COMPENDIUM: A Text Summarization System for Generating Abstracts of Research Papers

Elena Lloret[1,⋆], María Teresa Romá-Ferri[2], and Manuel Palomar[1]

[1] Dept. Lenguajes y Sistemas Informáticos
Universidad de Alicante
Apdo. de correos, 99, E-03080 Alicante, Spain
{elloret,mpalomar}@dlsi.ua.es
[2] Department of Nursing, University of Alicante, Spain
mtr.ferri@ua.es

Abstract. This paper presents COMPENDIUM, a text summarization system, which has achieved good results in extractive summarization. Therefore, our main goal in this research is to extend it, suggesting a new approach for generating abstractive-oriented summaries of research papers. We conduct a preliminary analysis where we compare the extractive version of COMPENDIUM (COMPENDIUM$_E$) with the new abstractive-oriented approach (COMPENDIUM$_{E-A}$). The final summaries are evaluated according to three criteria (content, topic, and user satisfaction) and, from the results obtained, we can conclude that the use of COMPENDIUM is appropriate for producing summaries of research papers automatically, going beyond the simple selection of sentences.

Keywords: Human Language Technologies, NLP Applications, Text Summarization, Information Systems.

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) [20]. In particular, TS has been shown to be very useful as a stand-alone application [2], as well as in combination with other systems, such as text classification [19].

⋆ This research has been supported by the FPI grant (BES-2007-16268) from the Spanish Ministry of Science and Innovation, under the project TEXT-MESS (TIN2006-15265-C06-01) and project grant no. TIN2009-13391-C04-01, both funded by 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 Cruañes, Fernando Peregrino, Héctor Llorens, Felipe Sellés, and Rubén Izquierdo for their help in the manual evaluation of the summaries.

R. Muñoz et al. (Eds.): NLDB 2011, LNCS 6716, pp. 3–14, 2011.

Among the wide range of applications of TS, one especially interesting concerns the generation of abstracts of research papers. In the research context, every article published in a conference, journal, etc. must include an abstract written by the author, which is a summary of the main topics and findings addressed in the research presented. These abstracts are very important to provide an idea of what the article is about, and they can be used not only by humans, but also by automatic systems for indexing, searching and retrieving information without having to process the whole document. TS can be very useful for automatically generating such abstracts. However, to carry out this process is very challenging and difficult. This is shown by the fact that, although there have been some attempts to generate abstracts in the recent years [4], [16], most of the current work on TS still focuses on extractive summarization[1] [10], [12], [21]. 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.

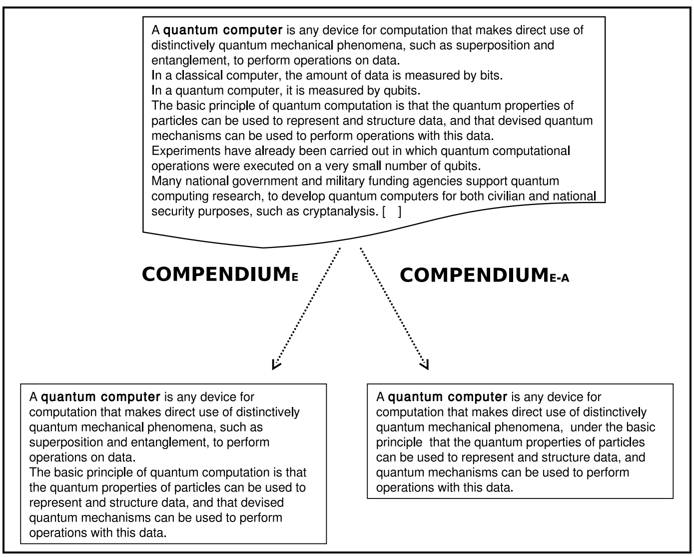

Fig. 1. Intuitive illustration of the COMPENDIUM system

In this paper, our goal is to analyze to what extent TS is useful for a specific application: the automatic generation of research paper abstracts. In particular, we propose two generic TS approaches: one based on a pure extractive summarizer (COMPENDIUM$_E$) and a novel one that combines extractive and abstractive techniques (COMPENDIUM$_{E-A}$). In this preliminary research work, we want to

[1] Extractive summarization consists of selecting the most important sentences of a document to form the final summary, whereas abstractive implies also the generation of new information.

study to what extent such approaches are appropriate to 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. Results show that extractive and abstractive-oriented summaries perform similarly as far as the information they contain, but the latter is more appropriate from a human perspective, when a user satisfaction assessment is carried out. Figure 1 illustrates the differences between an extractive summary (COMPENDIUM$_E$) and an abstractive one (COMPENDIUM$_{E-A}$) in an intuitive way.

