

# Personalized Text Summarization Based on Important Terms Identification

Róbert Móro, Mária Bielíková

Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technologies,  
Slovak University of Technology, Ilkoviova 3, 842 47 Bratislava, Slovakia

Email: robert.moro, maria.bielikova@stuba.sk

**Abstract**—Automatic text summarization aims to address the information overload problem by extracting the most important information from a document, which can help a reader to decide whether it is relevant or not. In this paper we propose a method of personalized text summarization which improves the conventional automatic text summarization methods by taking into account the differences in readers characteristics. We use annotations added by readers as one of the sources of personalization. We have experimentally evaluated the proposed method in the domain of learning, obtaining better summaries capable of extracting important concepts explained in the document when considering the relevant domain terms in the process of summarization.

**Keywords**—automatic text summarization; personalization; annotations; relevant domain terms

## I. INTRODUCTION

Information overload is one of the most serious problems of the present-day web. There are various approaches addressing this problem; we are interested mainly in two: automatic text summarization and personalization.

Automatic text summarization aims to extract the most important information from a document, which can help readers (users) to decide whether it is relevant for them and they should read the whole text or not. However, the classical (generic) summarization methods summarize the content of a document without considering the differences in users, their needs or characteristics, i.e. their interests, goals or knowledge. On the other hand, personalization aims to adapt the content presented to an individual user or the way she accesses the content based on her characteristics.

In this paper we propose a method of personalized summarization which extracts from the document information that we assume to be the most important or interesting for a particular user. Because annotations (e.g. highlights) can indicate a users interest in the specific parts of the document [1], we use them as one of the sources of personalization. Our proposed method is sufficiently general to be used independently of the chosen domain; however we focus on the summarization for revision in the domain of learning.

## II. RELATED WORK

The first method of automatic text summarization was proposed by Luhn [7]; it was based on term frequency which he used to compute the significance of terms. The idea was to extract from a document the most significant sentences, which contained the highest number of occurrences of significant terms. Edmundson [5] considered not only the frequency of

terms, but also their location. Terms in the title, the first and the last paragraph and the first and the last sentence in each paragraph were assigned positive weights.

Gong and Liu [6] were first to use latent semantic analysis (LSA) for the text summarization. This method is capable of finding salient topics or concepts in the document and also their relative importance within the document. Each concept (topic) is represented in the summary by a sentence, which captures it the best. However, Steinberger and Jeek [10] showed that this approach fails to include into the summary sentences, which capture many concepts well, but have the highest score for none of them. They proposed a modification of the selection of sentences; sentences are selected based on their overall score computed as a combination of scores for each concept (topic).

All the methods mentioned so far summarize only the content of the document. However, there are many types of information which can indicate a relevance of the sentences to extract, especially on the Web, e.g. user activity or user-added annotations (comments, highlights, tags etc.). Sun et al. [11] utilized clickthrough data which records how users find information through queries; if a user clicks on a link to a web page which is one of the results of her query, it indicates that terms from the query describe the page and can be given more weight when summarizing the page. Park et al. [9] summarized not the content, but the comments (descriptions) and tags which users add when they create a bookmark using social bookmarking service such as Delicious. The advantage of this approach is that it can summarize also documents with no or minimum text, but with other multimedia content. On the other hand, it depends on the number of added bookmarks and is unable to summarize documents with no bookmarks.

Methods of personalized summarization also use additional information. However, their result is not a generic summary which is the same for all the users, but a summary that is adapted (personalized) to the characteristics of a particular user. Daz et al. [4] personalized summarization to mirror users interests represented by a user model in the form of a vector of weighted keywords; the disadvantage of this approach is that the users have to manually insert keywords and weights into the model. Campana and Tombros [3] automatically built a user model from the sentences of the documents recently read by a user. Summary is constructed from the sentences which are the most similar to the most representative sentences from the user model. Zhang et al. [14] utilized user annotations in the form of highlights by extending the classical tf-idf method.

Results of their study suggest that this approach is useful

for personalization of summarization. They identify determining a subset of annotations suitable for summarization and including of collaborative annotating as open problems.

### III. PERSONALIZED TEXT SUMMARIZATION

We propose a method of personalized text summarization based on a method of latent semantic analysis [6][10] which consists of the following steps:

- A. Pre-processing
- B. Construction of a personalized terms by sentences matrix
- C. Singular value decomposition (SVD)
- D. Sentences selection

#### A. Pre-processing

The pre-processing phase includes machine translation, tokenization, terms extraction and segmentation of the text into sentences.

