# Extractive Text Summarization: Can We Use the Same Techniques for Any Text?

Tatiana Vodolazova, Elena Lloret, Rafael Muñoz, and Manuel Palomar

Dept. Lenguajes y Sistemas Informáticos, Universidad de Alicante
Apdo de correos, 99, E-03080, Alicante, Spain
{tvodolazova,elloret,rafael,mpalomar}@dlsi.ua.es

**Abstract.** In this paper we address two issues. The first one analyzes whether the performance of a text summarization method depends on the topic of a document. The second one is concerned with how certain linguistic properties of a text may affect the performance of a number of automatic text summarization methods. For this we consider semantic analysis methods, such as textual entailment and anaphora resolution, and we study how they are related to proper noun, pronoun and noun ratios calculated over original documents that are grouped into related topics. Given the obtained results, we can conclude that although our first hypothesis is not supported, since it has been found no evident relationship between the topic of a document and the performance of the methods employed, adapting summarization systems to the linguistic properties of input documents benefits the process of summarization.

**Keywords:** text summarization, textual entailment, anaphora resolution.

## 1 Introduction

The first attempts to tackle the task of automatic text summarization were made as early as in the middle of the past century [17]. Since then the capabilities of modern hardware have increased enormously. However, nowadays when we talk about automatic text summarization we mostly focus on extractive summaries and hope to top the threshold of 50% on the recall [22]. Extractive summaries, as opposed to the abstractive ones that involve natural language generation techniques, consist of segments of the original text. The task sounds less challenging than it has been proven to be [22].

The extractive summarization systems developed so far have been tested on a number of different corpora [22]. There has been a significant number of systems proposed for the task of summarization of the newswire articles. Many of those systems emerged due to the Document Understanding Conference challenges (DUC)[1] [8,25]. Even that the last challenge has been held in 2007, the DUC data is still being used in research [15,16,24,26]. Some experiments were done

[1] http://duc.nist.gov/

E. Métais et al. (Eds.): NLDB 2013, LNCS 7934, pp. 164–175, 2013.

with the Reuters newswire corpus [2]. The short newswire articles differ from fiction. The summarization systems that target this niche have adapted to the particular characteristics of fiction. There has been research on short fiction summarization [12], fairy tales [16], whole books [19], etc. Due to the rapid growth of the amounts of web data, the need to summarize becomes even more acute. More recent research has focused on Web 2.0 textual genres, such as forum [30] and blog [11] summarization. The specific language used in blogs and forums makes the task being different to that of newswire article summarization. Between the blog and the newswire summarization we could place the e-mail summarization that ranges from summarizing a single e-mail message [20] to the whole thread of related e-mails [23]. Automatic text summarization has also been combined with speech recognition to summarize spoken dialogues [9,18].

Summarization systems have been adapted to a number of different domains. In particular, there has been an extensive research in summarizing medical documents [1]; a) medical journal articles [6,3]; b) healthcare documents for patients [7]. Another domain that attracted the attention is the legal domain. There have been some experiments with the documents from the European Legislation Website[2] [3].

However, text documents differ depending on genre, text type, domain, sublanguage, style, particular topic covered, etc. (for a detailed discussion see [13]). Personal style of a writer, their vocabulary size, word choice, use of expressive means and irony, sentence length and structure preferences are not less affecting. Dialogues and monologues, science fiction and love stories, technical reports and newswire articles, poems and legalese, use of metaphors and synonyms, anaphoric expressions and proper nouns all these carry with them their unique properties. Those properties may affect the quality of summaries generated using the techniques developed for the task of automatic text summarization. And in this paper we would like to study this issue.

We adapt out systems to specific domains, genres, text styles. We develop and implement different summarization techniques and heuristics. But to the best of our knowledge, so far there has been no attempt to treat documents in a collection differently from each other. If a system makes use of pronominal anaphora resolution module, it will try to resolve anaphora in all the documents. Now, what if the document contains only a few pronouns? The performance will slow down but the results will stay the same. What if a document contains a high number of pronouns and the chosen anaphora resolution module cannot handle them correctly? The performance will slow down and the resulting summary will be of a worse quality. If we consider word sense disambiguation task and some specific domains like e.g. legalese documents, the language used is so precise that synonymy disambiguation will probably introduce no improvement into the quality of summaries.