The remaining of the paper is structured as follows: Section 2 introduces previous work on TS approaches developed for specific applications. Section 3 describes in detail the data set employed, analysing also the human generated abstracts provided with research papers. Further on, in Section 4 we explain the suggested TS approaches and how we developed them. The experiments carried out, the results obtained together with a discussion for each type of evaluation are provided in Section 5. Finally, the conclusions of the paper together with the future work are outlined in Section 6.

2 Related Work

In this Section, we explain previous work on different approaches that have used TS techniques to address specific tasks (e.g. for generating Wikipedia articles, weather forecast reports, etc.).

Sauper and Barzilay [18] propose an automatic method to generate Wikipedia articles, where specific topic templates, as well as the information to select is 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. For instance, in Kumar et al. [9] 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 [4], 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.

Natural Language Generation (NLG) has been also applied for adding new vocabulary and language structures in summaries. In Yu et al. [22] 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 [1] also suggests a TS approach based on NLG, in order to generate weather forecasts reports automatically.

Another specific application of TS concerns the generation of summaries from biographies. The idea behind multi-document biography summarization is to produce a piece of text containing the most relevant aspects of a specific person. Zhou et al. [23] 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.

The generation of technical surveys has also been addressed in the recent years. In Mohammad et al. [14] citations are used to automatically generate technical surveys. They experimented with three types of input (full papers, abstracts and citation texts), and analyzed different already existing summarization systems to create such surveys, for instance LexRank [5]. Among the conclusions drawn from the experiments, it was shown that multi-document technical survey creation benefits considerably from citation texts.

Our research work focus on generating research abstracts, and in particular analyzing the capabilities of different TS techniques for such purposes. This problem has been addressed in previous literature. For instance, Pollock and Zamora [15] 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 be introduce relevant information, such as "*in conclusion*" or "*the aim of this paper*". Furthermore, Saggion and Lapalme [17] exploited also this kind of information, by means of a set of patterns that were later combined with the information extracted from the document. In our approach, we do not rely on specific patterns, nor we learn the struture or the content from already existing model abstracts. In contrast, we want to analyze the appropriateness of two generic TS approaches to generate abstracts of research papers in particular for the medical domain, and assess to what extent they would be appropriate for this task.

3 Description of the Data Set: Research Papers

We collected a set of 50 research articles from specialized journals of medicine directly from the Web. Each article contains a human-written abstract which will be used as a model summary. Furthermore, the articles also include an outline, a set of keywords, as well as several figures and tables. The data set was first preprocessed and only the main content of each document was kept for further processing. In other words, the outline, bibliographic entries, keywords, figures and tables were removed. Table 1 shows some properties of the data set after the preprocessing stage. Additionally, it also contains some figures corresponding to the human-written abstracts. As can be seen, the documents are rather long (more than 2,000 words on average), whereas the abstracts are shorter (162 words on average), thus being their compression ratio with respect to the whole article quite high (13%).

3.1 Analysis of the Model Summaries

Before conducting the TS process, an analysis of the human-written abstracts is carried out. The reason why such analysis is done is to quantify and understand

Table 1. Properties of the data set and the human-written abstracts

Avg. number of sentences per document	83.03
Avg. number of words per document	2,060
Avg. number of words per human abstracts	162.7
Compression ratio for abstracts wrt. documents	13%

the nature of human-written abstracts. As we expected, only a small number of abstracts (18%) have an extractive nature, containing at least one identical sentence expressed later in the article. Instead of a pure extractive nature, this can be considered a combination of extractive and abstractive strategies. The percentage of identical sentences ranged between 9% and 60%, depending on the abstract. On the contrary, the 82% of the abstracts have an abstractive nature, but these abstracts have not been created following a pure abstractive strategy either, since important fragments of information are selected identically from the source document, and then they are generalized and connected with others in a coherent way, through discourse markers and linking phrases. 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.

