Towards Web Browsing for Visually Impaired People

Gaël Dias and Bruno Conde
Centre for Human Language Technology and Bioinformatics
University of Beira Interior, Portugal

Abstract

Although a lot has been done for Visually Impaired People to access information with Braille screens, Braille keyboards, Braille PDAs and Text-to-Speech interfaces, very little has been made to reduce the amount of information they have to deal with. In this paper, we propose an automatic summarization system to ease web browsing for visually impaired people on PDAs.

1. Introduction

Visually impaired people (VIP) are info-excluded due to the overwhelming task they face to read information on the web. Unlike fully capacitated people, blind people can not read information by just scanning it quickly i.e. they can not read texts transversally. As a consequence, they have to come through all sentences of web pages to understand if a document is interesting or not.

Although a lot has been done for blind people to access information with Braille screens, Braille keyboards, Braille PDAs and Text-to-Speech interfaces, very little has been made to reduce the amount of information they have to deal with.

To solve this problem, three main approaches have already been proposed in the literature. First, some methodologies such as [1] [2] use simple but fast summarization techniques to produce results in real-time. Second, some works apply linguistic processing and rely on *ad hoc* heuristics [3] to produce compressed contents but can not be used in a real-time environment. Third, some approaches like [4] propose hybrid solutions by applying efficient algorithms for linguistic treatment [5] [6] that allow real-time processing and deeper linguistic analysis of web pages, thus producing quality content visualization. In this paper, we extend the work done by [4] by proposing a new summarization technique based on the TextRank algorithm [7] and implementing a Text-to-Speech module which will further be upgraded to a Speech-to-Speech module for full access for VIP.

2. The rw.idf

Once all Semantic Textual Units (STUs) [1] have been linguistically processed, the next step of the extractive summarization architecture is to extract the most important sentences of each STU. In order to make this selection, each sentence in a STU is assigned a significance weight. The sentences with higher significance become the summary candidate sentences.

Recently, [7] have proposed the TextRank algorithm for word weighting based on text graph representation. The basic idea of the algorithm is the same as the PageRank algorithm proposed by [8] i.e. the higher the number of votes that are cast for a vertex, the higher the importance of a vertex. Moreover, the importance of the vertex casting the vote determines how important the vote itself is, and this information is also taken into account by the ranking model. Hence, the score associated with a vertex is determined based on the votes that are cast for it, and the score of the vertices casting these votes. The score of a vertex V_i is defined as in Equation 1 [8]:

$$S(V_i) = (1-d) + d \times \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j) \quad (1)$$

where $In(V_i)$ is the set of vertices that point to it (i.e. predecessors), $Out(V_j)$ is the set of vertices that vertex V_j point to (i.e. successors) and d is a damping factor that can be set between 0 and 1, which has the role of integrating into the model the probability of jumping from a given vertex to another random vertex in the graph[1]. In our case, each STU is represented as an unweighted oriented graph being each word connected to its successor. following sequential order in the text.

After the graph is constructed (directed unweighted graph), the score associated to each vertex is set to an initial value of 1, and the ranking algorithm is run on the graph for several iterations until it converges. So, each word is then weighted as in equation 2 where *idf(.)* is the well-known Inverse Document Frequency [9]

$$rw.idf(w, stu) = S(w) \times idf(w) \quad (2)$$

and the sentence significance weight, *weight*(*S, stu*), is defined straightforwardly in Equation 3

$$weigth(S, stu) = \frac{\sum_{i=1}^{|S|} rw.idf(w_i, stu)}{|S|} \quad (3)$$

where $|S|$ stands for the number of words in S and w_i is a word in S.

[1] d was set to 0.85 as referred in [8].

3. Text-to-Speech Interface

The Text-to-Speech module is a crucial issue for accessibility of Visually Impaired People to web page contents. For this purpose, we have integrated the Microsoft Speech Server into our architecture using the SALT markup language following the architecture proposed in Figure 1.

