



Editorial

COMPENDIUM: A text summarization system for generating abstracts of research papers

Elena Lloret ^{a,*}, María Teresa Romá-Ferri ^b, Manuel Palomar ^a^a Department of Software and Computing Systems, University of Alicante, Spain^b Department of Nursing, University of Alicante, Spain

ARTICLE INFO

Available online 14 August 2013

Keywords:

Human Language Technologies

NLP applications

Text summarization

Information systems

ABSTRACT

This article analyzes the appropriateness of a text summarization system, COMPENDIUM, for generating abstracts of biomedical papers. Two approaches are suggested: an extractive (COMPENDIUM_E), which only selects and extracts the most relevant sentences of the documents, and an abstractive-oriented one (COMPENDIUM_{E-A}), thus facing also the challenge of abstractive summarization. This novel strategy combines extractive information, with some pieces of information of the article that have been previously compressed or fused. Specifically, in this article, we want to study: i) whether COMPENDIUM produces good summaries in the biomedical domain; ii) which summarization approach is more suitable; and iii) the opinion of real users towards automatic summaries. Therefore, two types of evaluation were performed: quantitative and qualitative, for evaluating both the information contained in the summaries, as well as the user satisfaction. Results show that extractive and abstractive-oriented summaries perform similarly as far as the information they contain, so both approaches are able to keep the relevant information of the source documents, but the latter is more appropriate from a human perspective, when a user satisfaction assessment is carried out. This also confirms the suitability of our suggested approach for generating summaries following an abstractive-oriented paradigm.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The vast amount of information currently available has fuelled research into systems and tools capable of managing such information in an effective and efficient manner. That is the case of Text Summarization (TS), whose aim is to produce a condensed new text containing a significant portion of the information in the original text(s) [40]. In particular, TS has been shown to be very useful as a stand-alone application [39,9], as well as in combination with other systems, such as information retrieval [7], information extraction [8], text classification [38], or automatic rating prediction [36].

TS has a wide range of applications. For instance, in the scientific context, TS can be used for generating abstracts of research papers automatically, thus avoiding authors the difficult task of having to synthesize in a small paragraph the main topics addressed in their research. These abstracts are very important not only to provide readers an overview of the article, but also to be used by automatic systems for indexing, searching and retrieving information without having to process the whole document. Moreover, another application of TS would be the automatic generation of newsletters, containing information which is of interest for a particular group of experts. In the biomedical domain, this would allow health workers and professionals to be updated of recent findings, such as new drugs, novel treatments, etc.

However, the process of TS, and in particular, the automatic generation of abstracts, is still very challenging. The relevant information has to be identified first, and then it has to be generalized, with the purpose of stating the gist of the whole document in

* Corresponding author at: Apdo. de correos, 99, E-03080 Alicante, Spain. Tel.: +34 965903772; fax: +34 965909326.
E-mail address: elloret@dlsi.ua.es (E. Lloret).

one paragraph. This difficulty is shown by the fact that, although there have been some attempts to generate abstractive summaries in recent years [32,34], most of the current work on TS still focuses on extractive summarization [19,22,42]. The differences between them are that extractive summarization consists of selecting the most important sentences of a document to form the final summary and they are extracted as they are in the source documents. On the contrary, abstractive summarization also implies the generation of new information, in the form of sentence compression or fusion, or even natural language generation. The main problem associated to extractive summarization is the lack of coherence resulting summaries exhibit, partly due to non-resolved coreference relationships, and the wrong link between sentences.

Therefore, in this article, our goal is to analyze to what extent TS is useful for a specific application: the automatic generation of research paper abstracts in the biomedical domain. In particular, we develop a TS system: COMPENDIUM, which is capable of producing two types of generic summaries: extracts (COMPENDIUM_E), which follows a pure extractive approach; and abstractive-oriented summaries (COMPENDIUM_{E-A}), which combines extractive and abstractive techniques (Fig. 1). In this manner, we study to what extent such approaches are appropriate to produce summaries that serve as a surrogate of the source document or, in contrast, if they can only be used in order to provide users with an idea of what is important in the document. Therefore, the specific objectives within the scope of this article we want to analyze are: i) if COMPENDIUM produces good summaries in this domain; ii) which is the best strategy to adopt for generating this type of summaries (extractive or abstractive-oriented); and iii) the opinion of real users towards automatic summaries.

The structure of the article is as follows: Section 2 introduces previous work on state-of-the-art TS systems, as well as TS approaches that were developed for addressing specific applications. In Section 3, our proposed TS system, COMPENDIUM, is explained in detail, comprising the two approaches for generating extractive and abstractive-oriented summaries, respectively. Section 4 describes the experimental framework. Within this section, we provide the description of the data set employed, together with the analysis of the human abstracts of the research papers on the one hand, and we outline the experiments performed, on the other hand. Then, the evaluation conducted, the results obtained together with a discussion for each type of evaluation are provided in Section 5, and finally, the conclusions of the paper together with the future work are outlined in Section 6.

2. Related work

In this section, we explain existing TS systems that produce summaries automatically, as well as previous work on different approaches that have used TS techniques to address specific tasks (e.g. for generating Wikipedia articles [37], weather forecast reports [3], etc.).

Regarding the TS systems, we can find a wide range of them for generating different types of summaries. One of the most cited summarizers is MEAD [31]. This system produces extractive single- and multi-document generic or query-focused summaries in English and Chinese, thus being also a multilingual summarizer. For determining the most important sentences, it relies on the calculation of

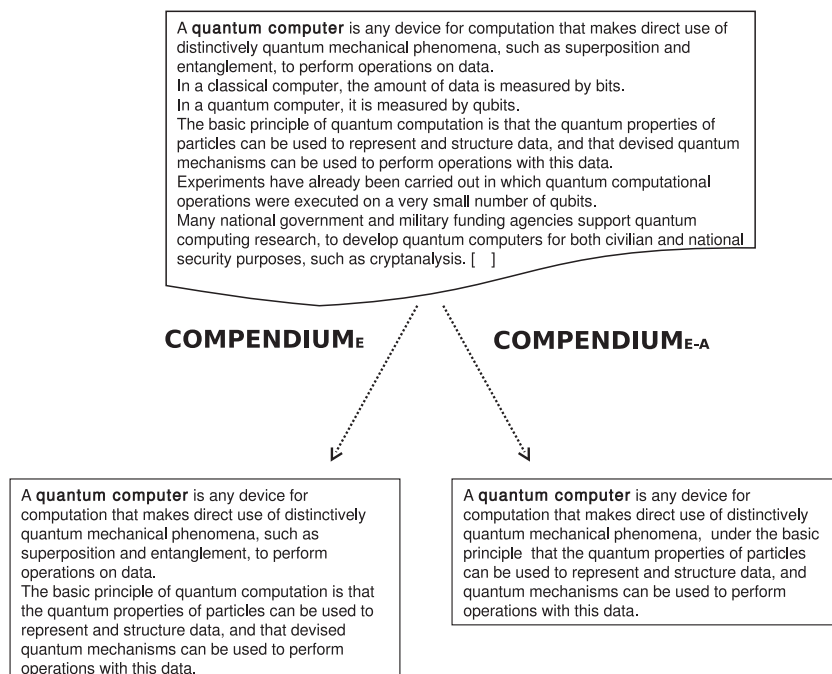


Fig. 1. Illustration of the functionalities of COMPENDIUM system.

