

Comparing Commercial Tools and State-of-the-Art Methods for Generating Text Summaries

René Arnulfo García-Hernández¹, Yulia Ledeneva², Griselda Matías Mendoza³, Ángel Hernández Domínguez⁴,
Jorge Chavez⁵, Alexander Gelbukh⁶, José Luis Tapia Fabela⁷

^{1,2,3,7}Software Engineering Department
Autonomous University of the State of Mexico
Santiago Tianguistenco, México

¹renearnulfo@hotmail.com, ²yledeneva@yahoo.com,
³gris_9123@hotmail.com, ⁷joseluis.fabela@gmail.com

⁴Computer Science Department
Juarez Autonomous University of Tabasco
ahedom@gmail.com

⁵Computational Systems Engineering
Iguala Institute of Technology
Iguala, Guerrero
gara_men@hotmail.com

⁶Natural Language Laboratory
CIC, National Polytechnic Institute
DF, Mexico
gelbukh@gelbukh.com

Abstract—*Nowadays, there are commercial tools that allow automatic generation of text summaries. However, it is not known the quality of the generated summaries and the method that it is used for the generation of the summaries using these commercial tools. This paper provides a study about the commercial tools such as Copernic Summarizer, Microsoft Office Word Summarizer 2003 and Microsoft Office Word Summarizer 2007, with the objective to detect which of them gives the summaries more similar to those made by a human. Furthermore, the comparison between commercial tools and state-of-the-art methods is realized. The experiments were carried out using DUC-2002 standard collection which contains 567 news in English.*

Keywords- *Automatic Text Summarization; Copernic Summarizer; Microsoft Office Word Summarizer; Open Office Summarizer.*

I. INTRODUCTION

The amount of documents in a digital format is increased by 30% annually [1]. A lot of relevant and interesting documents are not read by the user due to the large amount of information. That is why the necessity of tools that automatically generate summaries appears. These tools are not just for professionals who need to find the information in a short time but also for large searching companies such as Google, Yahoo!, AltaVista, and others, which could obtain a lot of benefits in its results if they use automatic generated summaries. After that, the user only will require the interesting documents, reducing the flow of information.

According to Lloret [2], a summary is a text that it is generated from one or more documents, which contains the most significant information and also it is not larger than half of the original documents. However, we consider that a good summary also must have coherence. The size of the

summary depends on the user needs and from the size of the original document. Therefore, the size of the summary must be the most flexible parameter. Also, we consider that to achieve a good summary, the tool should work mostly with text content and to a less degree with the document format. It is also desirable that the tools can work independently of domain and language of a given document, indeed it is not necessary that the original document is grammatically well written.

Another important feature is the format in which the summary is presented to the user, the most important key phrases or sentences can be highlighted within the summary or original document, without deleting the context in which such phrases occur. Also, a good tool to generate summaries should have a friendly interface.

Currently, there are commercial tools that automatically generate summaries compressing main ideas of a document. The first objective of the paper is to know which of the commercial tools produces summaries most similar to a human. The second objective is to compare the commercial tools to the state-of-the-art methods.

The methods for generation of automatic summaries are generally classified into abstractive and extractive summaries. Humans create abstractive summaries, in this case is necessary first understand the content of the document and then it is generated summary, but the new text could have words or sentences that does not exist in the original text. In contrast, extractive summaries are generated using the selection of important units of the text, such as key phrases, sentences or paragraphs of the original document. The generation of extractive summaries does not require the understanding of the text. In this paper, we generate extractive summaries using commercial tools and state-of-the-art methods.

II. DESCRIPTION OF COMMERCIAL TOOLS

Currently, there are several commercial tools that help us in generating automatic summaries, among the most popular are the following tools.

Copernic Summarizer. In this paper, we use version 2.1 which was installed on the Microsoft Windows operating system. This software was developed exclusively for the generation of automatic summaries. It is a flexible and suitable tool. It offers different options to make summaries: 5%, 10%, 25% and 50% of words of the original document; 100, 250 and 1000 words.