In this paper we address two issues. The first one is concerned with the problem of preliminary document analysis and how the linguistic properties of a text may affect the performance of a number of automatic text summarization

[2] `http://eur-lex.europa.eu/en/legis/index.htm`

techniques. We have focused on the basic linguistic characteristics of text, such as the noun ratio, pronoun ratio and personal noun ratio. And we have analyzed how they affect the summarization systems that use textual entailment and anaphora resolution tools to aid in the summarization process. The final goal would be to develop a system that chooses the best summarization techniques based on the linguistic properties of a document. Moreover, we have divided our corpus in groups according to the topic covered. The second goal of this research is to analyze whether the performance of a text summarization engine depends on the topic of a document.

This paper is structured as follows. Section 2 reports on the related work. Section 3 describes in detail the system used for the experiment. The corpus is described in Section 4. The results are discussed in Section 5. Finally, conclusion and future work are outlined in Section 6.

## 2 Related Work

With the evolution of technology different methods and heuristics have been used to improve extractive summarization systems. The early systems relied on the simple heuristics: i) *sentence location* (sentences located in the beginning or end of the text, headings and the sentences highlighted in bold among others are considered to be more important and are included in the final summary) [5]; ii) *cue phrases* (presence of previously defined words and phrases as "concluding", "argue", "propose" or "this paper") [5,28]; iii) *segment length* (sentences with the length below some predefined threshold can be automatically ignored) [28]; iv) *the most frequent words* (exploring the term distribution of a document allows to identify the most frequent words that are assumed to represent at the same time the most important concepts of the document) [17].

Today we apply various methods to structure information that we extract from documents and to analyze it intelligently. Graph theory has been successfully applied to represent the semantic contents of a document [24] . Latent Semantic Analysis, that involves term by sentences matrix representation and singular value decomposition has also been proven to benefit the task of extractive summarization [26,10]. A number of machine learning algorithms such as decision trees, rule induction, decision forests, Naive Bayes classifiers and neural networks among others have been adapted to this task as well [20,4]. Part-of-speech taggers [20], word sense disambiguation algorithms [24], anaphora resolution [26], textual entailment [27,15] and chunking [20] are among the most frequently used linguistic analysis methods.

To the best of our knowledge there has been no attempt to analyze the impact of shallow linguistic properties of the original text on the quality of automatically generated summaries.

However, there has been a related work involving automatic text summarization and sentence structure. Nenkova et al. [21] focused on how sentence structure can help to predict the linguistic quality of generated summaries. The authors selected a set of structural features that include:

- *sentence length*
- *parse tree depth*
- *number of fragment tags in the sentence parse*
- *phrase type proportion*
- *average phrase length*
- *phrase type rate* was computed for prepositional, noun and verb phrases as diving the number of words of each phrase by the sentence length
- *phrase length* was computed for prepositional, noun and verb phrases as diving the number of phrases of the given type that appeared in the sentence by the sentence length
- *length of NPs/PPs contained in a VP*
- *head noun modifiers*

Though the set of features is different and more diverse, the phrase type ratio and phrase length can be probably compared to the noun, pronoun and proper noun ratios selected for our research. A ranking SVM was trained using these features. The summary ranking accuracy of the ranking SVM was compared to other linguistic quality measures, that include Coh-Metrix, language models, word coherence and entity coherence measures. The evaluation of results was done on the system and input levels. Whereas in the former all participating systems were ranked according to their performance on the entire test set, and in the latter all the summaries produced for a single given input.

Structural features proved to be best suitable for input-level human summaries, middle of the range for input level system summaries and about the worst class of features for system-level evaluation of automatic summaries. At the same time being the most stable set of features and ranging the least across the chosen evaluation settings.