surface features, such as sentence position, sentence length, similarity with the first sentence, and similarity with a centroid, which are then combined linearly. SUMMA [33] follows a similar approach, also relying on the combination of statistical, positional and similarity features. In particular, the main features are: i) sentence similarity to the centroid; ii) sentence position and iii) similarity with the lead part of the document where the sentence comes from. It uses a vector space representation, where each vector position contains the term and its tf-idf, and it is capable of producing single- and multi-document extractive summaries, as well as generic or query-focused summaries. QCS system [10] integrates a summarization approach within a broad process for retrieving and clustering information. For generating summaries from a set of documents, it first generates single-document summaries, and then a second summarization process is applied to these summaries in order to obtain the final one. For detecting relevant sentences, Hidden Markov Models are employed after some sentences have been trimmed. The sentences with highest probability in the model are chosen for the summary. The AZOM text summarization system for Persian [45] combines statistical and conceptual properties of unstructured documents and generates a summary with a specific structure. Other systems, such as SummGraph [27], employ graph-based algorithms and knowledge databases for detecting relevant content in the documents. In particular, this system has been proven to work successfully in the newswire, biomedical, and tourist domain, and the techniques used have been also successfully employed for retrieving information from medical records [28].

Furthermore, we can also find systems that are especially targeted to produce a specific type of summaries (e.g., sentiment-based summaries) or deal with multi-linguality. That is the case of CBSEAS [4] which generate multi-document extractive sentiment-based summaries, or MUSE – Multilingual Sentence Extractor [21], that employs language-independent techniques for generating summaries in English and Hebrew.

In other contexts, TS techniques and approaches have been used for solving specific tasks. For instance, Balahur and Montoyo [2] use opinion mining techniques for extracting opinion features from customer reviews and then summarizing them. Sauper and Barzilay [37] propose an automatic method to generate Wikipedia articles, where specific topic templates, as well as the information to select are learnt using machine learning algorithms. The templates are obtained by means of recurrent patterns for each type of document and domain. For extracting the relevant content, candidate fragments are ranked according to how representative they are with respect to each topic of the template. Other approaches that also rely on the use of templates to organize and structure the information previously identified, are based on information extraction systems. In Kumar et al. [18], reports of events are generated from the information of different domains (biomedical, sports, etc.) that is stored in databases. In such research, human-written abstracts are used, on the one hand, to determine the information to include in a summary, and on the other hand, to generate templates. Then, the patterns to fill these templates in are identified in the source texts. Similarly, in Carenini and Cheung [6], patterns in the text are also identified, but since their aim is to generate contrastive summaries, discourse markers indicating contrast such as “although”, “however”, etc. are also added to make the summary sound more naturally. To generate summaries from the complete biography of a person is also interesting. This can also be considered a particular type of multi-document summarization, since its goal is to produce a piece of text containing the most relevant aspects of a specific person. Instead of using templates for detecting which information should be of interest from a person, Zhou et al. [46] analyzed several machine learning algorithms (*Naïve Bayes*, *Support Vector Machines*, and *Decision Trees*) to classify sentences, distinguishing between those ones containing biographic information (e.g. the date/place of birth) from others that do not.

Natural Language Generation (NLG) has been also applied for adding new vocabulary and language structures in summaries. In Yu et al. [43] very short summaries are produced from large collections of numerical data. The data is presented in the form of tables, and new text is generated for describing the facts that such data represent. Belz [3] also suggests a TS approach based on NLG, in order to generate weather forecast reports automatically.

Another interesting approach is to use citations from articles. In Kan et al., [17] it was shown that from bibliographic entries it was possible to produce an indicative summary. The main idea behind this assumption is that such entries contain informative as well as indicative information, for example, details about the resource or metadata, such as author or purpose of the paper. In their research, a big annotated corpus (2000 annotated entries) is developed for such purposes. Following the idea of generating summaries from this input information, in Qazvinian and Radev [30] citations are analyzed to produce a single-document summary from scientific articles. The final objective is to generate summaries about a specific topic.

The generation of technical surveys has also been addressed in the recent years. In Mohammad et al. [24] citations are used to automatically generate technical surveys. They experimented with three types of inputs (full papers, abstracts and citation texts), and analyzed different already existing summarization systems to create such surveys, such as LexRank [11], Trimmer [44], or C-RR and C-LexRank [30]. Among the conclusions drawn from the experiments, it was shown that multi-document technical survey creation benefits considerably from citation texts.

Focused on the biomedical domain, a wide range of approaches that deal with this domain have emerged in recent years [16,14,1]. Specifically, we can also find several research works that address the generation of abstracts. For instance, Pollock and Zamora [29] used cue words for generating abstracts. This technique consists of determining the relevance of a sentence by means of the phrases or words it contains that could introduce relevant information, such as “in conclusion” or “the aim of this paper”. Furthermore, Saggion and Lapalme [35] exploited also this kind of information, by means of a set of patterns that were later combined with the information extracted from the document.

Our contribution with respect to state-of-the-art systems is that COMPENDIUM is able to deal with extractive summarization, but in addition, a novel strategy for producing abstractive-oriented summaries is proposed. Contrary to existing research works for the biomedical domain, we do not rely on specific patterns, nor we learn the structure or the content of the document for generating summaries. Moreover, we carry out an analysis with the purpose of determining which approach performs better, not only according to the information contained in summaries, but also, to user’s assessment.

3. Text summarization approach: COMPENDIUM

In this Section, the two suggested approaches that COMPENDIUM TS system is able to follow for generating summaries are explained. First we described COMPENDIUM_E, a pure extractive TS approach (Section 3.1), and then we take this extractive approach as a basis in an attempt to improve the final summaries by integrating abstractive techniques, leading to COMPENDIUM_{E-A} (Section 3.2).

3.1. COMPENDIUM_E: extractive summarization

For generating extractive summaries, COMPENDIUM relies on four main stages, as it can be seen in Fig. 2:

- *Surface Linguistic Analysis*: first of all, a basic linguistic analysis is carried out in order to prepare the text for further processing. Such analysis comprises tokenization, sentence segmentation, part-of-speech tagging, and stop word identification, since stop words will not be taken into consideration for the remaining stages.
- *Redundancy detection*: a Textual Entailment (TE) tool [12] is used to detect and remove repeated information. The main idea behind the use of TE for detecting redundancy is that those sentences whose meaning is already contained in other sentences can be discarded, as the information has been already mentioned. Therefore, by applying TE we can obtain a set of sentences from the text which do not hold an entailment relation with any other, and then keep this set of sentences for further processing. Although in a different manner, TE has been successfully applied to summarization in previous research [41].
- *Topic Identification*: In this stage, the most relevant topic/topics of the document are identified by means of their frequency of appearance in the text. Term Frequency (TF) calculation is employed for achieving this goal, because it has been shown in previous work that high frequent words are indicative of the topic of a document [23,25], and, what is more, they are very likely to appear in human-written summaries [26]. Therefore, the frequency of a word, without considering stop words, is computed and it is used in the next stage of the TS process for determining the overall relevance of a sentence.
- *Relevance detection*: this stage computes a score for each sentence depending on its importance, relying on the topic identification stage previously explained and a cognitive-based feature: the Code Quantity Principle (CQP) [15]. The CQP states that the most important information within a text is expressed by a high number of units (for instance, words or noun phrases). In our TS approach, we select as units noun phrases because they are flexible coding units and can vary in the number of elements they contain depending on the information detail one wants to provide. Therefore, in order to generate the summary, sentences containing longer noun phrases of high frequent terms (i.e., the most representative topics of a text) are considered to be more important, thus having more chances to appear in the final summary.
- *Summary generation*: finally, having computed the score for each sentence, in this last stage of the TS process sentences are ranked according to their relevance and the highest ones are selected and extracted in the same order as they appear in the original document, thus generating an extractive summary, and leading to COMPENDIUM_E variant.