According to [3], Copernic Summarizer it uses the following methods:

- a. *Statistical model (S-Model)*. This model is used in order to find the vocabulary of the text.
- b. *Knowledge Intensive Processes (K-Process)*. Consider the way in which human make summary texts by taking into account the following steps:
- c. *Language detection*. It detects the language (English, German, French or Spanish) of the document for applying specific processes.
- d. *The limits of sentence recognition*.
- e. *Concept extraction*. Copernic Summarizer uses machine learning techniques to extract keywords.
- f. *Document Segmentation*. Copernic Summarizer organizes the information that it can be divided into larger related segments.
- g. *Sentence Selection*. Sentences are selected according to their importance (weight) discarding those that decrease readability and coherence.

Microsoft Office Word Summarizer. This tool can be found in versions of Microsoft Office Word 2003 and Microsoft Office Word 2007. This tool can generate summaries of 10 or 20 sentences, 100 or 500 words (or less) or in percentages of 10%, 25%, 50% and 75% of words of the original document. If some of the percentages are not appropriate, the user can change as needed. This tool offers various ways of visualizing summaries. One is highlighting the color of important sentences in the original document.

The summary created by this tool is the result of an analysis of key words; the selection of these is done by assigning a score to each word. The most frequent words in the document will have highest scores which will be considered as important. The sentences containing these words will be included in the summary.

OpenOffice Summarizer¹. The software of OpenOffice is available free. The tool for generating summaries only works when the text has paragraphs with a predetermined format. This format is required manually marking. This tool only takes the header and the first lines of each paragraph as a summary.

StarOffice Summarizer. This tool is very similar to OpenOffice with one difference that is a commercial version of Sun Microsystems. This tool has been discarded from the comparison because it does not use the content of the document, and only uses the format and structure of a document to generate the summary.

Finally, in this paper, the tools considered to analyze and compare are: Copernic Summarizer, Microsoft Office Word Summarizer 2003 and Microsoft Office Word Summarizer 2007.

III. DESCRIPTION OF THE STATE-OF-THE-ART METHODS

The next state-of-the-art methods are not commercial but let to obtain good results.

TextRank [4]. This method consists of weighted graphs. Mihalcea [4] constructs the graph to represent the text, so the nodes are words (or other text entities) interconnected by vertices with meaningful relationships. For the task of extracting sentences, the goal is to qualify whole sentences and order them from most to least importance. Therefore, vertices are added to the graph for each sentence in the text. To make the connections between sentences, a relation of similarity is defined, where the relationship between two sentences can be seen as a process of "recommendation".

A sentence that points to a certain concept in the text gives the reader a "recommendation" to refer to other sentences in the text that point to the same concepts, and then link a can be established between any two sentences that share a common content. Since this method can determine the importance of each of the sentences, it was used to generate documents summaries.

Baseline. This heuristic ensures that the most important information of a document is in the early sections [6] for which is taken the first n sentences of the document to form the summary. This simple heuristic has been shown to generate very good summaries in the field of news documents.

Maximal Frequent Sequences (SFMs) [7, 8]. This paper presents a method to generate extractive summaries from a single document based on statistics, which is independent of the domain and language. Ledeneva *et al.* [7, 8] experimentally shows that the words which are parts of bigrams (2-word sequences) which are repeated more than once in the text are good terms to describe the content of that text, so also called the maximal frequent sequences (sequences of words that are repeated a number of times and also are not contained in other frequent sequences). This work also shows that the frequency of the term as ranking of terms gives good results (while only count the occurrences of a term in repeated bigrams). Ledeneva *et al.* applies a method which has 4 stages for generating the summary. These steps are term selection, term weighting, sentence weighting and sentence selection. In term selection step, SFMs, repetitive bigrams (must appear at least twice in the text), and unigrams (simply words) are extracted. In term

¹ <http://www.openoffice.org/>

weighting step, the frequency of the term is used, which is the number of times the term occurs in the text. In sentence weighting, only the weight of all the terms contained in that sentence is calculated. Finally, sentence selection that composes the summary is performed by two criteria. First, the k sentences with bigger weight are selected. Second, k sentences with bigger weight are selected and completed with the first sentences (similar to baseline heuristic) that appear in the document (combined version). The best result is obtained with combined version when k=1, reaching 47% of similarity with the summaries made by a human.

