. The scoring module is applied to the original version of remaining sentences, thus making it possible to evaluate the impact of both MFS and UKB algorithms on the resulted data.
- NOSW ISF AR WSD TE: is similar to the previous one with the difference that AR is applied before the WSD and ISF is used for scoring. The scoring module is again applied to the original version of the remaining sentences.

– SSW: using the standard stopword list

– SSW AR: SSW combined with the AR

– SSW AR-BART: SSW combined with the BART AR system

– SSW TE: SSW combined with the TE

– SSW WSD TE: similar to the NOSW WSD TE, but using the SSW for filtering

– SSW AR WSD TE: similar to the NOSW ISF AR WSD TE, but using the SSW for filtering and the simple term or concept frequency for scoring.

Each of the listed settings is further combined with either TF or CF scoring strategy as shown in the Table 1 (Note 3).

Vodolazova et al. [8] reported on the poor results when the stopwords filtering is combined with the inverse sentence and term frequencies, thus the introduced set of experiments does not include the ISF and ITF settings for the SSW option.

Table 1. ROUGE-1 Recall values using the standard stopword list for filtering and without stopwords filtering

	TF	CF-MFS		CF-UKB	
		NVA	N	NVA	N
NOSW	0.39467	0.39323	0.39541	0.39394	0.39666
NOSW ITF	0.39676	0.39555	0.39555	0.39872	0.39872
NOSW ISF	0.40629	0.40395	0.40395	0.40293	0.40293
NOSW AR	0.38260	0.38312	0.38353	0.38371	0.38334
NOSW TE	0.39509	0.39740	0.39885	0.39683	0.39894
NOSW WSD TE	0.39509	0.39556	0.39685	0.39586	0.39780
NOSW ISF AR WSD TE	0.40325	0.40497	0.40497	0.40396	0.40396
SSW	0.40906	0.40869	0.41004	0.41100	0.40718
SSW AR	0.40432	0.40432	0.40207	0.40628	0.40264
SSW AR-BART	0.31395	0.31619	0.32439	0.32443	0.32556
SSW TE	0.41027	0.40965	0.40951	0.41125	0.40917
SSW WSD TE	0.41027	0.41031	0.41061	0.41168	0.40873
SSW AR WSD TE	0.42413*	0.42339*	0.42379*	0.42556*	0.42350*

4.2 Evaluation corpus and metrics

The developed system was tested on the data for single-document summarization task of the Document Understanding Conference 2002 (Note 1). The standard evaluation metrics of ROUGE [19] was used to assess the quality of generated summaries as compared to the human model summaries provided by DUC 2002. In the present research we include both recall and precision values for the ROUGE-1.

5 Results and discussion

The summaries generated for all the system settings described in Section 4.1.5 were evaluated using the ROUGE metrics. The evaluation results for ROUGE-1 recall and precision are provided in the Tables 1 and 2 respectively (Note 3).

Table 2. ROUGE-1 Precision values using the standard stopword list for filtering and without stopwords filtering

	TF	CF-MFS		CF-UKB	
		NVA	N	NVA	N
NOSW	0.43646	0.43539	0.43725	0.43641	0.43839
NOSW ITF	0.44647	0.44582	0.44582	0.44934	0.44934
NOSW ISF	0.46192	0.46033	0.46033	0.45901	0.45901
NOSW AR	0.41542	0.41593	0.41646	0.41645	0.41615
NOSW TE	0.44715	0.44208	0.44319	0.44191	0.44359
NOSW WSD TE	0.44715	0.44645	0.44924	0.44785	0.45061
NOSW ISF AR WSD TE	0.43495	0.43608	0.43608	0.43589	0.43589
SSW	0.45863	0.45693	0.45901	0.45994	0.45706
SSW AR	0.43804	0.43804	0.43501	0.44072	0.43569
SSW AR-BART	0.33287	0.33494	0.34363	0.34347	0.34498
SSW TE	0.46369	0.46268	0.46350	0.46419	0.46359
SSW WSD TE	0.46369	0.46374	0.46497	0.46513	0.46348
SSW AR WSD TE	0.45988	0.45902	0.45960	0.46188	0.45979

The combination of stopwords filtering and TF-based scoring represent a simple approach to tackle the TS task that nevertheless yields fairly good results. It was decided to selects this feature combination as the baseline. 0.40906 is slightly lower than the DUC baseline of 0.41132 [6]. Despite this fact none of the system settings without the stopwords filtering could outperform it. Even the combination of AR, WSD and TE could not reach it. It can be concluded that for the TS systems based on the unigrams as opposed to the multiword descriptions [16] stopwords filtering is essential. The best result in this range of settings is 0.40629. It was obtained calculating the plain ISF as in Formula 2. Combining ISF and concept frequency (Formula 5) also yields very competitive results for this range of settings and outperforms the more sophisticated setup including the WSD and TE. However, combining ISF with AR, WSD and TE decreases the quality of generated summaries. The same tendency is observed in the ROUGE-1 precision values for the ISF: they reveal even stronger contrast to the remaining figures in that range. The precision value of 0.46192 is not only the best value in the range, but also outperforms the baseline. Thus in the absence of stopword list, ISF can be used as the basic summary generation approach.