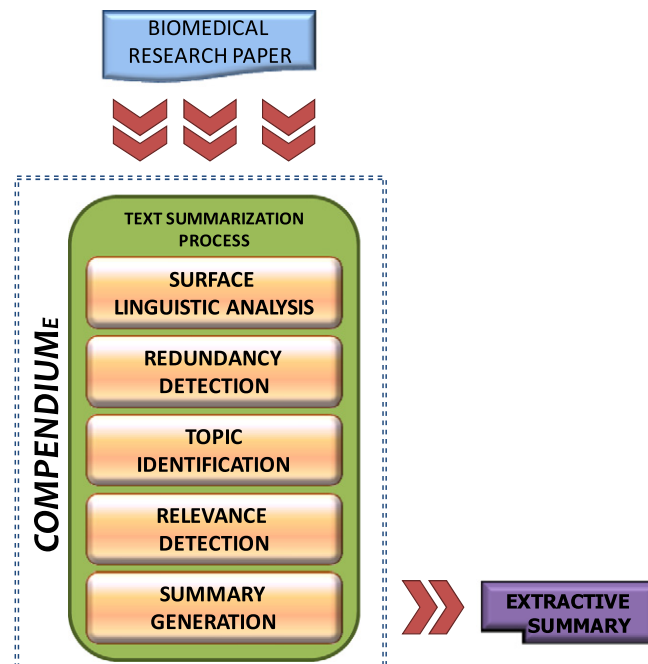


Fig. 2. Overview of the stages involved in COMPENDIUM_E.

3.2. COMPENDIUM_{E-A}: abstractive-oriented summarization

The second type of summaries COMPENDIUM is able to generate is abstractive-oriented summaries, by means of COMPENDIUM_{E-A}. In this approach, extractive and abstractive techniques are combined in the following manner: we take as a basis the COMPENDIUM_E approach described in the previous subsection (Section 3.1), and we integrate an *information compression and fusion stage* after the relevant sentences have been identified and before the final summary is generated, thus generating abstractive-oriented summaries. The goal of this stage is to generate new sentences in one of these forms: either a compressed version of a longer sentence, or a new sentence containing information from two individual ones. The main steps involved in this stage are (Fig. 3):

- **Word graph generation:** for generating new sentences, we rely on word graphs adopting a similar approach to the one described in Filippova [13]. Specifically in our approach, we first generate an extractive summary in order to determine the most relevant content for being included in the summary. Then, a weighted directed word graph is built taking as input the generated extract, where the words represent the nodes of the graph, and the edges are adjacency relationships between two words. The weight of each edge is calculated based on the inverse frequency of co-occurrence of two words and taking also into account the importance of the nodes they link, through the PageRank algorithm [5]. Once the extract is represented as a word graph, a pool of new sentences is created by identifying the shortest path between nodes (e.g. using Dijkstra's algorithm), starting with the first word of each sentence in the extract, in order to cover its whole content. The reason why we used the shortest path is twofold. On the one hand, it allows sentences to be compressed, and on the other hand, we can include more content in the summary, in the case several sentences are fused.
- **Incorrect paths filtering:** this stage is needed since not all of the sentences obtained by the shortest paths are valid. For instance, some of them may suffer from incompleteness (*"Therefore the immune system."*). Consequently, in order to reduce the number of incorrect generated sentences, we define a set of rules, so that sentences not accomplishing all the rules are not taken into account. Three general rules are defined after analyzing manually the resulting sentences derived from the documents in the data set, which are: i) the minimal length for a sentence must be 3 words, since we assume that three words (i.e., subject + verb + object) is the minimum length for a complete sentence; ii) every sentence must contain a verb; and iii) the sentence should not end in an article (e.g. a, the), a preposition (e.g. of), an interrogative word (e.g. who), nor a conjunction (e.g. and).
- **Given and new information combination:** the objective of the last step is to decide which of the new sentences are more appropriate to be included in the final summary. Since we want to develop a mixed approach, combining extractive and abstractive techniques, the sentences of the final summary will be selected following a strategy that maximizes the similarity between each of the new sentences and the ones that are already in the extract, given that the similarity between them is above a predefined threshold. For this, we use the cosine measure to compute the similarity between two sentences, and a threshold has been empirically set to 0.5. Finally, in the cases where a sentence in the extract has an equivalent in the set of new generated sentences, the former will be substituted for the latter; otherwise, we take the sentence in the extract.

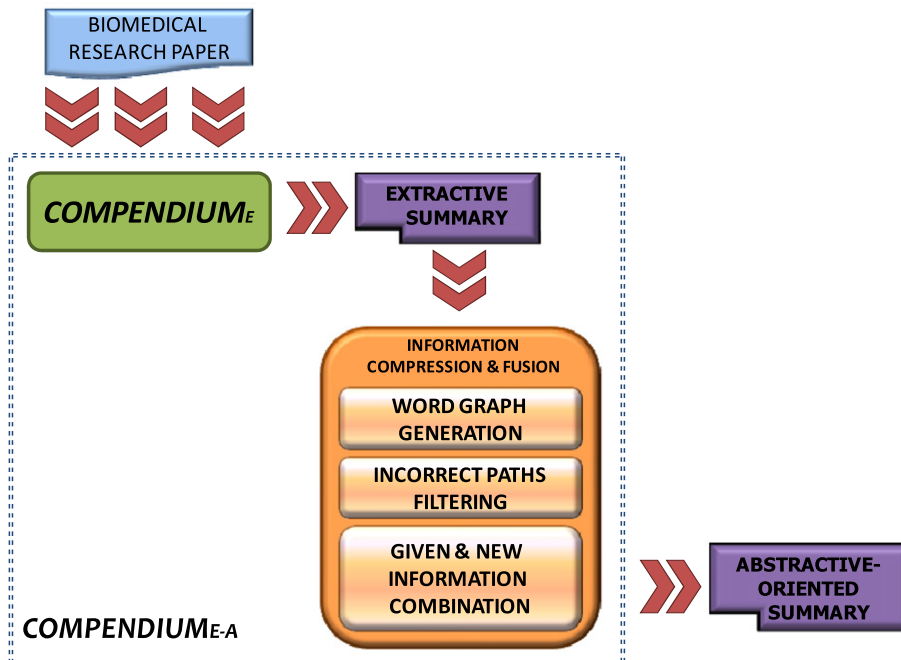


Fig. 3. Overview of the stages involved in COMPENDIUM_{E-A}.