## 3 Summarization System

To analyze the impact of proper noun, pronoun and noun ratios we have chosen the summarization system described in [29]. The system allows a modular combination of anaphora resolution, textual entailment and word sense disambiguation tools with the term or concept frequency based scoring module. In this research we focused on textual entailment and anaphora resolution.

*Textual Entailment.* The task of textual entailment is to capture the semantic inference between text fragments. There has been a number of summarization systems utilizing textual entailment to aid in summarization process. Both in the process of evaluating the final summary and in the process of summary generation. In the latter case textual entailment is often applied to eliminate the semantic redundancy of a document [15].

*Anaphora resolution.* Powerful pronominal anaphora resolution tool relates pronouns to their nominal antecedents. This is of use to all the summarization methods that rely on term overlap, from the simple term frequency to latent semantic analysis. Steinberger et al. [26] report an increase of 1.5% for their summarization system based on latent semantic analysis when anaphoric information is included.

The system that we have chosen for this experiment consists of 4 modules: anaphora resolution, textual entailment, word sense disambiguation and scoring modules (see Figure 1). The scoring module is essential. It is the last step in the process when the final sentence scores are calculated. The remaining 3 modules can be freely combined pairwise with each other and/or with the scoring module. All the 4 modules can be applied at once as well. This suits well our purpose of analyzing the impact of different shallow linguistic properties of a text on various kinds of summarization techniques, that are represented by modules in our case. For this study we do not use all the possible combinations of modules (for the precise list of combinations see Section 5.1)

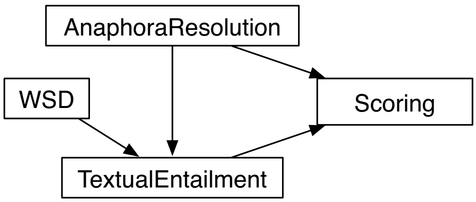


**Fig. 1.** Interaction of semantic components

## 4 Topics and Linguistic Properties of the Data Set

Out data is a set of newswire articles, taken from the Document Understanding Conference challenge of 2002. The original set consists of about 530 articles, that are grouped topicwise into a set of 59 subgroups. The original grouping involves some duplicate articles and there are some topics/events that are represented by more than one such group.

One of the goals of this research is to investigate whether the topic of a document affects the quality of a generated summary. Due to that fact we have selected about 270 articles. The duplicates were removed. The articles were manually reviewed and grouped trying to keep the original DUC grouping whenever possible. Below is the list of the resulted topics. The number of documents in each group is stated in the round brackets.

1. battleship explosion (11)
2. ferry accidents (9)
3. IRA attack (8)
4. earthquake Iran (15)
5. China flood (10)
6. Hurricane Gilbert (13)
7. Mount Pinatubo volcano (5)
8. North American drought (9)
9. thunderstorm US (11)
10. Checkpoint Charlie (5)
11. abortion law (6)
12. Germany reunification (14)
13. Honecker protest (11)
14. Iraq invades Kuwait (27)

15. Robert Maxwell companies (10)
16. striking coal miners (12)
17. US ambassadors (11)
18. Super Bowl (10)
19. marathon (9)
20. Olympics (10)
21. Leonard Bernstein (13)
22. Lucille Ball (14)
23. Margaret Thatcher (10)
24. Sam Walton (7)
25. Gorbachev (10)

We have further grouped the selected articles according to the more general topic covered, e.g. *marathon*, *Olympics*, *Super Bowl* were assigned to the topic on *sports*, etc. This yielded 5 groups, covering the general topics on *accidents*, *natural disasters*, *politics*, *sports* and *famous people* (please see Table 1 for more details).

Having grouped the data in different topics, we proceeded with their linguistic analysis. The selected documents were processed using a part-of-speech tagger to obtain the average *noun* (NR), *pronoun* (PR) and *proper noun ratios* (PNR) for each of the 25 topics. These ratios were calculated by diving the number of words of the respective word class by the total number of words in a document.

