

Personalized Text Summarization

Róbert MÓRO*

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

One of the most serious problems of present-day Web is information overload. Since we can find almost everything on the Web, it has become very problematic to find what we actually want or need – to find relevant information. Also, the term “relevant information” is subjective, because users differ in their interests, goals and knowledge. Automatic text summarization aims to address the information overload problem. The idea is to extract the most important information from documents, which can help users to decide, whether it is relevant for them and they should read it or not.

The problem with classical (generic) automatic text summarization methods is that they do not consider different user goals, interests or knowledge. Our idea is to personalize text summarization since much information about user interests can be inferred from their browsing (or in this case reading) behavior [1]. If a user reads a document about a particular topic, it serves as implicit feedback that he or she is interested in the topic [2]. Also, with the arrival of Web 2.0, users are no longer passive consumers of web content, but they can create content and add metadata, such as annotations or tags. These can be used as another important source for personalization.

Annotation of documents is a technique widely used by people especially when reading printed documents. They highlight or underline important parts of text, add explanations or different formulations or even references to other documents. This way, annotations can indicate reader (or user’s in the context of the web) interest in that particular part of the document [3]. We can take into account not only user annotations but also those of similar users, including the users’ collaboration into the process of text summarization.

Tags, as a special type of annotations, can also be considered. They are usually generalized descriptions of the topics contained in a document and directly reflect the users’ vocabulary and their understanding of the document.

We propose a personalized summarization method for web documents. Figure 1 shows the summarization process, which is common for all extractive summarization methods. The main difference, by which various methods can be distinguished, is usually in the keyword scoring and sentence selection. Also, additional preprocessing

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

of documents and/or post-processing of produced summaries can be employed to improve readability and text cohesion.

Our goal is to explore what additional information inferred from user provided metadata can be used (and how) to give higher scores to keywords reflecting user interests, goals or knowledge thus resulting in personalized summarization; and also to produce summaries adapted to actual user needs or context (e.g., reflected in summary length). We plan to evaluate our proposed method in the domain of e-learning in ALEF (Adaptive Learning Framework); however, the method itself will be designed and implemented as domain-independent using an enhanced proxy server [1].

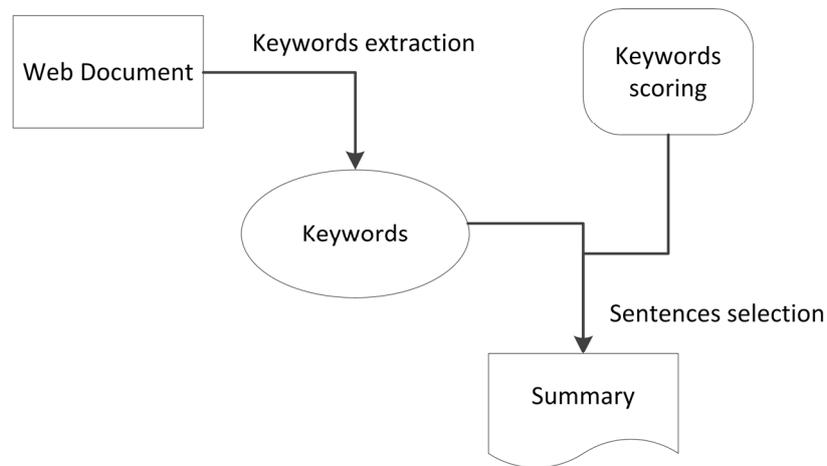


Figure 1. Overview of web document summarization.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

- [1] Barla, M., Bielikova, M.: Ordinary Web Pages as a Source for Metadata Acquisition for Open Corpus User Modeling. In *Proc. of IADIS WWW/Internet 2010*. IADIS Press, pp. 227-233, 2010.
- [2] Campana, M., Tombros, A.: Incremental Personalized Summarization with Novelty Detection. In *Proc. of the 8th Int. Conf. on Flexible Query Answering Systems*, Springer-Verlag Berlin, Heidelberg, pp. 641-652, 2009.
- [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 Morristown, NJ, USA, pp. 41-48, 2003.