Table 1

Fragment of the text of an article of Autoimmunity Reviews journal.

Fragment of a biomedical article:
<p>1: Histologic examination of lesions plays a key role in the diagnostics of cutaneous lupus erythematosus (LE). 2: LE has a broad spectrum of histological signs which are related to the stages of the lesions, but some signs apply to all stages (e.g., mucin deposition). [...] 8: From the histological standpoint, LE can be classified only into early, fully developed, late LE, and special manifestations of LE. [...] 26: If interface dermatitis becomes manifest at its maximal expression, Rowell's syndrome must be considered. [...] 33: In subtypes of LE with epidermal involvement, damage to keratinocytes is a distinctive sign which is very helpful in establishing the diagnosis. [...] 50: An important histological differential diagnosis is the polymorphous light eruption (PLE), which exhibits prominent edema of the papillary dermis with occasional vesiculation and absence of mucin deposition. [...] 78: LE has a broad spectrum of histopathological signs, which are related to the stages of the lesions. [...] 81: LE may be associated with Kikuchi–Fujimoto disease (histiocytic necrotizing lymphadenitis). [...]</p>

4. Experimental framework

In this section, we provide a description of the corpus employed (Subsection 4.1), we analyze the abstracts included in the research papers (Subsection 4.2) and we finally explain the experiments performed (Subsection 4.3).

4.1. Description of the data set: research articles

For our experiments, a set of 50 research articles from a specialized journal of medicine was collected directly from the Web. Specifically, these articles belonged to the Autoimmunity Reviews journal.¹

The structure is similar for all of the articles in the corpus: title, authors and affiliations, keywords, abstract, outline and a variable number of sections (the content of these section is what we considered as the text of the article), depending on each article. At the end of each one, there are also two special sections: one is the “take-home messages” and the last one contains the references. In some articles, an acknowledgement section is included as well. In addition, each article may contain figures and tables, which are not taken into consideration for generating the summaries.

A fragment of an article together with its complete abstract is shown in Tables 1 and 2, respectively. For clarity reasons we have numbered the sentences according to its position within the text of the article and the abstract. The abstracts included in the articles were generated by the authors, and therefore, we use them as gold-standard for comparing them with respect to the results obtained by our TS system COMPENDIUM.

For using the abstracts as gold-standards, we carry out an analysis of different properties concerning the number of words and sentences the text of the article and the abstract have, as well as its compression ratio. In this manner, the articles have on average 2000 words (minimum = 949, maximum = 4464) and 83 sentences (minimum = 44, maximum = 151), whereas the abstracts only 162 words (minimum = 64, maximum = 341) and 6 sentences (minimum = 2, maximum = 15). This means that the compression ratio for the abstracts with respect to the text of the articles is around 8%. As can be deduced, this compression ratio is very high, since the articles are very long compared to the length of the abstracts.

4.2. Analysis of the model summaries

Besides analyzing the statistical properties of the corpus, we are also interested in the narrative style of the abstracts, in order to quantify and understand their nature with respect to the types of summaries COMPENDIUM produces.

From this analysis, we found out that 82% of the abstracts had an abstractive nature, and they were created by identifying important fragments of information in the text (instead of complete sentences), and then, generalizing and connecting them with others in a coherent way, through discourse markers and linking phrases. On the contrary, the remaining 18% of the abstracts had an extractive nature, thus containing different sentences that appeared in the text. In these cases, the percentage of identical sentences ranged between 9% and 60% the length of the abstract, and therefore, we can observe that in some cases, more than half of the content of the abstract was extracted directly from the text of the article.

In order to illustrate this fact, Table 3 shows a fragment of an article and its abstract together with the correspondence of their sentences. This fragment of the article and its abstract are the same as the ones reported in Tables 1 and 2. On the one hand, 50% of the sentences in the abstract are found identically within the content of the article (sentences 1 and 2 in the abstract correspond to sentences 1 and 78 in the article) and the remaining sentences (sentences 3 and 4) were generated from relevant pieces of information that appears in the original document (e.g. sentence 3 contains information from sentences 8, 26, and 81). On the other hand, sentence 4 has been created by the author and consequently, it does not have a direct correspondence to any other sentence in the text of the article. As it can be noticed, it synthesizes the discussion of several diagnoses in just one sentence.

Therefore, as a result of this analysis, one may think that a TS approach that combines extractive with abstractive techniques together can be more appropriate to tackle this task. In this article, we analyze two types of summaries generated by COMPENDIUM: extractive (COMPENDIUM_E) and abstractive-oriented (COMPENDIUM_{E-A}), where the latter is, indeed, a hybrid approach, which combines extractive and abstractive information.

¹ <http://www.sciencedirect.com/science/journal/15689972>

Table 2

Abstract of the article shown in Table 1.

Abstract included in an article:
<p>1: Histologic examination of lesions plays a key role in the diagnostics of cutaneous lupus erythematosus (LE). 2: LE has a broad spectrum of histopathological signs, which are related to the stages of the lesions. 3: In addition to the main subtypes of LE, we report on special manifestations like Rowell's-syndrome and Chilblain LE, and give an account of Kikuchi–Fujimoto disease (histiocytic necrotizing lymphadenitis), which may be associated with systemic LE. 4: Furthermore the most considerable histopathologic differential diagnoses are discussed.</p>

4.3. Experiments

COMPENDIUM is employed for producing generic summaries of the data set previously explained. In particular, we generate two kinds of summaries: i) extractive summaries with COMPENDIUM_E and ii) abstractive-oriented summaries with COMPENDIUM_{E-A}, so that we can analyze on the one hand if COMPENDIUM is capable of determining the most relevant information, and on the other hand, which summarization approach would be more appropriate, and if they could be used instead of the abstract of the article. In the biomedical domain, articles usually contain an abstract no longer than 250 words. Since it was previously analyzed that the average length of the abstracts in our data set was 162 words, we take this length as a reference for the summaries generated with COMPENDIUM, since it is even shorter than the standard length for abstracts (i.e., 250 words).

It is worth mentioning that for generating the summaries neither the keywords of the original article nor the information in the titles or in the abstract have been taken into consideration. Moreover, before passing the articles through COMPENDIUM, the data set is cleaned and only the main content of each article is kept for further processing. In other words, the outline, bibliographic entries, keywords, figures and tables are removed.

Table 4 shows two examples of the types of summaries generated with COMPENDIUM_E and COMPENDIUM_{E-A}. The source document for these summaries is the article showed in Tables 1 and 3.

As it can be seen, the resulting summary for COMPENDIUM_{E-A} has also some sentences in common with COMPENDIUM_E, (e.g., sentences 2, 6, 7 and 8), whereas others have been compressed or merged (e.g., 1, 3, 4 and 5).

5. Evaluation and discussion

The aim of this section is to explain the evaluation performed and show the results obtained together with a discussion. For evaluating the summaries generated by COMPENDIUM, two types of evaluation were conducted: quantitative (Subsection 5.1) and qualitative (Subsection 5.2).

The goal of the quantitative evaluation is to assess whether the automatic summaries contain the main information from the source documents. In order to determine the most important information, we take into consideration the original abstracts of the articles and we compare them with the ones generated by COMPENDIUM_E and COMPENDIUM_{E-A} using ROUGE [20]. Moreover, apart from evaluating the information contained in the summaries, we also want to measure the user satisfaction towards the automatic summaries, thus carrying out a qualitative evaluation. Next, each of these evaluations is explained.