We use machine translation of the document to a reference language (in our case English) in order to maximize our methods independency of the summarized documents language. Because the methods of terms extraction and segmentation of the text into sentences performed during the pre-processing are language dependent, the use of machine translation enables us to provide their implementations only for one (reference) language and effectively cover a wide range of languages. Certainly, this can influence the quality of the resulting summaries; however, our experiments show that the commercially available machine translation (using Bing Translator, [www.microsofttranslator.com](http://www.microsofttranslator.com)) gives reasonable results.

#### B. Construction of a personalized terms by sentences matrix

We have identified the construction of a terms-sentences matrix representing the document as a step suitable for personalization of the summarization. In this step terms extracted from the document are assigned their respective weights. Our proposed weighting scheme extends the conventional weighting scheme based on tf-idf method by a linear combination of the multiple raters, which positively or negatively affect the weight of each term (see Fig. 1).

We formulate the weighting scheme as follows:

$$w(t_{ij}) = \sum_k \alpha_k(t_{ij}), \quad (1)$$

where  $w(t_{ij})$  is a weight of a term  $t_{ij}$  in the matrix and  $k$  is a linear coefficient of a rater  $R_k$ . Both the weights  $w(t_{ij})$  and the linear coefficients  $k$  can be any real number. The rater  $R_k$  is a function, which assigns each term from the extracted keywords set  $T$  its weight:

$$R_k : T \rightarrow \mathbb{R}_k. \quad (2)$$

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

- *Generic raters*: terms frequency rater, terms location rater and relevant domain terms rater
- *Personalized raters*: knowledge rater and annotations rater

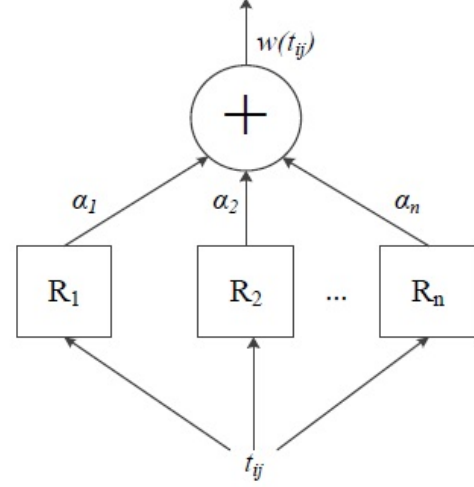


Fig. 1: Term weighting by a combination of raters.

Generic raters take into account the content of the document and some additional information to adapt summarization regardless of a particular user. On the other hand, personalized raters consider information about the specific user and her characteristics. Selection and combination of the raters (the values of their linear coefficients) depends on a specific summarization scenario and the types of available additional information.

Terms frequency rater and terms location rater are the two basic generic raters, the design of which was inspired by Luhn [7] and Edmundson [5]. The former assigns the weights based on tf-idf method, the latter based on the location of terms; terms in the title and the first and the last sentence of a document are given positive weights.

Because we focus on the domain of learning, we have identified three main sources of personalization and adaptation of the summarization suitable for the chosen domain:

- Domain conceptualization in the form of the relevant domain terms
- Knowledge of the users
- Annotations added by users, i.e. highlights or tags

We have designed a relevant domain terms rater to utilize information contained in a domain model of an adaptive system [13]. Domain models are usually constructed manually by domain experts (however, it is useful to utilize folksonomies for this purpose as well, as we have shown in [8]) by capturing their knowledge of the domain in the form of important

concepts (relevant domain terms) and relationships among them which makes them valuable sources of information for adapting the summarization (if they are available).

Let  $C_d$  be a set of concepts associated with a document  $d$ ; each concept in the set is represented by ordered pair  $(t_i, w_i)$ , where  $t_i$  is a relevant domain term  $i$  and weight  $w_i$  represents a measure of association between document  $d$  and concept  $i$  and it is a real number from the interval  $0, 1$ .

We formulate the relevant domain terms rater as follows:

$$\begin{aligned} w(t_{ij}) &= w_i & \text{if } t_i \in S_j \cap C_d \\ w(t_{ij}) &= 0 & \text{else,} \end{aligned} \quad (3)$$

where  $S_j$  is a set of terms contained in the sentence  $j$ .

Educational systems usually use overlay user models to record the level of knowledge of each concept from the domain for each particular user. We have designed a knowledge rater as a personalized version of the previous one. It uses information captured in the user model [1]; so instead of  $w_i$  representing the measure of association between document  $d$  and concept  $i$ , we now use  $k_{iu}$  reflecting a level of knowledge of the concept  $i$  by a user  $u$ . This way, concepts which are better understood by the user are given more weight, which is especially useful in the knowledge revision scenario.