The first and the second summaries in Table 2 correspond to a fragment of a medical article and a human abstract, respectively. It is worth noticing that 50% of the sentences in the human abstract are identical (sentences 1 and 2 correspond to sentences 1 and 78 in the original article) and 50% are new sentences. Moreover, as it can be seen the third and the fourth sentence in the human abstract have been generated from relevant pieces of information that appears in the original document (e.g. sentence 3 contains information from sentences 8, 26, 33 and 81).

4 Text Summarization Approach: COMPENDIUM

In this Section, the two suggested approaches for generating summaries are explained. First we described COMPENDIUM$_E$, a pure extractive TS approach (Section 4.1), 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 4.2).

Besides a fragment of a medical article and a human abstract, Table 2 shows two summaries generated with COMPENDIUM$_E$ and COMPENDIUM$_{E-A}$, which are explained in the next subsections. It is worth mentioning that these approaches produce generic summaries, and for generating them neither the keywords of the original article nor the information in the titles or in the abstract have been taken into consideration. As it can be seen, the first two sentences in the summary generated with COMPENDIUM$_E$ are also presented in the human abstract. Moreover, the resulting summary for COMPENDIUM$_{E-A}$ has also some sentences in common with COMPENDIUM$_E$, whereas others have been compressed or merged.

Table 2. Example of different types of summaries and the corresponding source article

Fragment of the original medical 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 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 it's 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. [...] **81:** LE may be associated with Kikuchi-Fujimoto disease (histiocytic necrotizing lymphadenitis). [...]
Human Abstract: **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.
Extractive summary generated with COMPENDIUM$_E$: **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 . **3:** Histologic findings of skin lesions are essentially identical for systemic lupus erythmatosus SLE and cutaneous LE. **4:** From the histological standpoint, LE can be classified only into early, fully developed, late LE, and special manifestations of LE. **5:** The early histologic findings of LE lesions are sparse superficial perivascular lymphocytic infiltrates, neutrophils and sometimes nuclear dust immediately beneath the dermoepidermal junction. **6:** Few individual necrotic keratinocytes and focal vacuolar alteration of basal cells may occur. **7:** According to Kuhn et al. the presence of even slight epidermal or junctional involvement should exclude LE tumidus.
Abstractive-oriented summary generated with COMPENDIUM$_{E-A}$: **1:** LE lesions plays a key role in the diagnostics. **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 . **3:** LE lesions are essentially identical for systemic lupus erythmatosus SLE. **4:** LE can be classified only into early histologic. **5:** LE lesions are sparse superficial perivascular lymphocytic infiltrates neutrophils and sometimes nuclear dust immediately beneath the dermoepidermal junction. **6:** Few individual necrotic keratinocytes and focal vacuolar alteration of basal cells may occur. **7:** According to Kuhn et al the presence of even slight epidermal or junctional involvement should exclude.

4.1 COMPENDIUM$_E$: Extractive Summarization

This TS approach relies on four main stages for producing extractive summaries:

- **Preprocessing:** 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 removal.
- **Redundancy removal:** a Textual Entailment (TE) tool [6] is used to detect and remove repeated information. In this sense, two sentences containing a true entailment relationship are considered equivalent, and therefore, the one which is entailed is discarded.
- **Sentence relevance:** this stage computes a score for each sentence depending on its importance, relying on two features: Term Frequency (TF) [13] and the Code Quantity Principle (CQP) [8]. On the one hand, TF allows us to determine the topic of a document[2], whereas on the other hand, the CQP states that the most important information within a text is expressed by

[2] Stop words are not taken into account.

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 are considered 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.