5.1. Quantitative evaluation

In this type of evaluation, the informativeness of the summaries with respect to their content is assessed, i.e., if the summaries are able to capture the most relevant information of the biomedical research papers. For determining which content should be considered more relevant, we take into account the original abstracts provided with the articles for two main reasons: i) the abstracts provided with the articles are representative of the main contents of the whole document; and ii) they have been manually generated, and consequently, humans have decided the essential information of the article to be included in the abstract. Therefore, the content of the summaries generated by COMPENDIUM and the original abstracts is compared using the ROUGE tool. Such comparison is made in terms of n-gram co-occurrence. We employ this tool because it is automatic and it has been widely adopted by the research community for the evaluation of summaries. The most common ROUGE metrics are: ROUGE-1 and ROUGE-2, which compute the number of overlapping unigrams and bigrams, respectively; ROUGE-L, which calculates the longest common subsequence between two summaries; and finally, ROUGE-SU4, which measures the overlap of skip-bigrams an automatic summary contains with respect to a model one, with a maximum distance of four words between them. In addition, in order to broaden the evaluation, and analyze how a state-of-the-art summarizer would perform, we also use MS-Word Summarizer 2007² for producing summaries of the same length. This allows us to compare our approaches with another state-of-the-art summarizer. Table 5 shows the results obtained in percentages, remarking the statistical significances according to a *t*-test performed ($p \leq .05$).

As it can be seen, our both TS approaches (extractive and abstractive, with COMPENDIUM_E and COMPENDIUM_{E-A}, respectively) are comparable with respect to the state-of-the-art TS tool (i.e., MS-Word 2007 Summarizer). Regarding COMPENDIUM_{E-A}, it is worth mentioning that the precision obtained is higher for ROUGE-1 and ROUGE-L metrics, compared to the remaining approaches, being also statistically significant for ROUGE-1. However, its recall is lower, so in the end the final value of F-measure is negatively affected,

² <http://www.microsoft.com/education/autosummarize.aspx>

Table 3

Sentence correspondence between an article and its abstract.

Abstract included in a biomedical article:	Fragment of the article:
1: Histologic examination of lesions plays a key role in the diagnostics of cutaneous lupus erythematosus (LE). 2: LE has a broad spectrum of histopathological signs, which are related to the stages of the lesions. 3: In addition to the main subtypes of LE, we report on special manifestations like Rowell's-syndrome and Chilblain LE, and give an account of Kikuchi–Fujimoto disease (histiocytic necrotizing lymphadenitis), which may be associated with systemic LE. 4: Furthermore the most considerable histopathologic differential diagnoses are discussed.	1: Histologic examination of lesions plays a key role in the diagnostics of cutaneous lupus erythematosus (LE). 78: LE has a broad spectrum of histopathological signs, which are related to the stages of the lesions. 8: From the histological standpoint, LE can be classified only into early, fully developed, late LE, and special manifestations of LE. 26: If interface dermatitis becomes manifest at its maximal expression, Rowell's syndrome must be considered. 81: LE may be associated with Kikuchi–Fujimoto disease (histiocytic necrotizing lymphadenitis).

performing COMPENDIUM_E better than COMPENDIUM_{E-A} for most of the metrics. This is due to the fact that for building the abstractive summaries, we rely on the sentences detected as important in the relevance detection stage, and we compress or merge some information within them. Therefore, the resulting summaries are shorter than the extracts, and since no extra information is added, the recall value will never be higher than it is for COMPENDIUM_E . One possible solution to address this issue would be to rely on the source document and generate the new sentences from it instead of the most relevant sentences. Another strategy would be to include in the COMPENDIUM_{E-A} summary the next highest ranked sentence in the document according to the relevance detection stage that were not included in the extract, because of summary length restrictions. Regarding COMPENDIUM_E , it achieves higher results than MS-Word 2007 for most of the metrics, except for ROUGE-L; however, there are no statistical differences between them, except for ROUGE-1.

Although from the results obtained we show that COMPENDIUM is capable to identify and extract the most relevant information from documents, we need to evaluate the generated summaries with respect to other criteria, in order to confirm whether these summaries are good.

As a consequence, for evaluating also the information contained in the summaries, we relied on the keywords provided with the articles. In this manner, we wanted to check which ones appeared identically in the summaries. Our hypothesis was that since keywords represent essential concepts of the document, if summaries included them, this would mean that they would capture the important information. Therefore, for this experiment, we took into account the keywords of each article and we analyzed whether they appeared or not in the summaries (the original abstract, and the summaries generated by COMPENDIUM_E and COMPENDIUM_{E-A}). For each article, the author assigned 4 keywords on average (minimum = 2, maximum = 9). The preliminary analysis conducted over the original abstracts of the articles, showed that only 2 abstracts (4% of the abstracts) included all the keywords, whereas 31 abstracts (62%) included less than the 50% of them. Moreover, a high percentage of abstracts (34%) did not include any keyword at all.

In light of these results, we decided to expand the keywords with equivalent concepts, by using the MeSH thesaurus,³ and analyze to what extent the summaries contained also synonyms of the keywords or equivalent terms. Again, we first analyzed the original abstracts, and we found out that, as in the previous analysis, most of the expanded keywords were not present in the original abstracts, nor in the summaries generated by COMPENDIUM . In the biomedical domain, keywords synthesize concepts that globally describe the content of the article, and they have to be selected from a normalized set of keywords (i.e., MeSH). It normally occurs that when authors do not find a keyword for the specific concept in these sets, they select a general term, being therefore more ambiguous. In addition, since we deal with a technical context, the variability of vocabulary decreases, and instead of using synonyms of the terms, authors tend to use their own abbreviations, which do not appear in the aforementioned normalized sets (e.g., polymorphous light eruption (PLE) taken from the fragment of the biomedical article shown in Table 1 does not appear in MeSH). This strategy helps them to ensure that the abstract and the text do not surpass the maximum number of words allowed with respect to the journal requirements.

Therefore, in light of the analysis conducted and the results obtained, it is necessary to conduct a qualitative evaluation where we measure the user satisfaction of the generated summaries with respect to different aspects, including issues regarding the most important topics stated in the summaries, so that we can confirm the appropriateness of the summaries produced using COMPENDIUM in its extractive and abstractive variants (COMPENDIUM_E and COMPENDIUM_{E-A} , respectively). The next Subsection explains in detail the qualitative evaluation performed.

5.2. Qualitative evaluation

In this evaluation, we aim at assessing the user satisfaction with respect to the generated summaries with COMPENDIUM . For this purpose, we performed a qualitative evaluation. A group of 10 assessors were involved in the process, where they had to evaluate

³ <http://www.ncbi.nlm.nih.gov/mesh>

Table 4

Examples of summaries generated with COMPENDIUM.

