Item Summarization in Personalisation of News Delivery Systems*

Alberto Díaz and Pablo Gervás

[1] CES Felipe II – Universidad Complutense de Madrid
C/ Capitán 39, 28300 Aranjuez, Madrid, Spain
Email: adiaz@cesfelipesegundo.com
[2] Dep. Sistemas Informáticos y Programación
Universidad Complutense de Madrid, Spain
Email: pgervas@sip.ucm.es

Abstract. The designer of an information filtering system based on user preferences formulated as user models has to decide what method to use to provide summaries of the available documents without losing information that may be significant to a particular user even if it would not be considered as such in general terms. In this paper we describe a personalised summarization facility to maximise the density of relevance of information sent by the system. The selection uses a relevance feedback mechanism that captures short term interests as indicated by a user's acceptance or rejection of the news items received. Controlled experiments were carried out with a group of users and satisfactory and insightful results were obtained, providing material for further development. The experimental results suggest that personalised summaries perform better than generic summaries at least in terms of identifying documents that satisfy user preferences.

1 Introduction

Personalised information systems typically send to the users the title and the first lines of the items that are detected as interesting, and links to the full text. This is in most cases insufficient for a user to detect if the item is relevant or not, forcing him to inspect the full text of the document. An interesting approach is to replace the first sentences sent as a sample of a document by a proper summary or extract.

Personalised summarization is understood as a process of summarization that preserves the specific information that is relevant for a given user profile, rather than information that truly summarises the content of the news item. The potential of summary personalization is high, because summaries obtained in a generic manner may mislead the user into disregarding documents that may catch his atention if sentences that match the user interest are selected during summarization.

In this paper we defend the use of a personalised summarization facility to maximise the density of relevance of selections sent by a personalised information system to a given user. Two sources of information about the user are employed to construct the summaries. The long term user model allows the introduction of user-defined keywords. An additional short-term component was added to the user model. The user gets the possibility of voting for or

* Supported by the Spanish Committee of Science & Technology (TIC2002-01961).

Petr Sojka, Ivan Kopeček, and Karel Pala (Eds.): TSD 2004, LNAI 3206, pp. 49–56, 2004.

against the news items received. Significant keywords are then extracted from the read item and fed back into the user model. A personalised summary is obtained for each news item that has been deemed relevant. These summaries give the user immediate access to the part of the news item that is relevant to his profile, providing information on a news item – beyond its title – for a quick decision on whether the full item might be interesting. Users were asked to evaluate the total collection of news items from which the summaries had been obtained. This collection of data is used to evaluate the adequacy of the summaries in an information filtering context.

2 Automatic Summarization

Automatic summarization is the process through which the relevant information from one or several sources is identified in order to produce a briefer version intended for a particular user – or group of users – or a particular task [1]. According to their scope and purpose [2], the summaries considered in this paper will be restricted to a single document – rather than a set of documents – and they will be indicative – their aim is to anticipate for the user the content of the text and to help him to decide on the relevance of the original document. With respect to their focus, we can distinguish between: *generic abstracts*, if they gather the main topics of the document and they are addressed to a wide group of readers, and *user adapted abstracts*, if the summary is constructed according to the interests – i.e. previous knowledge, areas of interest, or information needs – of the particular reader or group of readers that the system is addressing. It has already been shown that in an information retrieval environment summaries adapted to the user query outperform other kinds of summary [3].

Techniques for selection and extraction of phrases are very attractive due to their domain and language independence. In these techniques the segments of text – usually sentences or paragraphs – that contain the most significant information are selected based on linear combination of the weights resulting from the application of a set of heuristics applied to each of the units of extraction. These heuristics may be *position dependent*, if they take into account the position that each segment holds in the document; *linguistic*, if they look for certain patterns of significant expressions; or *statistical*, if they include frequencies of occurrence of certain words. The summary results from concatenating the resulting segments of text in the order in which they appear in the original document [4].

3 Applying Long and Short Term User Models to Personalise Summaries

The user model has to be adapted to the different aspects of each domain, in order to allow a better definition of the user interests [5]. The specific domain chosen for this work is the electronic newspaper domain. The system uses various reference frameworks to obtain from the user different views or descriptions of his interests.

In completing his user profile, each user is asked to type in a number of *chosen keywords*. The occurrence of these keywords in a news item is taken to indicate it may be of interest to the user. For each keyword the user also indicates a rough weighting to show the degree of his interest. The keywords typed in by the user are represented as term weight vectors [6], using the weight assigned to each word in the current user model. This constitutes the long term part of the user model.