Clustering sentences with SFMs [9]. In the previous method, sentences which have bigger weight are selected for composing the summary. However, if the sentences that are chosen in that order may include very similar sentences and do not provide new information in the summary. The work [9] uses a clustering algorithm based on SFMs to make sets of sentences, from which it selects the most representative sentence from each group to compose the summary.

Baseline-Random [8]. This heuristic far from seeking to obtain best summaries attempts to determine the quality of the summaries when only set of sentences as summary are taken at random. The idea is to determine how significant the results can be achieved.

IV. EXPERIMENTATION

For comparing the abovementioned applications the collection Document Understanding Conference (DUC) 2002 [5] was used, which was created by the National Institute of Standards and Technology (NIST) for use of researchers in automatic text summarization. This collection has 567 news in English of various lengths, about technology, food, politics, finance, etc. For each document in the collection was created two summaries by two human experts with a minimum length of 100 words.

ROUGE2 1.5.5, proposed by Lin [10, 11], is the tool used for the automatic comparison of summaries with the option of n-gram where n = 1, which has the ability to measure similarity and determine the quality of an automatic summary compared to the one created by a human. Using this tool, it is possible to compare the summaries generated by commercial tools.

4.1 Configuration of Comparison

Commercial tools were evaluated in the operating system Windows XP Professional Service Pack 2 (SP2). Each file was manually selected and applied to generated summary of 100-words. In the case of Microsoft Office Word 2003 and 2007 is not possible to use the option of 100-word summaries because it generates summaries with until 100

words but sometimes produces less than 100-word, putting in disadvantage this tool. For such problem was necessary to calculate the adequate percentage to produce a summary with minimum 100 words, calculated as follows: $(\text{Number of desirable words} / \text{Number of total words}) \times 100$.

4.2 Evaluation of Commercial Tools

The evaluation results of commercial tools are realized using ROUGE.

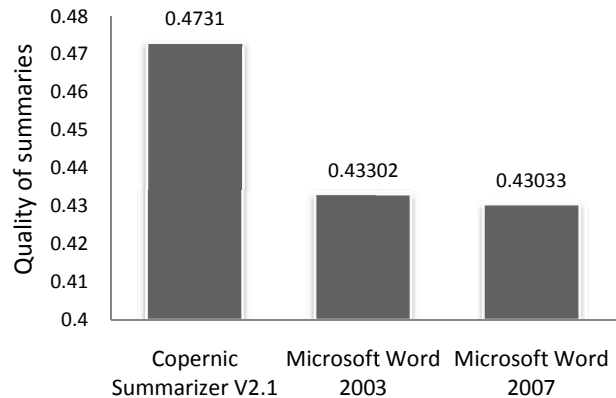


Figure 1. The evaluation results of commercial tools using the operating system Windows XP Professional SP2.

According with Figure 1 Copernic Summarizer outperforms the two versions of Microsoft OfficeWord, although Microsoft Office Word 2003 was slightly better than its 2007 version.

However, during the experimentation an inconsistency in Microsoft Office Word was observed because the generated summaries change depending on the operating system. In order to verify this fact the same setup package of Microsoft Office Word 2003 was installed on Windows XP Professional SP2, which also was done with Microsoft Office Word 2007. The resulting abstracts were assessed with the same version of ROUGE and obtained the results shown in Figure 2.

In Figure 2, we can observe the slight difference between the tools of auto summary of Microsoft Office Word. In contrast, Copernic Summarizer tool show the same results in both operating systems. We can conclude that Copernic Summarizer is independent of the operating system).