**Table 1.** Linguistic properties of the original documents

| | | **PNR** | **NR** | **PR** | **size** |
|---|---|---|---|---|---|
| accidents | 1. battleship explosion | 0.11466 | 0.34381 | 0.03540 | **670.0** |
| | 2. ferry accidents | 0.10874 | 0.34006 | 0.04024 | 423.666 |
| | 3. IRA attack | 0.10666 | 0.31853 | **0.05897** | **599.625** |
| natural disasters | 4. earthquake Iran | **0.15052** | **0.35761** | 0.02933 | 444.8 |
| | 5. China flood | 0.12880 | **0.38535** | 0.02032 | 383.8 |
| | 6. Hurricane Gilbert | **0.13867** | **0.36818** | 0.02501 | **730.923** |
| | 7. Mount Pinatubo volcano | 0.10716 | 0.33642 | 0.03041 | **672.4** |
| | 8. North American drought | 0.10402 | 0.34050 | 0.02485 | 398.0 |
| | 9. thunderstorm US | 0.13165 | **0.36461** | 0.02791 | **718.7** |
| politics | 10. Checkpoint Charlie | **0.16148** | 0.34312 | **0.04406** | 513.2 |
| | 11. abortion law | 0.11954 | 0.33815 | **0.06449** | 545.833 |
| | 12. Germany reunification | **0.14561** | 0.33543 | 0.03163 | 558.5 |
| | 13. Honecker protest | **0.15005** | 0.34585 | 0.03887 | 286.545 |
| | 14. Iraq invades Kuwait | **0.16509** | **0.36821** | 0.03189 | 552.555 |
| | 15. Robert Maxwell companies | **0.17074** | **0.37558** | 0.03782 | 444.1 |
| | 16. striking coal miners | 0.09267 | **0.36252** | 0.02651 | 507.083 |
| | 17. US ambassadors | **0.20187** | **0.38561** | 0.03811 | 415.545 |
| sports | 18. Super Bowl | **0.17758** | **0.39363** | 0.03184 | 438.8 |
| | 19. marathon | 0.13454 | 0.33495 | **0.05224** | **810.555** |
| | 20. Olympics | **0.15359** | **0.35780** | 0.04059 | **607.5** |
| famous people | 21. Leonard Bernstein | **0.19529** | **0.38679** | **0.04427** | **596.923** |
| | 22. Lucille Ball | 0.13095 | 0.32332 | **0.08895** | **848.714** |
| | 23. Margaret Thatcher | 0.13329 | 0.31561 | **0.06571** | **624.4** |
| | 24. Sam Walton | 0.10693 | 0.32362 | **0.05967** | **566.714** |
| | 25. Gorbachev | 0.09943 | 0.31251 | **0.05673** | **745.9** |
| average | | 0.13718 | 0.35031 | 0.04183 | 564.191 |

This shallow linguistic analysis methods were chosen in agreement with the summarization system described in Section 3. The noun ratio has been chosen since a topic of a document is usually characterized in the form of noun phrases and textual entailment (with or without the word sense disambiguation) can be used to eliminate the semantic redundancy. The anaphora resolution process involves analyzing the pairs of nouns, pronouns and proper nouns in a document.

Table 1 contains the results for the selected features topic-wise. The figures higher than the average are highlighted in bold. Already on this shallow analysis level it can be seen, that different topics have different tendencies. The documents that cover political issues and sports tend to have a higher number of proper nouns. The articles about famous people contain a lot of pronouns. The latter led us to the hypothesis, that summarization systems that involve anaphora resolution would yield summaries of a better quality for those articles. While the former suggested to rather apply a textual entailment heuristics. The actual results obtained when applying the selected summarization system to the set of 25 groups of documents are discussed in Section 5.