Although learning systems usually assume that modeled knowledge only grows in time, users in fact do forget a part of their acquired knowledge [2]. The knowledge revision represents a means of re-acquiring the forgotten knowledge. We believe that summarization is suitable for this scenario, because it can help users to remind them of the important concepts explained in the documents.

Annotations rater takes into account the fragments of the text highlighted by a particular user. First, we construct a set of all the sentences, fragments of which were highlighted by a particular user:

$$S_H^u = \{S_j | S_j \cap H_u \neq \emptyset\} \quad (4)$$

where  $S_j$  is  $j$ -th sentence of the document and  $H_u$  is a set of all the highlights made by the particular user  $u$ . Similarly, we construct a set of all the sentences, fragments of which were highlighted by all the users, taking only those for which the following condition stands true:

$$|h_j| \geq \frac{\max_l |h_l|}{2} \quad (5)$$

where  $|h_j|$  is a number of times  $j$ -th sentence  $S_j$  was highlighted. Lastly, we make a union of these two sets and assign positive weights to those terms of the document, which are located in the sentences in the resulting set  $S_H$ :

$$w(t_{ij}) = 1 \text{ if } S_j \in S_H, w(t_{ij}) = 0 \text{ else,} \quad (6)$$

where  $S_j$  is  $j$ -th sentence of the document.

### C. Singular value decomposition

Constructed personalized terms-sentences matrix  $A$  is decomposed using a singular value decomposition [6]:

$$A = U \Sigma V^T, \quad (7)$$

where  $U = (u_{ij})$  is a matrix, the columns of which represent left singular vectors and the rows terms of the document,  $\Sigma = \text{diag}(1, 2, \dots, n)$  is a diagonal matrix with diagonal elements representing the non-negative singular values in the descending order and  $V = (v_{ij})$  is a matrix, the columns of which represent right singular vectors and its rows represent sentences of the document.

### D. Sentences selection

As the final step, we select sentences with the highest score computed by a method proposed in [10] (sentences are selected from the original, not the translated document):

$$S_k = \sqrt{\sum_{i=1}^n v_{ki}^2 \sigma_i^2} \quad (8)$$

where  $s_k$  is a score of  $k$ -th sentence,  $v_{ki}$  is a value from matrix  $V$  that measures how well concept  $i$  is represented by sentence  $k$  and  $i$  is a singular value that represents relative relevance of the concept  $i$  in the document;  $n$  is a number of dimensions and it is a parameter of the method. The length of the generated summary is a parameter of our method as well.

## IV. EVALUATION

We have experimented in the domain of learning choosing the knowledge revision as our specific use case. Our dataset has consisted of the educational materials from the Functional and Logic Programming course in the educational system ALEF.

ALEF (Adaptive Learning Framework) [12] is an application framework which merges the concepts of the traditional web-based educational systems with the principles of the Web 2.0, i.e. it brings several interactive features such as tagging, commenting, collaborative task solving etc. We have integrated our implementation of the summarizer with ALEF (see Fig. 2).

In our experiment, we have focused on evaluation and comparison of the two variants of summaries:

- generic summarization and
- summarization considering the relevant domain terms identified by a domain expert.

We have generated both variants for each document in our dataset; because we focus on the summarization for revision, we have chosen the length of generated summaries to be approximately one third of the document length.

We have asked the Functional and Logic programming course students to evaluate the quality of the generated summaries; 17 students have participated in our experiment. Their task has been to read the educational texts in ALEF and rate

Fig. 2: Example screenshot of ALEF (in Slovak) with the integrated summarizer (1 highlighted by the border). Current user rating is shown in the right top corner (2). Users rated summaries on a five-point scale using stars; after each rating they were automatically asked a follow-up question. They could have also added feedback in the form of free text (3) or navigate themselves to the next summary by clicking the Next button in the right bottom corner (4).

presented summaries on a five-point scale without knowing what variant of summary is presented to them. Summaries have been placed underneath the text so that the students would read the document first and its summary second. In the real usage the summary is positioned above the text so that it gives users (students) an overview of the topics of the text and it can help them to decide whether to read the whole text or not.

After each summary rating, the students have been asked a follow-up question to further evaluate the summary quality. We have asked them whether the sentences selected for the summary are representative, whether the summary is suitable for revision or whether it could help them to decide the document relevance. We have also inquired whether the length of the summary is suitable given the length and content of the document and if it is readable and comprehensible.

Furthermore, we have chosen a group of five excellent students as an expert group to compare their summary evaluation to that of the other students. In contrast to the other participants, they have been presented both summary variants (in random order) for each educational text in order to decide which variant is better or whether they are equal. We have gathered summaries for 79 educational texts (explanation learning objects) in Slovak, 278 summary ratings and 154 summary variants comparisons from students-experts. Moreover, students have answered 275 follow-up questions.