The messages sent to the user with the daily selection of news items allow the user to provide positive or negative feedback on each news item received. Each news item that has been selected appears with two underlined links which the user can click on to send either positive or negative feedback to the system. This information is stored, a set of *feedback keywords* is extracted from the chosen news items, and these feedback keywords are used to update the short term part of the user model [7].

Documents are downloaded from the web of the newspaper as HTML documents. For each document, title, category, URL and text are extracted and stored for later processing. A term weight vector representation for a document d (d_d) is obtained by application of a stop list, a stemmer, and the *tf-idf* formula for computing actual weights [6].

Our system uses three phrase-selection heuristics to build summaries: two to construct generic summaries, and one for personalized summaries. To generate summaries a value is assigned to each sentence of the text being summarized, obtained as a weighted combination of the results of the three heuristics. This value is used to select the most relevant sentences, which will be used to form an extract of the news item later used as summary.

The *position heuristic* assigns the highest value to the first five sentences of the text [8]. The specific values chosen in our system are shown in Table 1. Decreasing values have been chosen to account for the traditional structure of news items as inverted pyramid with most relevant sentences at the beginning. Sentences from the 6th on are assigned the value 0. These provide the weights A_{oi} for each sentence o of a news item i using the position heuristic. These values are independent from the particular user j being considered.

Table 1. Values assigned to the first 5 sentences of a document for the position heuristic

Sentence number	1	2	3	4	5
Assigned value	1.00	0.99	0.98	0.95	0.90

Each text has a number of thematic words, which are representative of its content[3]. The *thematic words heuristic* extracts the M non-stoplist most significant words of each text and checks how many of these thematic words are found in each sentence. In this way, a higher value will be assigned to sentences that hold a highest number of thematic words [4,9].

To obtain the M most significant words of each document, documents are indexed to provide the weight of each word in each document using the *tf-idf* method [6]. The eight words with highest weight are selected for each document ($M = 8$).

To obtain the value for each sentence o within the document i using the thematic words heuristics (B_{oi}), the number of thematic words appearing in the sentence is divided by the total number of words in the sentence. This is intended to give more weight to sentences with a higher density of thematic words [9]. The values obtained in this way are also independent from the particular user j being considered.

[3] This set of content based keywords for a document should not be confused with the set of keywords specified by a user to define his interests.

The *personalization heuristic* boosts those sentences that are more relevant to a particular user model. The user model provides a vector of weighted terms (k_j) corresponding to the chosen keywords of the long term model and a vector of weighted terms (f_j) corresponding to the feedback keywords of the short term model. This information is used to calculate the similarity between the user model j and each sentence o of news item i, assigning a final weight to the sentence by means of the following formula:

$$C_{oij} = \frac{\chi sim(s_{oi}, k_j) + \beta sim(s_{oi}, f_j)}{\chi + \beta} \quad (1)$$

where s_{oi} is the term weight vector representing the sentence o of news item i and sim is the cosine formula of the Vector Space Model [6].

The following equation is applied to combine the values resulting from each of the three heuristics and provide a single value for each sentence:

$$Z_{oij} = \frac{\mu A_{oi} + \nu B_{oi} + \sigma C_{oij}}{\mu + \nu + \sigma} \quad (2)$$

The parameters μ, ν and σ allow relative fine-tuning of the different heuristics, depending on whether position (μ), thematic key words (ν) or relevance to the user model (σ) is considered more desirable. Values of σ determine the degree of personalisation of the summaries: if σ is 0, the resulting summaries are generic, and for σ greater than 0 personalisation increases proportionally to σ. In order for this combination to be significant, the relevance obtained for each framework must be normalised with respect to the best results for the collection of documents under consideration.

The summary is constructed by selecting the top 20% of the ranking of sentences by the value Z_{oij} and concatenating them according to their original order of appearance in the document.

4 Evaluation of Personalized Summarization

The main issue regarding the use of summarization in an information dissemination setting is to what extent the use of summaries instead of the complete document involves a loss of significant information for the user.

