Hybridization of Social Spiders and Extractions Techniques for Automatic Text Summaries

Mohamed Amine Boudia, Department of Computer Science, Laboratory Knowledge Management and Complex Data (GeCoDe Lab), Dr. Moulay Tahar University Saida, Saida, Algeria

Reda Mohamed Hamou, Department of Computer Science, Laboratory Knowledge Management and Complex Data (GeCoDe Lab), Dr. Moulay Tahar University Saida, Saida, Algeria

Abdelmalek Amine, Department of Computer Science, Laboratory Knowledge Management and Complex Data (GeCoDe Lab), Dr. Moulay Tahar University Saida, Saida, Algeria

Mohamed Elhadi Rahmani, Department of Computer Science, Laboratory Knowledge Management and Complex Data (GeCoDe Lab), Dr. Moulay Tahar University Saida, Saida, Algeria

Amine Rahmani, Department of Computer Science, Laboratory Knowledge Management and Complex Data (GeCoDe Lab), Dr. Moulay Tahar University Saida, Saida, Algeria

ABSTRACT

The authors propose a new multilayer approach for automatic text summaries. In the first layer, they use two techniques of extraction, one after the other: scoring of phrases, and similarity that aims to eliminate redundant phrases without losing the theme of the text. While the second layer aims to optimize the results of the previous layer by a meta-heuristic based on social spiders. Its objective function of the optimization is to maximize the sum of similarity between phrases of the candidate summary in order to keep the theme of the text, minimize the sum of scores in order to increase the summarization rate; this optimization also will give a candidate's summary where the order of the phrases changes compared to the original text. The third and final layer concerned in choosing a best summary from all candidates summaries generated by optimization layer, we opted for the technique of voting with a simple majority.

Keywords: Automatic Summary Extraction, Data Mining, Optimization, Scoring, Similarity, Social Spider

DOI: 10.4018/IJCINI.2015070104
1. INTRODUCTION AND PROBLEMATIC

Every day, the mass of electronic textual information is increasing dramatically making it more and more difficult access to relevant information without the use of special tools. Additionally, access to the content of the texts by rapid and effective ways has become a necessity. It is becoming increasingly difficult to access relevant information without using specific tools, i.e. access to the content of texts by rapid and effective means has become a necessary task.

Currently, one of the major problems for computer scientists is access to the content of information, access itself or in other words the software and hardware infrastructure are no longer an obstacle, and the major problem is the exponential increase in the amount of textual information electronically. This requires the use of more specific tools i.e. access to the content of texts by rapid and effective means has become a necessary task.

A summary of a text is an effective way to represent the contents of the texts and allow quick access to their semantic content. The purpose of a summarization is to produce an abridged text covering most of the content from the source text. As a matter of fact, a summary of the text has rewritten the text in smaller way. Under constraint kept the semantics of a document that is minimized entropy semantics. The purpose of this operation is to help the reader to identify interesting information for him/her without reading the entire document.

“We cannot imagine our daily life, one day without summaries,” (Minel, J., 2004). Newspaper headlines, the first paragraph of a newspaper article, newsletters, weather, tables of results of sports competitions and library are all summarized. Even in the research, the authors of scientific articles must accompany their scientific articles by a summary written by themselves.

Automatic summaries can be used to reduce the search time to find the relevant documents or to reduce the treatment of long texts by identifying the key information.

To make an automatic summary, the current literature presents three approaches:

- Summarisation by extraction;
- Summarisation by understanding;
- Summarisation by classification.

Our current work uses automatic summarization by extraction as it is a simple method to implement that gives good results; Noting that they are three techniques in this approach: Scoring, similarity and prototype. In the previous work, produce the automatic summary by extraction approach consists to use only one technique at a time (Scoring of phrase, Similarity between phrase or prototype) and respects the order of the phrases in the original document.

The first part contribution of our work is to use two methods of summarization at the same time on the quality of summary. The second part and the most important of this work is the proposition to use a bio-inspired method based on the social spiders to the automatic summary. We aim to evaluate the impact of two part of contribution on the quality of summary.

For that we will proceed first in second section by a state of art to get an idea of anterior work: their advantage and lacks, In section three, then we will explain in details our approach proposes, as any research we will experiment our approach in the section four we will talk the experiment environment, then we will interpret the results to come out with a final conclusion.
Modified Gabor Wavelets for Image Decomposition and Perfect Reconstruction
www.igi-global.com/article/modified-gabor-wavelets-image-decomposition/37572?camid=4v1a

Semantically Linking Information Resources for Web-Based Sharing
Junsheng Zhang, Yingfan Gao, Yanqing He, Hongjiao Xu, Chongde Shi and Peng Qu (2013). *International Journal of Cognitive Informatics and Natural Intelligence* (pp. 65-79).
www.igi-global.com/article/semantically-linking-information-resources-for-web-based-sharing/101818?camid=4v1a