Another tendency revealed in the Tables 1 and 2 concerns CF- vs. TF-based scoring methods. In most of the settings, both involving stopwords filtering and not, at least one of the CF-based scoring heuristics outperforms the TF-based scoring result. And in most of the cases the best CF technique is based on the UKB WSD algorithm. As for the N-based CF scoring as opposed to the NVA-based one the results differ between the two WSD algorithms. The recall values for the settings with the SW filtering show that in 4 out of 6 cases the MFS-N performs better than the MFS-NVA. And in 5 out of 6 cases the UKB-NVA yields better results than the UKB-N CF scoring. Also, without the SW filtering WSD TE performs worse than the simple TE. But once the SWs are removed WSD TE outperforms the simple TE.

The worst results in term of both precision and recall values were obtained for the plain AR setting. The BART AR module does not improve over the JavaRAP results. This is due to the particular substitution heuristics applied to the coreference chains produced by BART system. The example below illustrates one of such coreference chains:

$$[a\ member, \quad he, \quad a\ protege, \quad Tanglewood\ founder\ Serge\ Koussevitzky, \quad Koussevitzky, \quad young\ Bernstein, \quad his\ successor, \quad Charles\ Munch, \quad John\ Williams, \quad John\ Mauceri, \quad Michael\ Tillson\ Thomas, \quad his, \quad Ozawa, \quad Bernstein]$$

We applied the longest member substitution (LMS) heuristics, such that all the remaining members of the chain are substituted by the longest chain member. The combination of the erroneously identified chain members, as in the example above, with the LMS heuristics cause these poor results for AR_BART settings. Instead of substituting with the longest member, the coreference chain needs to be analyzed to identify the proper nouns of the maximal length of 2-3 tokens and only the pronouns of the chain to be substituted.

Table 3. ROUGE-1 Recall values using the extended stopwords list [8]

	TF	CF-MFS		CF-UKB	
		NVA	N	NVA	N
ASW	0.41779	0.40976	0.41466	0.41785	0.41462
ASW ITF	0.36924	0.36828	0.36668	0.36890	0.36719
ASW ISF	0.38126	0.37985	0.37894	0.37924	0.37562
ASW AR	0.38945	0.38873	0.39077	0.39146	0.38991
ASW TE	0.41804	0.41807	0.41596	0.41897	0.41665
ASW WSD TE	0.41807	0.41796	0.41627	0.41894	0.41843
ASW AR WSD TE	0.43235*	0.43050*	0.42963*	0.43196*	0.43017*

The first results of current research were reported in reference 8. The system settings analyzed involved extending SW list with the most frequent words of the DUC 2002 data. Table 3 contains results for these settings. Except for the settings without the SW filtering where it yielded the second best results, the combination of AR and WSD TE generates the best summaries. The statistically significant results are indicated with the asterisks. However, both AR and WSD TE benefit more the process of redundancy detection. Table 3 demonstrates that TF-based scoring for the ASW AR WSD TE yields better results than any of CF-based scoring methods. And for the SSW AR WSD TE TF-based scoring takes the second best result after the UKB-based one.

6 Conclusions and future work

The goal of the present research is to study the interaction between a set of statistical and semantic features and their impact on the process of extractive text summarization with the final objective of selecting the most significant features and tactics to advance from extractive TS to abstractive TS. The obtained results have shown that sematic-based methods involving anaphora resolution, textual entailment and word sense disambiguation benefit the redundancy detection stage. Once the redundant information is detected and discarded, the statistical methods, such as term frequency and inverse sentence frequency offer a better machinery to select the most representative sentences to be included in the final summary.

A slightly different direction of the present research is concerned with the internal data representation. To draw the connection with the results obtained in the feature study we are planning to include the semantic features, such as resolved anaphora and word senses, on the level of data representation. This will allow us to carry out the redundancy detection on the concept level in the process of abstractive TS.

It has been shown that in the context of unigram-based text summarization approach as opposed to the multiword sequences, the stopwords filtering is essential.

The future work will also continue improving the anaphora resolution module.

