

A Visualization System for Multi-document Summarization

Jader B. P. Lima¹, Tao Li², Thiago A. S. Pardo¹

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo
Caixa Postal 668 – 13.560-970 – São Carlos – SP – Brazil

²School of Computer Science – Florida International University
Miami – FL – U.S.A.

jbplima@usp.br, taoli@cs.fiu.edu, taspardo@icmc.usp.br

1. Introduction

Nowadays, computational applications that are capable of processing the large amount of available information (mainly on the web) have become increasingly necessary. Consequently, the task of automatic multi-document summarization (MDS), which aims at producing a concise summary from a group of source texts on the same topic, receives a lot of research attention (Mani, 2001). For an overview of MDS and its current state-of-the-art techniques, please see Nenkova and McKeown (2011).

In MDS, after automatically producing a summary, the way the summary is presented and the available exploration possibilities are very important to the end user. Generally, it is not sufficient to simply present the summary in a plain text format. A good MDS should present more information about the data by both simple and advanced visualization features and allow users to explore the textual content in rich and varied ways. Some useful visualization features include but are not limited to the following: the presentation of the source texts from which the sentences in the summary were extracted, the display of illustrative images about the content of the summary, the definition of basic summary characteristics (as compression rate in relation to the size of the source texts), and the display of keywords for the summary. Moreover, these features may be tailored for different users, as different users may have different visualization preferences and information needs. Mani and Maybury (1999) suggested that users may be divided into two groups. The first group needs a broader view quantifying the data rather than detailing it. These users usually require the statistical and qualification information about the data, for example, the most frequent words in the source texts or the number of times each source text was read in the web. The second group of users usually requires more detailed information about the data. In this case, visualization techniques that allow the expansion of the information nugget (for example, the publication date or the full source texts) to be displayed to end users are needed.

In this short paper, we first briefly report our study of visualization features for MDS systems. We then propose a visualization system that might be used by any MDS system in order to improve the usage and interaction experience.

2. Visualization Features

We have studied many MDS systems with an emphasis on their visualization features, focusing on systems that are well-known in the area and on systems whose proposal was to allow for richer textual exploration, namely: Columbia Newsblaster (McKeown

et al., 2002), News in Essence (Radev et al., 2004), Ineats (Leuski et al., 2000), RSumm (Ribaldo et al., 2012) and SumView (Wang et al., 2012).

The studied features have been divided into three groups based on their main motivation. The first group contains features related to the basic information to be displayed to the user; the second contains more advanced features that are concerned with more refined information to be displayed; the third contains features regarding summarization parameters that may be input by the user. Table 1 summarizes the visualization features in different MDS systems. One may observe that some features are common to most of the systems (e.g., the “reference to the source texts” feature), while some of them are rarer (e.g., the “hot topics summarized by users”).

Table 1. Visualization features and MDS systems

	Columbia Newsblaster	News in Essence	Ineats	RSumm	Sum View
Group 1: information visualization					
1.1: Reference to the source texts	Yes	Yes	Yes	Yes	No
1.2: Highlight of a selected sentence in its source text	Yes	Yes	No	No	No
1.3: Adequacy of textual information (e.g., HTML cleaning)	Yes	No	Yes	Yes	Yes
1.4: Display of keywords in the summary	Yes	No	Yes	No	Yes
1.5: Display of size of the source texts	Yes	No	Yes	No	Yes
1.6: Display of publication date of the source texts	Yes	No	No	No	No
1.7: Display of local of publication of the source texts	No	No	Yes	No	No
Group 2: more advanced visualization options					
2.1: Hot topics summarized by users	No	Yes	No	No	No
2.2: Exhibition of full source texts	No	Yes	No	Yes	No
2.3: Display of illustrative images for the summary	Yes	No	No	No	Yes
2.4: Coloring of sentences according to their topics or sources	No	No	Yes	No	No
2.5: Grouping of source texts by topic	Yes	No	No	No	Yes
Group 3: Summary customization					
3.1: Parameterization of the size of the summary	No	Yes	Yes	Yes	No
3.2: Choice of relevant topics to be summarized	No	No	Yes	No	Yes

3. Visualization Interface

Figure 1 presents the prototype of the interface (named ViSUM) that is currently under development. It is important to notice that the proposed interface might present the results generated by any summarization system. In the figure, the prototype is displayed using the RSumm system.