4.2 COMPENDIUM$_{E-A}$: Abstractive-Oriented Summarization

Our second TS approach, COMPENDIUM$_{E-A}$, combines extractive and abstractive techniques in the following manner: we take as a basis the COMPENDIUM$_{E}$ approach described in the previous section, 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:

- **Word graph generation:** for generating new sentences, we rely on word graphs adopting a similar approach to the one described in [7]. 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 [3]. 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[3]; 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). In the future, we plan to extend the set of rules in order to detect a higher number of incorrect sentences, for instance, taking into account if the sentence ends with an adjective.

- **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[4]. 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.

5 Results and Discussion

In this Section, we explain the evaluation performed and we show the results obtained together with a discussion. We set up three different types of evaluation. In the first one, we use the human-written abstracts of the articles and we compare them with the ones generated by our approaches using a state-of-the-art evaluation tool. The second evaluation aims at determining to what extent our generated summaries contain the main topics of the research articles. In the third evaluation, we assess the summaries with regard to user satisfaction. Finally a discussion of the results is provided.

5.1 Comparison with Human Abstracts

The goal of this evaluation is to assess the informativeness of the summaries with respect to their content. Following the approaches described in Section 4, we generated summaries of approximately 162 words for the 50 documents pertaining to the data set (Section 3). We also use MS-Word Summarizer 2007[5] for producing summaries of the same length in order to allow us to compare our approaches with an state-of-the-art summarizer. We then use ROUGE-1 [11] for assessing how many common vocabulary there is between the generated summaries and the human-written abstract. Table 3 shows these results in percentages.

[3] We assume that three words (i.e., subject+verb+object) is the minimum length for a complete sentence.

[4] We use the cosine measure to compute the similarity between two sentences, and a threshold has been empirically set to 0.5.

[5] http://www.microsoft.com/education/autosummarize.aspx

Table 3. ROUGE-1 results for the different text summarization approaches

TS Approach	Recall	Precision	F-measure
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

As can be seen, our both TS approaches are comparable with respect to the state of the art TS tool (i.e., MS-Word 2007 Summarizer). Regarding our abstractive-oriented approach (COMPENDIUM$_{E-A}$), it is worth mentioning that it obtains higher precision than the remaining approaches. However, its recall is lower, so in the end the final value of F-measure is negatively affected. This is due to the fact that for this TS approach we take as input the extracts previously generated, and we compress or merge some information within them. Therefore, the summaries produced using COMPENDIUM$_{E-A}$ will be shorter than the extracts, and since no extra information is added, the recall value will be never higher than it is for COMPENDIUM$_E$. In order to solve this problem, we could either generate the new sentences from the original article instead of using the extract as input; or include in the COMPENDIUM$_{E-A}$ summary the next highest ranked sentence in the document according to the relevance detection stage (Section 4.1) that we could not include it in the extractive summary, because the desired length had already been reached.

5.2 Topic Identification

The objective of this evaluation is to assess the generated summaries with respect to the topics they contain (i.e., how indicative they are). Together with the content of the article and the abstract, a number of keywords were also included (5 on average). These keywords usually reflect the most important topics dealt in the article. Consequently, we want to analyze to what extent such keywords appear in the summaries generated by our suggested TS approaches (COMPENDIUM$_E$ and COMPENDIUM$_{E-A}$). If our summaries are able to contain such keywords, it will mean that they are indicative of the content of the source document, and therefore, they will be appropriate to provide an idea of what the article is about. In order to compute the number of keywords a summary contains, we calculate how many of them a summary contains, and we divide this result by the total number of keywords the corresponding article has. Table 4 shows the results obtained for this evaluation.

Table 4. Percentage of topics that resulting summaries contain

	% Correct Topics			
	< 25%	< 50%	< 75%	75-100%
COMPENDIUM$_E$	5%	12.5%	47.5%	35%
COMPENDIUM$_{E-A}$	7.5%	17.5%	42.5%	32.5%

As it can be seen, a considerable percentage of summaries are able to reflect at least half of the topics of the articles (82.5% and 75%, for COMPENDIUM$_E$ and COMPENDIUM$_{E-A}$, respectively). It is worth stressing upon the fact that our approaches produce generic summaries and in none of the cases, the keywords provided in the article have been taken into account in the summarization process. Some of summaries generated employing the COMPENDIUM$_{E-A}$ approach do not contain as many topics as the ones for COMPENDIUM$_E$. This occurs because in the former approach the resulting summaries contain sentences that may have been compressed, so in some of these cases, there is a loss of information, although minimal.