Also, it is important to mention that Microsoft Office Word 2007 with Windows XP Professional SP2 obtained the worst result, among the versions of Microsoft Office Word. Nevertheless, the best result, among the versions of Microsoft Office Word, was obtained with Microsoft Word 2003 with Windows Vista Home Premium SP1. This shows the dependence of these tools with respect to the operating system they are using.

² <http://belobog.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=W04-1013>.

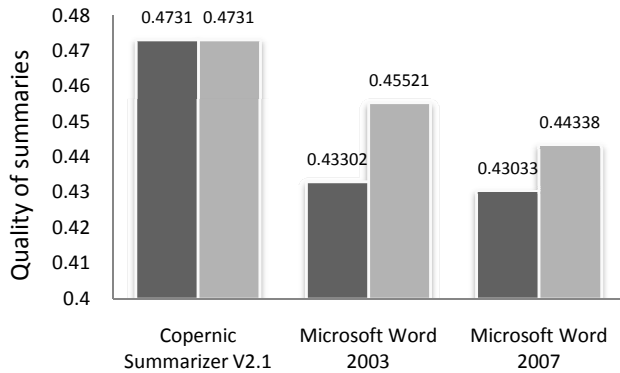


Figure 2. Results of the tools evaluated with the operating systems Windows XP Professional SP2 and Windows Vista Home Premium SP1.

4.3 Results of Commercial Tools and State-of-the-Art Methods

In order to see the quality of previous results compared to those obtained with state of the art methods, in Figure 3 are shown the best results obtained by commercial tools and the results reported by the state-of-the-art methods

Figure 3 shows clearly that the results of Copernic Summarizer is the highest score, just below the proposed method SFMs (1Best+First) and below of Sentence

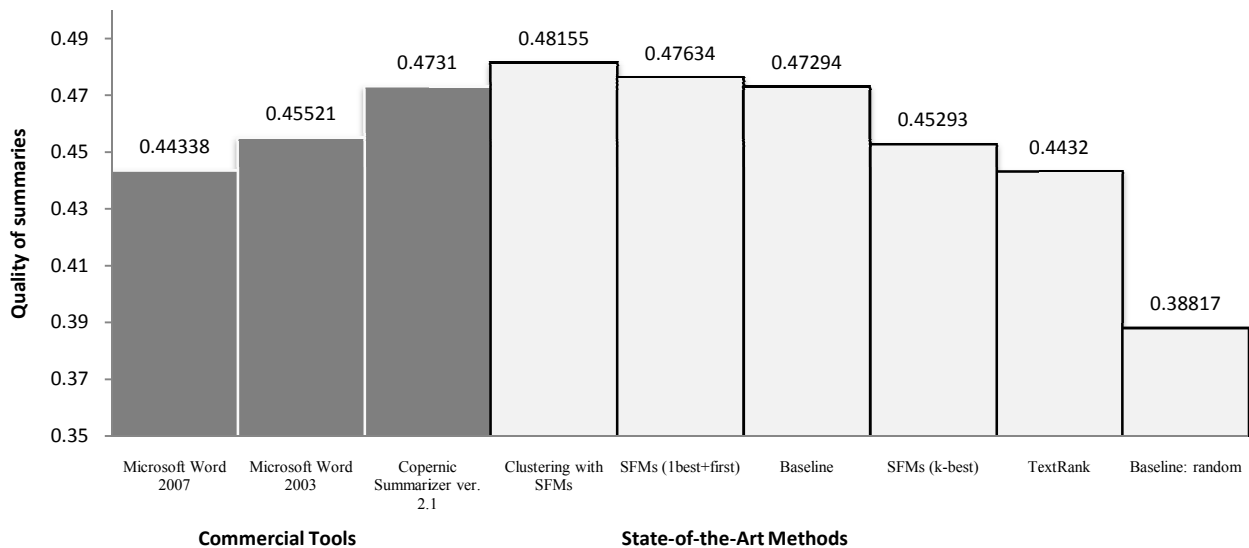


Figure 3. Results generated by ROUGE 1.5.5 for the collection of summaries obtained by commercial tools and the state-of-the-art methods.