Extractive summary generated with COMPENDIUM _E :	Abstractive-oriented summary generated with COMPENDIUM _{E-A} :
<p>1: Histologic examination of lesions plays a key role in the diagnostics of cutaneous lupus erythematosus LE.</p> <p>2: LE has a broad spectrum of histological signs which are related to the stages of the lesions, but some signs apply to all stages e.g. mucin deposition.</p> <p>3: Histologic findings of skin lesions are essentially identical for systemic lupus erythematosus SLE and cutaneous LE.</p> <p>4: From the histological standpoint, LE can be classified only into early, fully developed, late LE, and special manifestations of LE.</p> <p>5: The early histologic findings of LE lesions are sparse superficial perivascular lymphocytic infiltrates, neutrophils and sometimes nuclear dust immediately beneath the dermoepidermal junction.</p> <p>6: Few individual necrotic keratinocytes and focal vacuolar alteration of basal cells may occur.</p> <p>7: Fully developed lesions are characterized by moderately dense to dense perivascular and periappendageal lymphocytic infiltrates in the papillary and reticular dermis with abundant mucin deposition in the reticular dermis.</p> <p>8: According to Kuhn et al. the presence of even slight epidermal or junctional involvement should exclude LE tumidus.</p>	<p>1: LE lesions play a key role in the diagnostics.</p> <p>2: LE has a broad spectrum of histological signs which are related to the stages of the lesions, but some signs apply to all stages e.g. mucin deposition.</p> <p>3: LE lesions are essentially identical for systemic lupus erythematosus SLE.</p> <p>4: LE can be classified only into early histologic.</p> <p>5: LE lesions are sparse superficial perivascular lymphocytic infiltrates neutrophils and sometimes nuclear dust immediately beneath the dermoepidermal junction.</p> <p>6: Few individual necrotic keratinocytes and focal vacuolar alteration of basal cells may occur.</p> <p>7: Fully developed lesions are characterized by moderately dense to dense perivascular and periappendageal lymphocytic infiltrates in the papillary and reticular dermis with abundant mucin deposition in the reticular dermis.</p> <p>8: According to Kuhn et al. the presence of even slight epidermal or junctional involvement should exclude.</p>

independently the content of the extractive and abstractive-oriented summaries (COMPENDIUM_E and COMPENDIUM_{E-A}, respectively) with respect to the human abstracts and the text of the articles.

The evaluation focused on three questions (Q1: *The summary reflects the most important issues of the document*; Q2: *The summary allows the reader to know what the article is about*; and Q3: *After reading the original abstract provided with the article, the alternative summary is also valid*), and the user satisfaction was measured with respect to the content and appropriateness of the summaries by using a five-point Likert type scale (1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 agree; and 5 = strongly agree).

The first question (Q1) evaluates how informative a summary is, i.e., whether the summary contains relevant information or not. For evaluating it, the assessors have to read the article, in order to identify what information they consider more important, and check if it appears in the summary. The second question (Q2) allows us to determine to what extent the automatic summaries are indicative, thus providing a general idea of the contents of the article. Finally, the last question (Q3) aims at checking if the automatic summaries are good enough to substitute the original abstracts. Moreover, by rating the summaries, we can also analyze if the users prefer extractive or abstractive summaries, thus determining which approach would be more appropriate.

In Table 6, the statistics of the results for each question and each TS approach are shown. The best assessment is obtained when using abstractive-oriented summaries generated by COMPENDIUM_{E-A} for the three proposed questions.

It is also worth noting the values for the mode in both TS approaches. This value shows which value of the scale appears more frequently. Concerning this, we can observe that for all the questions the most frequent value for COMPENDIUM_{E-A} is 4, whereas for COMPENDIUM_E is 2 for Q1, and 1 for Q2 and Q3. This means that most of the users classified the abstractive-oriented summaries (COMPENDIUM_{E-A}) as “Agree”.

In this sense, in Table 7 we can observe that the percentage of summaries evaluated as “Agree” when using COMPENDIUM_{E-A} (36%, 36% and 28%, for questions Q1, Q2 and Q3, respectively) increases compared to the accumulated percentage obtained with COMPENDIUM_E for the criteria “Agree” and “Strongly Agree” (20%, 26%, and 16%, respectively). In other words, the assessors showed a higher satisfaction towards the summaries generated with COMPENDIUM_{E-A}, and these results were statistically significant with

Table 5

ROUGE results for the different text summarization approaches.

ROUGE	TS approach	Recall	Precision	F _β = 1
ROUGE-1	COMPENDIUM _E	44.02*	40.53	42.20*
	COMPENDIUM _{E-A}	38.66	41.81*	40.20
	MS-Word 2007	43.61	40.46	41.97
ROUGE-2	COMPENDIUM _E	14.46	13.21	13.81
	COMPENDIUM _{E-A}	11.49	12.63	12.03
	MS-Word 2007	13.67	12.85	13.24
ROUGE-L	COMPENDIUM _E	29.84	27.30	28.51
	COMPENDIUM _{E-A}	25.95	28.01	26.94
	MS-Word 2007	30.41	28.11	29.21
ROUGE-SU4	COMPENDIUM _E	18.92	17.29	18.07
	COMPENDIUM _{E-A}	15.60	16.95	16.24
	MS-Word 2007	17.81	16.60	17.18

* Means that the result is statistically significant ($p \leq .05$).

Table 6

Descriptive statistics for the qualitative evaluation.

	Q1		Q2		Q3	
	COMPENDIUM _E	COMPENDIUM _{E-A}	COMPENDIUM _E	COMPENDIUM _{E-A}	COMPENDIUM _E	COMPENDIUM _{E-A}
Mean	2.40	2.80	2.58	3.08	2.32	2.76
Standard dev.	1.088	1.161	1.279	1.175	1.269	1.153
Median	2.00	3.00	2.50	3.00	2.00	3.00
Mode	2	4	1	4	1	4 ^a

^a Q3 has two modes, but we only reflect the highest one.

respect to the ones obtained with COMPENDIUM_E according to a Chi-square test ($\chi^2 = 43.281$, $p = .000$; $\chi^2 = 38.808$, $p = .001$; $\chi^2 = 35.390$, $p = .004$ for each of the questions, respectively). This means that according to the results obtained when assessing the user satisfaction, summaries generated with COMPENDIUM_{E-A} included relevant information with regard to the text of the article, and they provided an overview of the article. In addition, these summaries were good enough for serving as a surrogate of the abstract written by the authors.

6. Conclusion and future work

In this paper we presented COMPENDIUM, a text summarization system applied to the generation of abstracts of research papers for the biomedical domain. In particular, two types of generic summaries were produced: extractive and abstractive-oriented, with the corresponding variants of COMPENDIUM: COMPENDIUM_E and COMPENDIUM_{E-A}, respectively. The extractive approach only selected and extracted the most relevant sentences, whilst for the abstractive-oriented one, we proposed a novel approach, which combined extractive and abstractive techniques, by incorporating an information compression and fusion stage once the most important content was identified.

With the aim of determining if COMPENDIUM was able to generate good summaries, as well as studying which approach should be more appropriate within the biomedical context, we carried out a quantitative and qualitative evaluation. On the one hand, the goal of the quantitative evaluation was to assess the information contained in the summaries, i.e., whether the system was able to identify the relevant information. On the other hand, the qualitative evaluation was conducted by means of a user satisfaction study where real users rated the generated summaries according to three criteria: their topics, content, and suitability for substituting the original abstracts in the research articles.