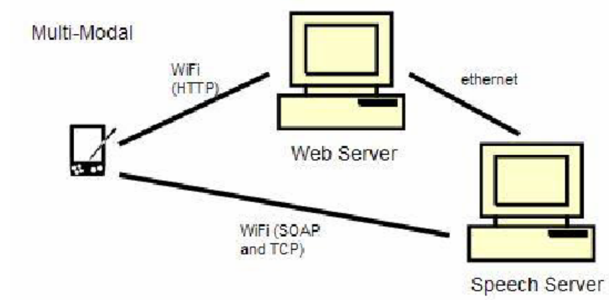

Figure 1. Text-to-Speech Interface

However, in future work, we will integrate a Speech-to-Speech module on the proper device in order to avoid the overload of the Microsoft Speech Server which has shown limitations for high amounts of requests.

8. Conclusion

In this paper, we showed the on-going progress of the work developed by [4] by introducing a new summarization technique and integrating a Text-to-Speech module. The first results are every encouraging in terms of (1) quality of the content of the summaries, especially with the rw.idf, (2) processing time although the architecture is not still distributed over different processing units and (3) user interaction satisfaction. The final result can be observed in Figure 2.

However, many improvements must be taken into account. In particular, the integration of language models, the identification of the structure of web pages and the reorganization of contents are under analysis. But most important, current work involves the integration of a Speech-to-Speech control interface which may provide a cheap integrated solution capable to compete with Braille PDAs that are expensive and difficult to use for VIP.

References

[1] O. Buyukkokten, H. Garcia-Molina and A. Paepcke. Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices. In Proc. of the 10th International World Wide Web Conference. (2000).

[2] C. Yang and F.L. Wang. Fractal Summarization for Mobile Devices to Access Large Documents on the Web. In Proc. of the International World Wide Web Conference. (2003).

[3] P. Gomes, S. Tostão, D. Gonçalves and J. Jorge. Web-Clipping: Compression Heuristics for Displaying Text on a PDA. In Proc. of 3rd Workshop on Human Computer Interaction with Mobile Devices. (2001).

[4] G. Dias and B. Conde. Efficient Text Summarization for Web Browsing On Mobile Devices. In Proc. of the Workshop on Ubiquitous User Modeling associated to the 17th European Conference on Artificial Intelligence. Riva del Guarda, Italy, August 28. (2006)

[5] T. Brants. TnT - a Statistical Part-of-Speech Tagger. In Proc. of the 6th Applied NLP Conference. (2000).

[6] A. Gil and G. Dias. Using Masks, Suffix Array-based Data Structures and Multidimensional Arrays to Compute Positional Ngram Statistics from Corpora. In Proc. of the Workshop on Multiword Expressions of the 41st ACL. (2003).

[7] R. Mihalcea and P. Tarau. TextRank: Bringing Order into Texts. In Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004). Barcelona, Spain, July. 2004.

[8] S. Brin and L. Page: The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7). 1998.

[9] G. Salton, C.S. Yang, and C.T. Yu. A Theory of Term Importance in Automatic Text Analysis. Amer. Soc. of Inf. Science. (26)1. (1975).

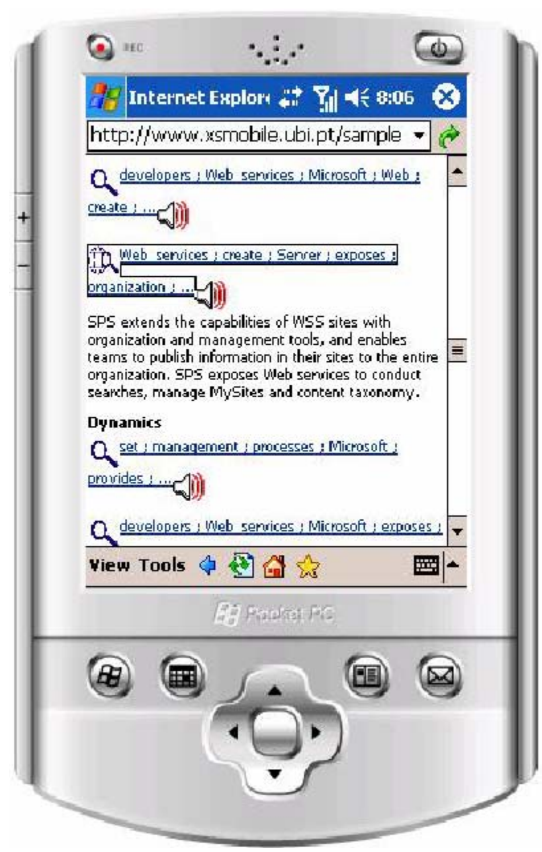

Figure 2. Screenshot of the XSMobile architecture