Summaries are evaluated using a technique of indirect evaluation [3]. The technique is based on the assumption that if a summarization process is good, the resulting summary should have retained as much as possible of the information that ensures correct retrieval according to the given user profile. For each user, a personalised version of given collection designed for the evaluation of an information filtering system [10] is built by summarising each news item (using the heuristic that is to be tested). A process of selection equivalent to that applied to the set of complete news items for evaluating the information filtering system is applied to the generated summaries. The selection mechanism employed combines short term and long term models. The overall relevance between a news item d – belonging to a section s – and a user model j is computed using the following formula:

$$r_{dj} = \frac{\alpha S_{sj} + \gamma sim(d_d, k_j) + \delta sim(d_d, f_j)}{\alpha + \gamma + \delta} \quad (3)$$

where Greek letters γ and δ show the significance assigned to each of the different references frameworks (γ for keywords, and δ for feedback keywords)[4]. In order for this combination to be significant, the relevance obtained for each framework must be normalised with respect to the best results for the collection of documents under consideration.

The hypothesis is that, if the summarization process employed preserves the information that is relevant for that user profile, the results obtained should mirror exactly those obtained for the collection with the complete news items, which are taken as reference value. Any deviations from that value indicate loss of information due to "leaks" during summarization, which have forced the resulting ranking for the input items to deviate from the one obtained using the complete news item as input. By applying this process to different summarization heuristics, this experiment should provide an explicit – though admittedly indirect – measure of its adequacy for personalised summarization.

The experiments presented here are carried out over the evaluation collection developed for the system described in [10], built from the sets of news items corresponding to five consecutive days (Monday to Friday) from the digital edition of the ABC newspaper, a major Spanish daily. A group of 11 users were asked to provide relevance judgements for the collection. For this evaluation, summaries have been generated for all the news items for each day for all the users.

In this working framework we are considering binary relevance as stated by the users (whether or not a news item is relevant) and a ranking of the news items provided by the system. This suggested [11,6] the use of normalised precision as metric. Additionally, for cases where equal relevance values are obtained for consecutive positions in the ranking, the average position number in the ranking has been chosen as position number for the whole conflicting set [6]. This adjustment avoids the problem of attributing a random relative ordering within the ranking to documents that have obtained equal relevance values.

Normalised precision is calculated using the following formula:

$$nP = 1 - \frac{\sum_{i=1}^{REL} \log RANK_i - \sum_{i=1}^{REL} \log i}{\log N!/(N - REL)!REL!} \tag{4}$$

where: REL is the number of relevant documents, $RANK_i$ represents the ranking of the ith most relevant item, and N is the total number of items.

We consider the results to be statistically significant if they pass the sign-test on paired samples at the 5% level ($p \leq 0.05$). This decision is based on the fact that there is no assumption about the underlying distribution, and, given the different normalization procedures being applied at various levels, the relative values rather than the actual magnitudes of relevance should be considered [6].

The questions to be answered are: how much is lost, in terms of information received, by sending a summary of a news item instead of the complete document; and which type of summary is better in that sense.

We designed the experiment to test whether summaries obtained by using only the personalization heuristic are better in terms of precision with respect to information selected

[4] The term αS_{sj} and the α factor correspond to information concerning newspaper sections that may be employed by the system for news item selection but plays no role in the summarization process. For more details about the selection process, see [10].

by the user than other summaries (including those given by the first lines of the document) but worse than the complete news item.

The following types of summaries are involved:

- **Fs.** (baseline reference) 20% first sentences of the corresponding news item
- **Gs.** using generic heuristics (position and keywords) ($\mu = 1, \nu = 1, \sigma = 0$)
- **Ps.** using personalization heuristics (combining short and long term models: $\chi = 1$ and $\beta = 1$) ($\mu = 0, \nu = 0, \sigma = 1$)
- **GPs.** using both types of heuristics ($\mu = 1, \nu = 1, \sigma = 1$)

Several different evaluation collections – consisting each one of summaries obtained from the news items in the original collection by applying a different summarization method – are built for each user. The selection procedure described above is applied to each one of these collections, using the corresponding user profile as source for user interests (formula 3 with $\alpha = 1, \gamma = 1, \delta = 1$).

If different summarization methods lead to different degrees of loss of relevant information, the resulting rankings will differ amongst them in a proportional way.

Evaluation of personalised summaries is more costly because each user must be evaluated separately. Generic summaries are the same for all users, so all the users can be evaluated simultaneously in a single system run, using the one set of generic summaries as input. Personalised summaries require a different procedure, since the selection for each particular user must be obtained from his own set of personal summaries.