Although the results concerning the quantitative evaluation benefited the extractive summaries, showing that the summaries generated with COMPENDIUM_{E-A} contained less similar information to the human-written abstracts included in the articles than the summaries produced with the extractive approach (COMPENDIUM_E), we can conclude that according to the satisfaction users have towards both kind of summaries, the abstractive-oriented summaries COMPENDIUM_{E-A} are the most appropriate. From these results, we can draw two main conclusions: i) COMPENDIUM is useful for producing automatic summaries of biomedical research papers, since both TS approaches are able to keep the most important information; ii) abstractive-oriented summaries generated with COMPENDIUM_{E-A} are better from a human perspective. Therefore, we showed the appropriateness of COMPENDIUM as a TS system, as well as our proposed strategy for facing abstractive summarization.

Table 7

Percentage of summaries evaluated according to each degree in the Likert type scale for the proposed questions.

	Likert scale	COMPENDIUM _E	COMPENDIUM _{E-A}	
Q1	Strongly disagree	22%	16%	$\chi^2 = 43.281$ ($p = .000$)
	Disagree	38%	28%	
	Neither agree nor disagree	20%	18%	
	Agree	18%	36%	
	Strongly agree	2%	2%	
Q2	Strongly disagree	26%	12%	$\chi^2 = 38.808$ ($p = .001$)
	Disagree	24%	20%	
	Neither agree nor disagree	24%	24%	
	Agree	18%	36%	
	Strongly agree	8%	8%	
Q3	Strongly disagree	32%	16%	$\chi^2 = 35.390$ ($p = .004$)
	Disagree	30%	28%	
	Neither agree nor disagree	22%	24%	
	Agree	6%	28%	
	Strongly agree	10%	4%	

In the future, we plan to analyze other variants of the proposed approach for building abstracts, such as taking the source document as a starting point instead of the extractive summary. Moreover, focusing also in the biomedical domain, we want to study other applications where automatic summaries are useful. For instance, the automatic generation of newsletters, that could help health experts and researchers to keep updated on the main novelties concerning their field of expertise, by just providing periodically a brief summary of recent biomedical articles, new treatments, drugs, etc.

Acknowledgment

This research was partially supported by the FPI grant (BES-2007-16268) and the project grants TEXT-MESS (TIN2006-15265-C06-01), TEXT-MESS 2.0 (TIN2009-13391-C04) and LEGOLANG (TIN2012-31224) from the Spanish Government. It has been also funded by the Valencian Government (grant no. PROMETEO/2009/119 and ACOMP/2011/001). The authors would like to thank Ester Boldrini, Paloma Moreda, Isabel Moreno, Helena Burruezo, Jesús Hermida, Jorge Cruaños, Fernando Peregrino, Héctor Llorens, Felipe Sellés, and Rubén Izquierdo for their help in the manual evaluation of the summaries.

References

- [1] M. Abulaish, L. Dey, Biological relation extraction and query answering from medline abstracts using ontology-based text mining, *Data & Knowledge Engineering* 61 (2007) 228–262.
- [2] A. Balahur, A. Montoyo, Multilingual feature-driven opinion extraction and summarization from customer reviews, *Proceedings of the 13th International Conference on Applications of Natural Language to Information Systems, Lecture Notes in Computer Science*, vol. 5039, Springer, 2008, pp. 345–346.
- [3] A. Belz, Automatic generation of weather forecast texts using comprehensive probabilistic generation-space models, *Natural Language Engineering* 14 (4) (2008) 431–455.
- [4] A. Bossard, M. Génèreux, T. Poibeau, CBSEAS, a summarization system integration of opinion mining techniques to summarize blogs, *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 2009, pp. 5–8.
- [5] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer Networks ISDN Systems* 30 (1998) 107–117.
- [6] G. Carenini, J.C.K. Cheung, Extractive vs. NLG-based abstractive summarization of evaluative text: the effect of corpus controversy, *Proceedings of the 5th International Natural Language Generation Conference*, 2008, pp. 33–40.
- [7] M. Cataldi, K.S. Candan, M.L. Sapino, Narrative-based taxonomy distillation for effective indexing of text collections, *Data & Knowledge Engineering* 72 (2012) 103–125.
- [8] W. Chen, P. Chundi, Extracting hot spots of topics from time-stamped documents, *Data & Knowledge Engineering* 70 (7) (2011) 642–660.
- [9] G. Costa, G. Manco, R. Ortale, E. Ritacco, Hierarchical clustering of xml documents focused on structural components, *Data & Knowledge Engineering* 84 (2013) 26–46.
- [10] D.M. Dunlavy, D.P. O'Leary, J.M. Conroy, J.D. Schlesinger, QCS: a system for querying, clustering and summarizing documents, *Information Processing and Management* 43 (6) (2007) 1588–1605.
- [11] G. Erkan, D.R. Radev, Lexrank: graph-based lexical centrality as salience in text summarization, *Journal of Artificial Intelligence Research* 22 (2004) 457–479.
- [12] O. Ferrández, D. Micol, R. Muñoz, M. Palomar, A perspective-based approach for solving textual entailment recognition, *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, 2007, pp. 66–71.
- [13] K. Filippova, Multi-sentence compression: finding shortest paths in word graphs, *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 322–330.
- [14] J. Gardner, L. Xiong, An integrated framework for de-identifying unstructured medical data, *Data & Knowledge Engineering* 68 (2009) 1441–1451.
- [15] T. Givón, *A Functional–typological Introduction*, II, John Benjamins, Amsterdam, 1990.
- [16] A. Hliaoutakis, K. Zervanou, E.G.M. Petrakis, The AMTex approach in the medical document indexing and retrieval application, *Data & Knowledge Engineering* 68 (2009) 380–392.
- [17] M.-Y. Kan, J.L. Klavans, K.R. McKeown, Using the annotated bibliography as a resource for indicative summarization, *Proceedings of the Language Resources and Evaluation Conference*, 2002, pp. 1746–1752.
- [18] M. Kumar, D. Das, S. Agarwal, A. Rudnick, Non-textual event summarization by applying machine learning to template-based language generation, *Proceedings of the 2009 Workshop on Language Generation and Summarisation*, 2009, pp. 67–71.
- [19] P. Lal, S. Rüger, Extract-based summarization with simplification, *Proceedings of the Workshop on Text Summarization in conjunction with the Association for Computational Linguistics*, 2002.
- [20] C.-Y. Lin, ROUGE: a package for automatic evaluation of summaries, *Proceedings of the Association for Computational Linguistics Text Summarization Workshop*, 2004, pp. 74–81.
- [21] M. Litvak, M. Last, M. Friedman, A new approach to improving multilingual summarization using a genetic algorithm, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 927–936.
- [22] M. Liu, W. Li, M. Wu, Q. Lu, Extractive summarization based on event term clustering, *Proceeding of the 45th Annual Conference of Association of Computational Linguistics*, 2007, pp. 185–188.
- [23] H.P. Luhn, The automatic creation of literature abstracts, in: Inderjeet Mani, Mark Maybury (Eds.), *Advances in Automatic Text Summarization*, MIT Press, 1958, pp. 15–22.
- [24] S. Mohammad, B. Dorr, M. Egan, A. Hassan, P. Muthukrishnan, V. Qazvinian, D. Radev, D. Zajic, Using citations to generate surveys of scientific paradigms, *Proceedings of the North American Chapter of the Association of Computational Linguistics*, 2009, pp. 584–592.
- [25] R. Montiel Soto, R. García-Hernández, Comparación de Tres Modelos de Texto para la Generación Automática de Resúmenes, *Procesamiento del Lenguaje Natural* 43 (2009) 303–311.
- [26] A. Nenkova, L. Vanderwende, K. McKeown, A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization, *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2006, pp. 573–580.
- [27] L. Plaza, *Uso de Grafos Semánticos en la Generación Automática de Resúmenes y Estudio de su Aplicación en Distintos Dominios: Biomedicina, Periodismo y Turismo*, (PhD thesis), 2011.
- [28] L. Plaza, A. Díaz, Retrieval of similar electronic health records using UMLS concept graphs, *Proceedings of the 15th International Conference on Applications of Natural Language to Information Systems, Lecture Notes in Computer Science*, vol. 6177, Springer, 2010, pp. 296–303.
- [29] J.J. Pollock, A. Zamora, Automatic abstracting research at chemical abstracts, in: Inderjeet Mani, Mark Maybury (Eds.), *Advances in Automatic Text Summarization*, MIT Press, 1999, pp. 43–49.
- [30] V. Qazvinian, D.R. Radev, Scientific paper summarization using citation summary networks, *Proceedings of the 22nd International Conference on Computational Linguistics*, 2008, pp. 689–696.
- [31] D. Radev, T. Allison, S. Blair-Goldensohn, J. Blitzer, A. Celebi, E. Drabek, W. Lam, D. Liu, J. Otterbacher, H. Qi, H. Saggion, S. Teufel, M. Topper, A. Winkel, Z. Zhang, MEAD — a platform for multidocument multilingual text summarization, *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 2004, pp. 699–702.
- [32] D. Rosaci, G. Terracina, D. Ursino, A framework for abstracting data sources having heterogeneous representation formats, *Data & Knowledge Engineering* 48 (2004) 1–38.
- [33] H. Saggion, SUMMA: a robust and adaptable summarization tool, *Traitement Automatique des Langues* 49 (2008) 103–125.