Clustering with MFSs, which confirms this software is one of the best of its kind. Also, it is possible to observe that Copernic Summarizer was the only one outperformed the Baseline of first sentences. It is possible to observe that Microsoft Office Word 2003 and Microsoft Office Word 2007 are in the 5th and 7th place respectively.

V. DISCUSSION AND CONCLUSIONS

In this paper, the evaluation of the automatic summaries generated by commercial tools (Copernic Summarizer, Microsoft Office Word Summarizer 2003 and Microsoft Office Word Summarizer 2007) was realized. The summaries were evaluated using the ROUGE system. The following conclusions can be given based on obtained results:

- Microsoft Office Word is inconsistent because it generates different summaries depending on the operating system.
- The results obtained with Microsoft Office Word 2003 and Microsoft Office Word 2007 with Windows Vista were better than with Windows XP.
- Microsoft Office Word 2003 gets a better result than Microsoft Office Word 2007 with Windows Vista operating system.
- Copernic Summarizer gets the best results of commercial tools.

VI. FUTURE WORK

There are other online tools capable of automatically summarize like Summarizer Online Tool³, Pertinence Summarizer⁴ and Shvoong Summarizer⁵. It would be interesting to know how good these tools are.

The experiments in this article utilized a news document collection which have a short extension, but if the collection is change by longer documents such as scientific articles or thesis. Does the baseline heuristic of first sentences would still be as good?

ACKNOWLEDGMENTS

Work done under partial support of Mexican Government (CONACyT, SNI, PROMEP, SIP). The authors thank Autonomous University of Guerrero, Juarez Autonomous University of Tabasco and Autonomous University of the State of Mexico for their assistance.

REFERENCES

- [1] Lyman, Peter and Hal R. Varian, "How Much Information", 2003. Retrieved from <http://www.sims.berkeley.edu/how-much-info-2003>
- [2] Lloret E., O. Ferrández, R. Muñoz y M. Palomar. 2008. Integración del Reconocimiento de la implicación textual en tareas automáticas de resúmenes de textos. *Procesamiento del Lenguaje Natural*, No. 41, pp. 183-190.
- [3] Copernic Summarizer, Technologies WhitePaper, 2003. <http://www.copernic.com/data/pdf/summarization-whitepaper-eng.pdf>
- [4] Rada Mihalcea; Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization; Department of Computer Science; University of North Texas; Texas; EUA; 2004.
- [5] DUC. Document Understanding Conference, www-nlpir.nist.gov/projects/duc.
- [6] Yulia Ledeneva, Alexander Gelbukh, René A. García-Hernández. Terms Derived from Frequent Sequences for Extractive Text Summarization, 9th Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2008), Lecture Notes in Computer Science, Springer-Verlag, Vol. 4919. pp. 593-604.
- [7] Yulia Ledeneva. Automatic Language-Independent Detection of Multiword Descriptions for Text Summarization, National Polytechnic Institute, PhD. Thesis, Mexico. 2009.
- [8] René Amulfo García-Hernández, Romyna Montiel, Yulia Ledeneva, Eréndira Rendón, Alexander Gelbukh, Rafael Cruz. Text Summarization by Sentence Extraction Using Unsupervised Learning, 7th Mexican International Conference on Artificial Intelligence (MICAIO8), Lecture Notes in Artificial Intelligence, Springer-Verlag, Vol. 5317 pp. 133-143, 2008.
- [9] Lin, C., y E. Hovy. 2003. Automatic evaluation of summaries using N-gram co-occurrence statistics. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*. Vol. 1, pp. 71-78.
- [10] Lin, C. 2004. ROUGE: A package for automatic evaluation of summaries. *Proceedings of the Association for Computational Linguistics 2004 Workshop*, pp. 74-81. Spain.

³<http://www.tools4noobs.com/summarize/>

⁴http://www.pertinence.net/index_en.html

⁵<http://www.shvoong.com/summarizer/>