## 5 Results and Discussion

### 5.1 Experiment Setup

The chosen summarization system as explained in Section 3 (see Figure 1) allows to freely combine anaphora resolution, textual entailment, word sense disambiguation and scoring modules. This suits well the purpose of this research since we can analyze how different document groups behave in different systems settings. This also allows us to see whether a single module yields better summaries than the combination of all the 4 modules for the selected 25 groups of documents. We have selected the following combinations of modules (please recall that the scoring module is the essential final step and thus common to all of them, so it is omitted from the description of combinations):

– **ASW** basic stopwords filtering
– **AR** pronominal anaphora are substituted by their antecedents prior to scoring
– **TE** redundant sentences are eliminated using textual entailment
– **TEWSD** members of the same WordNet[3] synset are replaced by the same synset representative and after that the redundant sentences are identified using the textual entailment module
– **ARTEWSD** pronominal anaphora are substituted by their antecedents, then the words of the resulting text are replaced by the chosen representative of the WordNet synset that they belong to and finally the redundant sentences are filtered out by the textual entailment module

Using these combination we generated summaries for all the 25 groups. Thereafter these summaries were evaluated using the ROUGE toolkit [14]. The average results group-wise are presented in Table 2.

[3] `http://wordnet.princeton.edu/`

**Table 2.** ROUGE-1 recall topicwise for generated summaries

| | | ASW | AR | TE | TEWSD | ARTE WSD |
|---|---|---|---|---|---|---|
| accidents | 1. battleship explosion | 0.41640 | 0.43224 | 0.41762 | 0.40645 | 0.42407 |
| | 2. ferry accidents | 0.40627 | 0.42393 | 0.38874 | 0.38874 | 0.41932 |
| | 3. IRA attack | 0.34765 | 0.38686 | 0.34955 | 0.34714 | 0.39822 |
| | average | 0.39010 | 0.41434 | 0.38530 | 0.38077 | 0.41387 |
| natural disasters | 4. earthquake Iran | 0.39851 | 0.40446 | 0.42258 | 0.42936 | 0.42390 |
| | 5. China flood | 0.44805 | 0.46006 | 0.48130 | 0.47697 | 0.46641 |
| | 6. Hurricane Gilbert | 0.40462 | 0.41458 | 0.40446 | 0.40584 | 0.44845 |
| | 7. Mount Pinatubo volcano | 0.32098 | 0.38782 | 0.31535 | 0.31535 | 0.36166 |
| | 8. North American drought | 0.42936 | 0.44494 | 0.40973 | 0.41773 | 0.41594 |
| | 9. thunderstorm US | 0.32589 | 0.37270 | 0.35346 | 0.36224 | 0.40338 |
| | average | 0.38790 | 0.41409 | 0.39781 | 0.40124 | 0.41995 |
| politics | 10. Checkpoint Charlie | 0.50284 | 0.42880 | 0.46839 | 0.47730 | 0.50036 |
| | 11. abortion law | 0.30283 | 0.38178 | 0.31252 | 0.31252 | 0.37216 |
| | 12. Germany reunification | 0.40949 | 0.42933 | 0.40967 | 0.40868 | 0.45277 |
| | 13. Honecker protest | 0.48695 | 0.48410 | 0.50062 | 0.49798 | 0.50517 |
| | 14. Iraq invades Kuwait | 0.38630 | 0.39488 | 0.40467 | 0.40404 | 0.41887 |
| | 15. Robert Maxwell companies | 0.42572 | 0.44064 | 0.42153 | 0.42411 | 0.44893 |
| | 16. striking coal miners | 0.46955 | 0.44690 | 0.49049 | 0.49049 | 0.49765 |
| | 17. US ambassadors | 0.45776 | 0.39415 | 0.47221 | 0.47242 | 0.45394 |
| | average | 0.43018 | 0.42507 | 0.43501 | 0.43594 | 0.45623 |
| sports | 18. Super Bowl | 0.50047 | 0.45308 | 0.53126 | 0.53318 | 0.51541 |
| | 19. marathon | 0.36541 | 0.35290 | 0.34855 | 0.34932 | 0.37978 |
| | 20. Olympics | 0.42524 | 0.42857 | 0.47455 | 0.48249 | 0.48283 |
| | average | 0.43037 | 0.41151 | 0.45145 | 0.45499 | 0.45934 |
| famous people | 21. Leonard Bernstein | 0.42669 | 0.42064 | 0.43319 | 0.43319 | 0.42277 |
| | 22. Lucille Ball | 0.37099 | 0.36333 | 0.38172 | 0.37673 | 0.38633 |
| | 23. Margaret Thatcher | 0.41783 | 0.42240 | 0.41403 | 0.43089 | 0.40990 |
| | 24. Sam Walton | 0.36775 | 0.37903 | 0.35351 | 0.35351 | 0.40549 |
| | 25. Gorbachev | 0.36199 | 0.37621 | 0.35525 | 0.35883 | 0.40405 |
| | average | 0.38905 | 0.39232 | 0.38754 | 0.39063 | 0.405708 |