Table 2. Global averages for normalised precision (nP) for complete news items (Nw) and various types of summaries (Ps, GPs, Gs, Fs)

	Nw	Ps	GPs	Gs	Fs
Averages					
nP	0,526	0,523	0,506	0,505	0,500

The analysis of the results presented in Table 2 shows that personalised summaries (**Ps**) give better results with respect to normalised precision of the selected information than generic summaries (4% **Gs**) and generic-personalised summaries (3% **GPs**). In both cases the improvement is statistically significant. The use of personalized summaries was shown to have worse precision than the full text document (1%). However, the difference shown is not statistically significant. Generic-personalised summaries (**GPs**) are better than generic summaries (**Gs**), and these (**Gs**) are better than baseline summaries (**Fs**), but in neither case is the difference statistically significant. The fact that personalised summaries performed significantly better (5%) than baseline summaries makes it possible to substitute the complete news item for a personalised summary with an acceptable loss of information where convenient.

5 Conclusions

In this paper, we have presented a summarization subsystem that generates different kinds of summaries adapted to the user, allowing the users to decide about the relevance of the received news items without inspecting the full text document.

It seems apparent from the results presented here that generic summaries perform very closely to summaries obtained by taking the first few lines of the news item. This seems to indicate that the position heuristic is overpowering the thematic word heuristic, which may be corrected by refining the choice of weights. In any case, although a first-sentences approach may provide good results for indicative summarization, it does not do so well in terms of personalised summarization (as defined above), where it is crucial to retain in the summary those specific fragments of the text that relate to the user profile.

This explains why the generic-personalised summaries perform so poorly in spite of being a combination of good techniques: given a fixed limit on summary length, the inclusion of sentences selected by the generic heuristics in most cases pushes out of the final summary information that would have been useful from the point of view of personalisation.

The methods proposed here ensure the efficient selection of relevant information for personalised summarization – the user receives an extract of the contents of a document that are related to his interests –, with no domain dependent assumptions. User adapted summaries are a useful tool to assist users in a personalization system. However, the information contained in these summaries cannot replace the full text document from an information retrieval point of view.

In future work, we will try to explore the possibility of obtaining feedback for the user models from the different kinds of summaries and explore its effectiveness. We are also interested in carrying out experiments with more users and during more days to extract more informative conclusions. Another line of research could be to add more information to the profile to improve the modeling of the users and to explore other techniques to perform the feedback.

References

1. Mani, I., Maybury, M., eds.: Advances in Automatic Text Summarization. The MIT Press, Cambridge, Massachusetss (1999).
2. Hahn, U., Reimer, U.: Knowledge-based text summarization: Salience and generalization operators for knowledge-based abstraction. In: Mani, I., Maybury, M., eds.: Advances in Automatic Text Summarization, Cambridge, Massachusetss, The MIT Press (1999).
3. Maña, M., de Buenaga, M., Gómez, J.M.: Using and evaluating user directed summaries to improve information access. In: Abiteboul, S., Vercoustre, A., eds.: Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries (ECDL'99), Lecture Notes in Computer Science, Vol. 1696, Springer-Verlag (1999) 198–214.
4. Kupiec, J., Pedersen, J.O., Chen, F.: A trainable document summarizer. In: Research and Development in Information Retrieval. (1995) 68–73.
5. Amato, G., Straccia, U.: User profile modeling and applications to digital libraries. In: Abiteboul, S., Vercoustre, A.M., eds.: Proc. 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL. Number 1696, Springer-Verlag (1999) 184–197.
6. Salton, G.: Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer. Addison-Wesley Publishing, Reading, Massachusets, US (1989).

7. Nakashima, T., Nakamura, R.: Information filtering for the newspaper. In: IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, August 20–22, 1997, Victoria, B.C., Canada (1997).
8. Edmundson, H.: New methods in automatic abstracting. Journal of the ACM **2** (1969) 264–285.
9. Teufel, S., Moens, M.: Sentence extraction as a classification task. In: Proceedings of ACL/EACL Workshop on Intelligent Scalable Text Summarization, Madrid, Spain (1997) 58–65.
10. Díaz, A., Gervás, P.: Dynamic user modeling in a system for personalization of web contents. In: Conejo, R., ed.: Proceedings of CAEPIA 2003. Lecture Notes in Artificial Intelligence 3104. Springer (2003).
11. Mizzaro, S.: A new measure of retrieval effectiveness (or: What's wrong with precision and recall). In: Ojala, T., ed.: International Workshop on Information Retrieval (IR'2001), Infotech Oulu. (2001) 43–52.