References

[1] Luhn, H. P. The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development. 1958; 2(2): 159-165. http://dx.doi.org/10.1147/rd.22.0159

[2] Edmundson, H. P. New Methods in Automatic Extracting. Journal of the Association for Com-puting Machinery. 1969; 16(2): 264-285. http://dx.doi.org/10.1145/321510.321519

[3] Lin, C.-Y., & Hovy, E. Identifying topics by position. Proceedings of the fifth conference on Applied natural language processing. 1997: 283-290. Morristown, NJ, USA: Association for Computational Linguistics. http://dx.doi.org/10.3115/974557.974599

[4] Larocca Neto, J., Santos, A. D., Kaestner, C. A. A., & Freitas, A. A. Document Clustering and Text Summarization. In N. Mackin (Ed.), Proceedings of the 4th International Conference Practical Applications of Knowledge Discovery and Data Mining (PADD-2000). 2000: 41-55). London: The Practical Application Company.

[5] Plaza, L., & Díaz, A. Using Semantic Graphs and Word Sense Disambiguation Techniques to Improve Text Summarization. Procesamiento de Lenguaje Natural. 2011; 47: 97-105.

[6] Steinberger, J., Poesio, M., Kabadjov, M. A., & Ježek, K. Two Uses of Anaphora Resolution in Summarization. Information Processing & Management. 2007; 43(6): 1663-1680. http://dx.doi.org/10.1016/j.ipm.2007.01.010

[7] Lloret, E., Ferráandez, Ó., Muñoz, R., & Palomar, M. A Text Summarization Approach under the Influence of Textual Entailment. Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science. 2008: 22-31.

[8] Vodolazova, T., Lloret, E., Muñoz, R., & Palomar, M. A Comparative Study of the Impact of Statistical and Semantic Features in the Framework of Extractive Text Summarization. In P. Sojka, A. Horák, I. Kopecek, & P. Karel (Eds.), Text, Speech and Dialogue, 15th International Conference (TSD 2012). 2012; 7499: 306–313. Brno, Czech Republic: Springer.

[9] Qiu, L., Kan, M.-Y., & Chua, T.-S. A Public Reference Implementation of the RAP Anaphora Resolution Algorithm. Proceedings of the 4th Language Resources and Evaluation Conference (LREC 2004). 2004; 2: 291-294). Lisboa, Potugal.

[10] Versley, Y., Ponzetto, S. P., Poesio, M., Eidelman, V., Jern, A., Smith, J., Yang, X., et al. (2008). BART: A Modular Toolkit for Coreference Resolution. Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008) (pp. 1-4).

[11] Gupta, V., & Lehal, G. S. A Survey of Text Summarization Extractive Techniques. Journal of Emerging Technologies in Web Intelligence. 2010; 2(3): 258-268. http://dx.doi.org/10.4304/jetwi.2.3.258-268

[12] Nenkova, A., Vanderwende, L., & McKeown, K. 2006. A Compositional Context Sensitive Multi-Document Summarizer. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’06 (p. 573). New York, New York, USA: ACM Press. http://dx.doi.org/10.1145/1148170.1148269

[13] Blake, C. 2006. A Comparison of Document, Sentence, and Term Event Spaces. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (pp. 601-608). Association for Computational Linguistics Stroudsburg, PA, USA

[14] Fellbaum, C. 1998. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

[15] Madnani, N., Zajic, D., Dorr, B., & Ayan, N. 2007. Multiple Alternative Sentence Compressions for Automatic Text Summarization. Proceedings of DUC, (April).

[16] Ledeneva, Y., Gelbukh, A., & García-Hernández, R. A. 2008. Effect of preprocessing on extractive summarization with maximal frequent sequences. In A. Gelbukh & E. Morales (Eds.), Proceedings of the 7th Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence (MICAI' 08). Atizapán de Zaragoza, Mexico: Springer-Verlag Berlin, Heidelberg.

[17] Carreras, X., Chao, I., Padró, L., & Padró, M. Freeling: An Open-Source Suite of Language Analyzers. Proceedings of the 4th Language Resources and Evaluation Conference (LREC 2004). 2004; 4: 239-242. PMid:15511719

[18] Agirre, E., & Soroa, A. Personalizing PageRank for Word Sense Disambiguation. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics on – EACL. 2009: 33-41. Morristown, NJ, USA: Association for Computational Linguistics. http://dx.doi.org/10.3115/1609067.1609070

[19] Lin, C.-Y., & Hovy, E. Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL '03). 2003; 1: 71-78. Edmonton, Canada: Association for Computational Linguistics. http://dx.doi.org/10.3115/1073445.1073465