## 5.2 Discussion

We pursued two different goals in this research. The initial hypothesis was that the topic of the original document and the quality of generated summary are related. The second one was that the quality of a generated summary rather depends on linguistic properties of the original text and how they interact with the particular summarization technique chosen to tackle this task.

Below is the analysis of the results with respect to both of the goals.

**Does the Topic of the Original Document affect the Quality of Generated Summaries?** On hand of the obtained results we couldn't prove the first hypothesis. If we for example consider the *sports* topic, it becomes evident that:

- already the starting values for the ASW setting range from 0.36541 to 0.50047

- the degree of improvement when adding additional modules differs between the three topics. For the *marathon* topic we started with the value of 0.36541 for the ASW setting and reached the maximum of 0.37978 for the ARTEWSD setting. Meanwhile for the *Olympics* the initial ROUGE value was 0.42524 and jumped up to 0.48283
- different modules and their combinations affected the quality of generated summaries in different ways. AR decreased the quality of summaries for the documents on *Super Bowl*, while TE and TEWSD setting noticeably improved it. The combination of both in the ARTEWSD setting yielded worse results than the mere TE and TEWSD settings. On the other hand, while the summaries of the documents on *Olympics* reveal the same tendencies for AR and TE/TEWSD settings, the combination of modules in ARTEWSD setting yielded the best results

No clear relation between the topic of a document and the summarization method performing best for it could be established. The same tendencies were observed for the remaining 4 general topics. Thus we can conclude that on this coarse-grained level of topic differentiation the performance of a summarization system does not depend on the topic of a document.

**Do the Linguistic Properties of the Original Document affect the Quality of Generated Summaries?** If we consider again the topic *Super Bowl* and the *Mount Pinatubo volcano*, we can see that the best results were obtained not for the combination of all the modules, but for the TE/TEWSD in the former case and the AR in the latter. The motivation for second objective of this research was to identify certain properties in the original text prior to summarizing to further chose the best summarization technique that will allow us to maximilly improve the quality of generated summaries. Though there are some exception, we could identify the following trends.

*Textual Entailment.* Whenever the noun ratio is above 0.35, textual entailment benefits the process of summarization (see for example the topics 4, 5, and 9 in Table 2). In the opposite case it worsens the results (as in topics 19, 24 or 25)

*Anaphora Resolution.* The chosen anaphora resolution tool seems to interact with the pronoun and proper noun ratios. It benefits when a) both values are low (e.g. topics 1, 2, 5, 7); b) proper noun ratio is low and pronoun ratio is high (as in topics 3, 11, 24); c) proper noun ratio is high and pronoun ratio is low (e.g. 4, 6, 12). The final thresholds for *high* and *low* ratios must be determined statistically with the larger corpus. In our case by *high* we mean "above the average" (thus highlighted in bold in Table 1) and *low* otherwise. When both ratios are high (topics 10 and 21), the anaphora resolution module tend to reduce the quality of generated summaries.

*Textual Entailment and Anaphora Resolution.* The interaction of textual entailment with anaphora resolution is less straightforward. In some cases when one should benefit and the other worsen the results, their combination still tops the results from the best performing module in isolation. For example, for the topic 25 AR improves the results of ASW. Both TE and TEWSD make them worse.