5.3 User Satisfaction Study

In the last evaluation, we aim at assessing the user satisfaction with respect to the generated summaries. For this purpose, we performed a qualitative evaluation and we asked 10 humans to evaluate our summaries[6] according to a 5-level Likert scale (1= Strongly disagree...5=Strongly Agree) for three questions concerning the appropriateness of the summaries. Specifically, these questions were:

Q1: The summary reflects the most important issues of the document.
Q2: The summary allows the reader to know what the article is about.
Q3: After reading the original abstract provided with the article, the alternative summary is also valid.

The percentage of summaries for each question in a scale 1 to 5 is shown Table 5. As it can be seen from the results obtained, our abstractive-oriented approach (COMPENDIUM$_{E-A}$) obtains better results than the pure extractive one (COMPENDIUM$_E$). Although the results concerning the information contained in the summaries generated with COMPENDIUM$_{E-A}$ were slightly lower than the extractive approach, taking into consideration their quality from a human point of view, the abstractive-oriented summaries are much better than the extractive ones. When we have a look at the different percentages of summaries that have been rated in one of each categories, we observe that there is a higher percentage of abstractive-oriented summaries that humans agree with, compared to the extractive summaries for the same rating (e.g. fourth row of the table - *Agree* -). Moreover, it is worth stressing upon the fact that, analogously, the percentage of summaries with lower ratings for strongly disagree and disagree also decrease when COMPENDIUM$_{E-A}$ is employed.

Furthermore, regarding the average results, COMPENDIUM$_{E-A}$ achieves at most 3.37/5 for Q2 and 3.1/5 for Q1 and Q3, whereas the maximum average value for COMPENDIUM$_E$ is 2.83/5 for Q2, the remaining questions obtaining values lower than 2.60/5. In light of the results obtained, it has been proved that the combination of extractive and abstractive techniques is more appropriate and leads to better summaries than pure extractive summaries.

[6] The humans were also provided with the original articles and their abstracts.

Table 5. User satisfaction results for the different text summarization approaches

%	TS Approach	Q1	Q2	Q3
1. Strongly disagree	COMPENDIUM$_E$	9.76	19.51	19.51
	COMPENDIUM$_{E-A}$	2.44	0	2.44
2. Disagree	COMPENDIUM$_E$	41.46	19.51	34.15
	COMPENDIUM$_{E-A}$	31.37	21.95	31.71
3. Neither agree nor disagree	COMPENDIUM$_E$	24.39	29.27	26.83
	COMPENDIUM$_{E-A}$	21.95	29.27	26.83
4. Agree	COMPENDIUM$_E$	21.95	21.95	7.32
	COMPENDIUM$_{E-A}$	41.46	39.02	34.15
5. Strongly agree	COMPENDIUM$_E$	2.44	9.76	12.20
	COMPENDIUM$_{E-A}$	2.44	9.76	4.88

6 Conclusion and Future Work

In this paper we presented COMPENDIUM, a text summarization system applied to the generation of abstracts of research papers. In particular, two generic text summarisation approaches were analysed: one based on a pure extractive summarizer (COMPENDIUM$_E$) and a novel approach (COMPENDIUM$_{E-A}$) which combined extractive and abstractive techniques, by incorporating an information compression and fusion stage once the most important content is identified. We carried out an evaluation based on three criteria: i) the information contained in the summaries; ii) the topics identified; and iii) the users' satisfaction. From the results obtained, we can conclude that COMPENDIUM is useful for producing summaries of research papers automatically. Although extractive and abstractive-oriented summaries perform similarly as far as the information and topics they contain, abstractive-oriented summaries are more appropriate from a human perspective. However, there are some issues that have to be tackled in the short-term. We plan to analyze other variants of the proposed approach for building abstracts, such as taking the source document as a starting point. Moreover, we are interested in studying other graph-based algorithms.

