Personalized Text Summarization

Róbert MÓRO*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
xmoror@stuba.sk

Information overload is one of the most serious problems of the present-day web. Automatic text summarization aims to address this problem by extracting the most important information from the document, which can help readers (users) to decide, whether it is relevant for them and they should read the whole text or not.

Conventional (generic) summarization methods summarize only the basic content of the document and don't take into account the differences in users' interests, goals or knowledge. Personalized summarization, on the other hand, uses this additional information about users' characteristics to produce summaries more suitable for a particular user's needs.

We propose a method of personalized summarization based on a method of *latent semantic analysis (LSA)* [1][2]. We have chosen LSA as a basis for our approach, because of its ability to provide better results compared to the other summarization methods. We have identified a construction of a terms-sentences matrix representing the document as a step suitable for summarization personalization. In this step terms extracted from the document are assigned their respective weights. Our proposed weighting scheme extends conventional weighting scheme based on tf-idf method by linear combination of multiple raters, which positively or negatively affect the weight of each term (see Figure 1).

We have proposed a set of raters which can be divided into two groups:

- *generic raters* take into account basic content of the document and some additional information to adapt summarization regardless of a particular user;
- *personalized raters* consider information about the specific user and her characteristics.

Our approach is language and domain independent; however, we focus on the domain of learning and the knowledge revision scenario. For this purpose we have designed the specific raters that take into account terms relevant for the domain or the level of knowledge of an individual user; we have also proposed a personalized method for selecting documents for revision. Because annotations (e.g. highlights) can indicate

* Supervisor: Mária Bieliková, Institute of Informatics and Software Engineering

Spring 2012 PeWe Workshop, April 1, 2012, pp. 73–74.

user's interest in the specific parts of the document [3], we use them as another source of personalization.

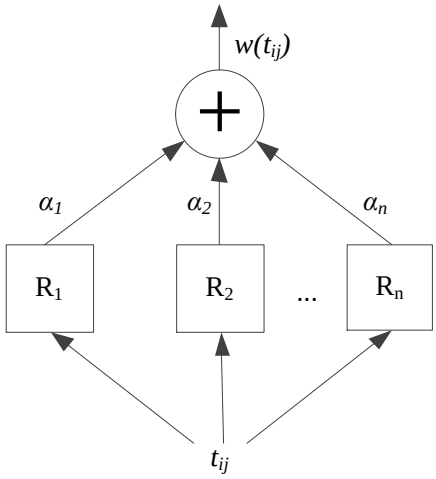

Figure 1. Term weighting by a combination of raters, where t_{ij} is a term, R_i is a rater with its linear coefficient α_i and $w(t_{ij})$ is a weight assigned to the term t_{ij}.

We have experimented with our proposed method. Our dataset has consisted of educational materials from the *Functional and Logic Programming (FLP)* course in the learning system ALEF. We have focused on evaluation and comparison of the two variants of summaries – *generic summarization* and *summarization considering the domain-relevant terms*. We have asked the FLP course students to evaluate quality of the generated summaries.

Our experimental results suggest that using the domain-relevant terms in the process of summarization leads to selecting representative sentences capable of summarizing the document for revision. We plan to experiment also with taking the users' annotations into consideration during the summarization process.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0971/11.

References

[1] Gong, X., Liu X.: Generic text summarization using relevance measure and latent semantic analysis. In: *SIGIR'01: Proc. of the 24th Int. ACM SIGIR Conf. on Research and Development in Inf. Retrieval*, ACM Press, (2001), pp. 19-25.

[2] Steinberger, J., Ježek, K.: Text summarization and singular value decomposition. In: *ADVIS'04: Proc. of Advances in Information Systems*, LNCS 3261. Springer, Berlin, (2005), pp. 245-254.

[3] Zhang, H., Ma, Z.C., Cai, Q.: A study for documents summarization based on personal annotation. In: *Proc. of the HLT-NAACL Workshop on Text summarization*, Association for Computational Linguistics, (2003), pp. 41–48.