- [34] H. Saggion, A classification algorithm for predicting the structure of summaries, *Proceedings of the Workshop on Language Generation and Summarisation*, 2009, pp. 31–38.
- [35] H. Saggion, G. Lapalme, Selective analysis for automatic abstracting: evaluating indicativeness and acceptability, *Proceedings of Content-Based Multimedia Information Access*, 2000, pp. 747–764.
- [36] H. Saggion, E. Lloret, M. Palomar, Can text summaries help predict ratings? A case study of movie reviews, *Proceedings of the 17th International Conference on Applications of Natural Language to Information Systems (NLDB 2012)*, *Lecture Notes in Computer Science*, vol. 7337, 2012, pp. 271–276.
- [37] C. Sauper, R. Barzilay, Automatically generating wikipedia articles: a structure-aware approach, *Proceedings of the 47th Annual Conference of the Association of Computational Linguistics*, 2009, pp. 208–216.
- [38] D. Shen, Q. Yang, Z. Chen, Noise reduction through summarization for web-page classification, *Information Processing and Management* 43 (6) (2007) 1735–1747.
- [39] S.B. Silveira, A. Branco, Extracting multi-document summaries with a double clustering approach, *Proceedings of the 17th International Conference on Applications of Natural Language to Information Systems (NLDB 2012)*, *Lecture Notes in Computer Science*, vol. 7337, 2012, pp. 70–81.
- [40] K. Spärck Jones, Automatic summarising: the state of the art, *Information Processing and Management* 43 (6) (2007) 1449–1481.
- [41] D. Tatar, A.D. Mihis, D. Lupsa, Text entailment for logical segmentation and summarization, *Proceedings of the 13th International Conference on Applications of Natural Language to Information Systems*, *Lecture Notes in Computer Science*, vol. 5039, Springer, 2008, pp. 233–244.
- [42] K.-F. Wong, M. Wu, W. Li, Extractive summarization using supervised and semi-supervised learning, *Proceedings of the 22nd International Conference on Computational Linguistics*, 2008, pp. 985–992.
- [43] J. Yu, E. Reiter, J. Hunter, C. Mellish, Choosing the content of textual summaries of large time-series data sets, *Natural Language Engineering* 13 (1) (2007) 25–49.
- [44] D. Zajic, B.J. Dorr, J. Lin, R. Schwartz, Multi-candidate reduction: sentence compression as a tool for document summarization tasks, *Information Processing and Management* 43 (6) (2007) 1549–1570.
- [45] A. Zamanifar, O. Kashefi, AZOM: a persian structured text summarizer, *Proceedings of the 16th International Conference on Applications of Natural Language to Information Systems*, *Lecture Notes in Computer Science*, vol. 6716, Springer, 2011, pp. 234–237.
- [46] L. Zhou, M. Ticeira, E. Hovy, Multi-document biography summarization, *Proc. of the International Conference on Empirical Methods in NLP*, 2004, pp. 434–441.



Dr. Elena Lloret — is a post-doctoral researcher at the University of Alicante at the European Project FIRST: A Flexible Interactive Reading Support Tool (grant no. FP7-287607). She is a Computer Science graduate and obtained her PhD at the University of Alicante. Her main field of interest is text summarization, text simplification and text comprehension. She is the author of over 25 scientific publications in relevant journals and international conferences. She has been collaborating with international researchers and has participated in a number of projects at a national level (TIN2006-15265-C06, TIN2009-13391-C04). She has also been collaborating with international groups in Wolverhampton, Sheffield and Edinburgh.



Prof. Dr. M Teresa Romá-Ferri — is a full professor of the Faculty of Health Sciences at the University of Alicante and a member of the Natural Language Processing and Information Systems Research Group of the same university. Her main teaching area focuses on health information systems and nursing informatics. She received her Nursing graduate and PhD in Computer Science at the University of Alicante, and her Master's degree in Information Science at the Open University of Catalonia (UOC). Her main field of interest is Nursing Science and Natural Language Processing (NLP) and in particular Text Summarization, Information Extraction and Ontology Development. She has been collaborating with international researchers, coordinated and has participated in a number of projects at a national level (PI051438, TIN2012-31224). She is the author of over 50 scientific publications in relevant journals and international conferences.



Prof. Dr. Manuel Palomar — is the University President of the University of Alicante and head of the Natural Language Processing and Information Systems Research Group of the same university. He is also a full professor of this university since 1991 and his main teaching area focuses on the analysis, design and management of databases, datawarehouses, and information systems. He received his Master's degree and PhD in Computer Science at the Polytechnic University of Valencia, Spain. His research interests are Human Language Technologies (HLT) and Natural Language Processing (NLP), in particular Text Summarization, Semantic Roles, Textual Entailment, Information Extraction and Anaphora Resolution. He has supervised more than 12 theses and he is the author of more than 70 scientific publications on international journals and conferences on different topics related to HLT and NLP. Furthermore, he has coordinated and been involved in a number of regional, national and international research projects funded by the Generalitat Valenciana (Valencian Government), the Ministry of Science and Innovation (Spanish Government) and the European Council.