But the combination of AR and TEWSD improves over the results yielded on AR. The opposite case is observed for the topic 1, when AR benefits, TE yields worse results, and their combination is still worse than the results of AR only. We thus assume that there are more linguistic properties and ways of interaction with the summarization techniques that are the subject to further research.

Nevertheless, it becomes clear that the linguistic properties of the original document affect the quality of generated summaries.

## 6 Conclusion and Future Work

We have pursued two objectives in this research: i) determine whether the performance of a summarization system depends on the topic of a document; ii) determine whether the quality of a generated summary depends on the linguistic properties of the original text and how they interact with different summarization techniques. Given the findings discussed in Section 5.2, we conclude that i) no clear relation between the topic of a document and the can be established; ii) a preliminary document analysis stage could benefit the summarization process. The latter is valid both for modular summarization systems as described in Section 3 and any other summarization system based on a single method (i.e. latent semantic analysis or graph-based approaches), provided that their developers are aware of the inherent properties of the text their system can handle the best. Therefore it becomes advisable to adapt summarization systems to the linguistic properties of input documents.

To the best of our knowledge there has been no other study of the impact of linguistic properties of the original text on the quality of generated summary. Thus this research focused on a few linguistic properties, such as noun, pronoun and proper noun ratio and their interaction with the summarization heuristics involving textual entailment and anaphora resolution. For the future work we are planning to include other structural and semantic text properties. It is worth investigating how the word ambiguity index of the original document affects summarization systems that use word sense disambiguation and textual entailment. Another direction would lead to the structural features that include phrase type ratio, instead of mere noun or pronoun ratio, and parse tree analysis, that can be combined with text simplification for automatic summarization.

**Acknowledgments.** This research work has been partially funded by the European Commission under the Seventh (FP7 - 2007-2013) Framework Programme for Research and Technological Development through the FIRST project (FP7-287607); the Spanish Government through the project TEXT-MESS 2.0 (TIN2009-13391-C04), "Análisis de Tendencias Mediante Técnicas de Opinión Semántica" (TIN2012-38536-C03-03 ) and "Técnicas de Deconstrucción en la Tecnologías del Lenguaje Humano" (TIN2012-31224); and by the Valencian Government through the project PROMETEO (PROMETEO/2009/199). This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## References


1. Afantenos, S., Karkaletsis, V., Stamatopoulos, P.: Summarization from medical documents: a survey. Artificial Intelligence in Medicine 33, 157–177 (2005)
2. Amini, M.-R., Gallinari, P.: The Use of Unlabeled Data to Improve Supervised Learning for Text Summarization. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2002, p. 105. ACM Press, New York (2002)
3. Ceylan, H., Mihalcea, R., Öyertem, U., Lloret, E., Palomar, M.: Quantifying the Limits and Success of Extractive Summarization Systems Across Domains. In: Human Language Technologies, pp. 903–911. Association for Computational Linguistics, Stroudsburg (2010)
4. Chuang, W.T., Yang, J.: Text Summarization by Sentence Segment Extraction Using Machine Learning Algorithms. In: Terano, T., Liu, H., Chen, A.L.P. (eds.) PAKDD 2000. LNCS, vol. 1805, pp. 454–457. Springer, Heidelberg (2000)
5. Edmunson, H.: New methods in automatic extracting. Journal of the ACM 16(2), 264–285 (1969)
6. Elhadad, N., McKeown, K., Kaufman, D., Jordan, D.: Facilitating physicians access to information via tailored text summarization. In: AMIA Annual Symposium, pp. 226–230 (2005)
7. Elhadad, N., Kan, M.-Y., Klavans, J.L., McKeown, K.R.: Customization in a Unified Framework for Summarizing Medical Literature. In: Artificial Intelligence in Medicine, vol. 33, pp. 179–198 (2005)
8. Filippova, K., Mieskes, M., Nastase, V.: Cascaded Filtering for Topic-Driven Multi-Document Summarization. In: Proceedings of the Document Understanding Conference, Rochester, N.Y., pp. 30–35 (2007)
9. Galley, M.: Automatic Summarization of Conversational Multi-Party Speech. In: The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, pp. 1914–1915. AAAI Press, Boston (2006)
10. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), pp. 19–25. ACM Press, New York (2001)
11. Hu, M., Sun, A., Lim, E.: Comments-Oriented Blog Summarization by Sentence. In: Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management, pp. 901–904. Association for Computational Linguistics, New York (2007)
12. Kazantseva, A.: Automatic Summarization of Short Fiction, Master thesis (2006), http://www.site.uottawa.ca/ ankazant/pubs/thesis_corrected_18_12_06_OK.pdf
13. Lee, D.: Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle. Language and Computers 5, 37–72 (2002)
14. Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries. In: Proceedings of the Workshop on Text Summarization, p. 89 (2004)
15. Lloret, E., Ferrández, O., Muñoz, R., Palomar, M.: A Text Summarization Approach Under the Influence of Textual Entailment. In: 5th International Workshop on NLPCS, pp. 22–31 (2008)