The second variant (summarization considering the relevant domain terms) has on average gained approximately 7.2% higher score on a five-point scale compared to the first vari-

ant (generic summarization using only terms frequency and location rater) as can be seen in Tab. 1.

TABLE I: SUMMARY VARIANTS RATINGS

Statistic	Generic	Relevant domain terms
No. of ratings	143	135
Mean	3.538	3.793
Variance (n-1)	1.518	1.419

We have also computed average score for each summary variant for each document. The second variant has scored more in comparison to the first one in 48% of the cases, the same in 11% and less in 41%. The comparison of summaries by the students-experts has given us similar results. The second variant has been evaluated as better in 49% of the cases, as equal in 20% and worse in 31% (see Fig. 3).

The relatively high percentage of cases when the generic variant has been evaluated as better can be attributed to the fact that even though the summaries that have considered the relevant domain terms have managed to extract more relevant information, they have been sometimes less comprehensible and readable compared to the generic variant.

Still, our results suggest that considering the relevant domain terms during the summarization process leads to better summaries in comparison to the baseline generic variant. However, they also suggest that we have to pay more attention to the natural language processing problems, such as anaphora resolution in the future.

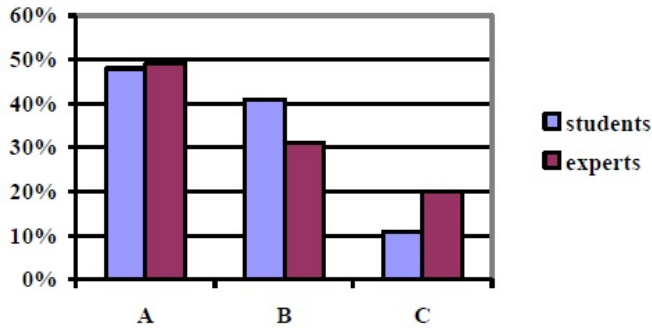


Fig. 3: Comparison of the summary variants, where A means that summary considering the relevant domain terms has been evaluated as better, B that generic summary has been evaluated as better and C that they have been evaluated as equal.

Lastly, we have evaluated the students answers to the follow-up questions. They suggest that our method in general has managed to choose representative sentences and the summaries could be in many cases used for revision or to help the students to decide whether the summarized document is relevant. Furthermore, the answers to the follow-up questions show that the notion of the summary quality is subjective. Therefore, we believe that it is useful to personalize summarization and we can get significantly better results when we take students annotations into consideration.

## V. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a method of personalized summarization, which extends existing summarization methods by considering various user characteristics including context.

Our contribution lies in the proposal of

- the specific raters that take into account terms relevant for the domain or the level of knowledge of an individual user
- the method of the raters combination which allows considering various parameters or context of the summarization.

Even though our approach is domain independent, it allows us to identify the sources of personalization and adaptation of the summarization specific for the chosen domain and to adapt the method for a particular scenario (in our case students knowledge revision).

We work towards providing the users with better summaries reflecting their particular needs by taking into account also their annotations (highlights) that indicate users interest in the fragments of a document. By considering not only a users personal annotations, but the most popular ones as well, we can potentially deal with a situation when the user has not yet added any annotations to the document and also utilize the wisdom of the crowd.

We have evaluated our approach in the domain of learning in the knowledge revision scenario. In the first phase of the evaluation, we have focused on the comparison of the two summary variants: the generic summarization and the summarization considering the relevant domain terms. Our experimental results suggest that using the relevant domain terms in the process of summarization can help selecting representative sentences capable of summarizing the document, even for revision.

As our future work, we plan to carry out more experiments with the two raters: the knowledge rater and the annotations rater. We believe that considering the knowledge as well as users annotations in the summarization process will lead to summaries better adapted for a particular users needs. We also see a potential in extending our method so that the raters combination parameters would not have to be set manually, but automatically and dynamically based on the reliability of the raters sources of information or the category of the summarized document.

## ACKNOWLEDGEMENT

This work was partially supported by the grants VG1/0971/11/2011-2014, APVV-0208-10 and it is the partial result of the Research & Development Operational Programme for the project Research of methods for acquisition, analysis and personalized conveying of information and knowledge, ITMS 26240220039, co-funded by the ERDF.

## REFERENCES

- [1] H. Zhang, Z. C. W.-y. Ma, and Q. Cai, "A study for documents summarization based on personal annotation," in *Proceedings of the HLT-NAACL 03 on Text summarization workshop-Volume 5*, pp. 41–48, Association for Computational Linguistics, 2003.