References

1. Belz, A.: Automatic Generation of Weather Forecast Texts Using Comprehensive Probabilistic Generation-space Models. Natural Language Engineering 14(4), 431–455 (2008)
2. Bouras, C., Tsogkas, V.: Noun retrieval effect on text summarization and delivery of personalized news articles to the user's desktop. Data Knowledge Engineering 69, 664–677 (2010)
3. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks ISDN Systems 30, 107–117 (1998)
4. Carenini, G., Cheung, J.C.K.: Extractive vs. NLG-based abstractive summarization of evaluative text: The effect of corpus controversiality. In: Proc. of the 5th International Natural Language Generation Conference, pp. 33–40 (2008)
5. Erkan, G., Radev, D.R.: Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research (JAIR) 22, 457–479 (2004)

6. Ferrández, Ó., Micol, D., Muñoz, R., Palomar, M.: A perspective-based approach for solving textual entailment recognition. In: Proc. of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp. 66–71 (2007)
7. Filippova, K.: Multi-sentence compression: Finding shortest paths in word graphs. In: Proc. of the 23rd International Conference on Computational Linguistics, pp. 322–330 (2010)
8. Givón, T.: A functional-typological introduction, vol. II. John Benjamins, Amsterdam (1990)
9. Kumar, M., Das, D., Agarwal, S., Rudnicky, A.: Non-textual Event Summarization by Applying Machine Learning to Template-based Language Generation. In: Proc. of the 2009 Workshop on Language Generation and Summarisation, pp. 67–71 (2009)
10. Lal, P., Rüger, S.: Extract-based summarization with simplification. In: Workshop on Text Summarization in Conjunction with the ACL (2002)
11. Lin, C.Y.: ROUGE: a package for automatic evaluation of summaries. In: Proc. of ACL Text Summarization Workshop, pp. 74–81 (2004)
12. Liu, M., Li, W., Wu, M., Lu, Q.: Extractive summarization based on event term clustering. In: Proc. of the 45th ACL, pp. 185–188 (2007)
13. Luhn, H.P.: The automatic creation of literature abstracts. In: Mani, I., Maybury, M. (eds.) Advances in Automatic Text Summarization, pp. 15–22. MIT Press, Cambridge (1958)
14. Mohammad, S., Dorr, B., Egan, M., Hassan, A., Muthukrishan, P., Qazvinian, V., Radev, D., Zajic, D.: Using citations to generate surveys of scientific paradigms. In: Proc. of the North American Chapter of the ACL, pp. 584–592 (2009)
15. Pollock, J.J., Zamora, A.: Automatic abstracting research at chemical abstracts. In: Mani, I., Maybury, M. (eds.) Advances in Automatic Text Summarization, pp. 43–49. MIT Press, Cambridge (1999)
16. Saggion, H.: A classification algorithm for predicting the structure of summaries. In: Proc. of the Workshop on Language Generation and Summarisation, pp. 31–38 (2009)
17. Saggion, H., Lapalme, G.: Selective analysis for automatic abstracting: Evaluating indicativeness and acceptability. In: Proceedings of Content-Based Multimedia Information Access, pp. 747–764 (2000)
18. Sauper, C., Barzilay, R.: Automatically generating wikipedia articles: A structure-aware approach. In: Proc. of the 47th Association of Computational Linguistics, pp. 208–216 (2009)
19. Shen, D., Yang, Q., Chen, Z.: Noise Reduction through Summarization for Web-page Classification. Information Processing and Management 43(6), 1735–1747 (2007)
20. Spärck Jones, K.: Automatic summarising: The state of the art. Information Processing & Management 43(6), 1449–1481 (2007)
21. Wong, K.F., Wu, M., Li, W.: Extractive summarization using supervised and semi-supervised learning. In: Proc. of the 22nd International Conference on Computational Linguistics, pp. 985–992 (2008)
22. Yu, J., Reiter, E., Hunter, J., Mellish, C.: Choosing the Content of Textual Summaries of Large Time-series Data Sets. Natural Language Engineering 13(1), 25–49 (2007)
23. Zhou, L., Ticrea, M., Hovy, E.: Multi-document biography summarization. In: Proc. of the International Conference on Empirical Methods in NLP, pp. 434–441 (2004)