16. Lloret, L., Palomar, M.: A Gradual Combination of Features for Building Automatic Summarisation Systems. In: Proceedings of the 12th International Conference on Text, Speech and Dialogue (TSD), Pilsen, Czech Republic, pp. 16–23 (2009)
17. Luhn, H.P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development 2(2), 157–165 (1958)
18. McKeown, K., Hirschberg, J., Galley, M., Maskey, S.: From Text to Speech Summarization. In: International Conference on Acoustics, Speech, and Signal Processing, pp. 997–1000. IEEE, Philadelphia (2005)
19. Mihalcea, R., Ceylan, H.: Explorations in Automatic Book Summarization. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 380–389 (2007)
20. Muresan, S., Tzoukermann, E., Klavans, J.L.: Combining Linguistic and Machine Learning Techniques for Email Summarization. In: Proceedings of the 2001 Workshop on Computational Natural Language Learning (ConLL 2001). Association for Computational Linguistics, Stroudsburg (2001)
21. Nenkova, A., Chae, J., Louis, A., Pitler, E.: Empirical Methods in Natural Language Generation. Springer, Heidelberg (2010)
22. Nenkova, A.: Automatic Summarization. Foundations and Trends in Information Retrieval 5, 103–233 (2011)
23. Nenkova, A., Bagga, A.: Facilitating Email Thread Access by Extractive Summary Generation. In: Nicolov, N., Bontcheva, K., Angelova, G., Mitkov, R. (eds.) Recent Advances in Natural Language Processing III, Selected Papers from RANLP 2003, pp. 287–296. John Benjamins, Amsterdam (2003)
24. Plaza, L., Díaz, A.: Using Semantic Graphs and Word Sense Disambiguation. Techniques to Improve Text Summarization. Procesamiento del Lenguaje Natural 47, 97–105 (2011)
25. Saggion, H.: Topic-based Summarization at DUC 2005. In: Proceedings of the Document Understanding Workshop, Vancouver, B.C., Canada, pp. 1–6 (2005)
26. Steinberger, J., Poesio, M., Kabadjov, M.A., Ježek, K.: Two Uses of Anaphora Resolution in Summarization. Information Processing and Management 43(6), 1663–1680 (2007)
27. Tatar, D., Tamaianu-Morita, E., Mihis, A., Lupsa, D.: Summarization by Logic Segmentation and Text Entailment. In: 33rd CICLing, pp. 15–26 (2008)
28. Teufel, S., Moens, M.: Sentence extraction as a classification task. In: ACL/EACL 1997 Workshop on Intelligent Scalable Text Summarization, pp. 58–65. Association for Computational Linguistics, Madrid (1997)
29. 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: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) TSD 2012. LNCS, vol. 7499, pp. 306–313. Springer, Heidelberg (2012)
30. Yang, J., Cohen, A.M., Hersh, W.: Automatic summarization of mouse gene information by clustering and sentence extraction from MEDLINE abstracts. In: AMIA Annual Symposium, pp. 831–835 (2007)