In order to allow for a systematic development of the interface, we started by including the features that are used by most MDS systems and are moving towards the inclusion of all of them (while the system usability is not affected). In the interface, the summary is on the left panel and the source texts are on the right. It is possible to navigate among the content of summary and the content of the source texts, because the corresponding sentences in both sides are highlighted when the user positions the mouse pointer over them (considering extractive summaries). Other features are the display of illustrative images (retrieved from a search engine API), the possibility of parameterization of the summary and the display of extra information as the most frequent words. In the near future, after it is fully developed, the system will be evaluated using traditional usability measures.

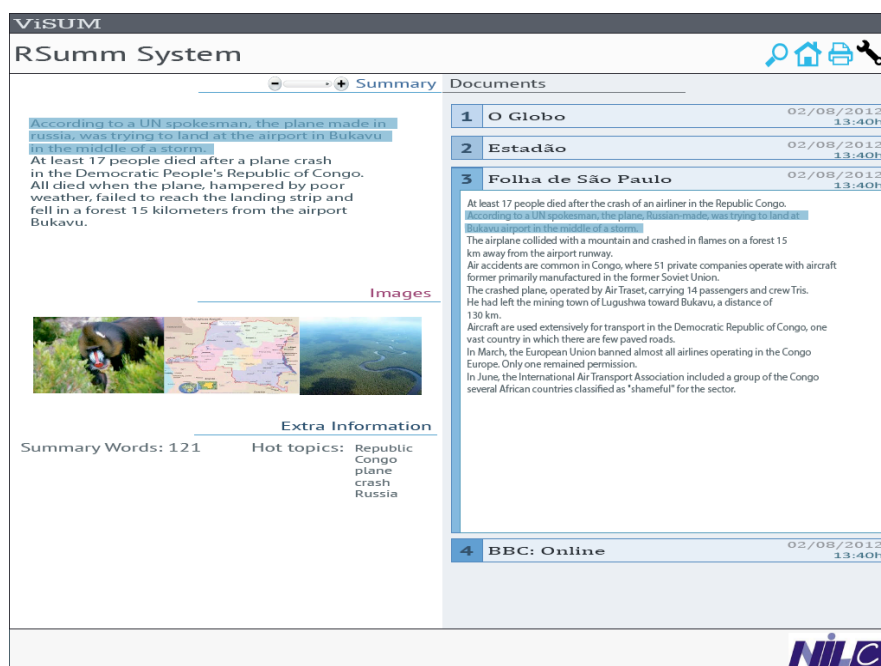


Figure 1. Prototype interface under development

Acknowledgements

The authors are grateful to FAPESP (2011/23200-0 and 2012/16220-7) and US NSF (HRD-0833093) for supporting this work.

References

- Leuski, A; Lin, C.Y.; Hovy, E.H. (2003). iNeATS: Interactive Multidocument Summarization. In the *Proceedings of 41st Annual Meeting on Association for Computational Linguistics*, pp. 125-128.
- Mani, I. (2001). *Automatic Summarization*. John Benjamins Publishing Co., Amsterdam.
- Mani, I and Maybury, M.T. (1999). *Advances in automatic text summarization*. MIT Press, Cambridge, MA.
- McKeown, K.R.; Barzilay, R.; Evans, D.K.; Hatzivassiloglou, V.; Klavans, J.L.; Nekova, A.; Sable, C.; Schiffman, B.; Sigelman, S. (2002). Tracking and summarizing news on a daily basis with Columbia's Newsblaster. In the *Proceedings of the second international conference on Human Language Technology Research*, pp. 280-285.
- Nenkova, A. and McKeown, K. (2011). Automatic Summarization. *Foundations and Trends in Information Retrieval*, Vol. 5, N. 2 & 3.
- Radev, D.; Jing, H.; Syvs, M.; Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing and Management*, Vol. 40, pp. 919-938.
- Ribaldo, R.; Akabane, A.T.; Rino, L.H.M.; Pardo, T.A.S. (2012). Graph-based Methods for Multi-document Summarization: Exploring Relationship Maps, Complex Networks and Discourse Information. In the *Proceedings of the 10th International Conference on Computational Processing of Portuguese (LNAI 7243)*, pp. 260-271.
- Wang, D.; Li, T.; Zhu, S. (2013). SumView: A web-based engine for summarizing product reviews and customer opinions. *Expert Systems and Applications*, Vol. 40, N. 1, pp. 27-33.