Abstract

Information on the World Wide Web and in other electronic form is increasing tremendously. Therefore there is a need for some form of information compression which can be achieved by various mining tasks like classification, clustering and summarization that help in understanding the information. Large amount of web content is news. News websites are daily overwhelmed with plenty of news articles. This paper presents an effective approach for single document news article summarization to help people obtain the most important information in the shortest time. The proposed approach is query based news article summarization. The results from web based on user query are filtered and refined and then result is directed to user. The technique used for summarization is keyword based extractive summarization. Keywords are the index terms that contain the most important information. The summarization technique identifies different features like thematic terms, named entity, title terms, numbers etc that are relevant to news articles to construct keyword table. This knowledge base is then used to score sentences and then top ranked sentences are presented as summary to the user. For evaluation of summary generated, extrinsic technique by question answering system is used. The purpose of using this evaluation technique is to test if the summary can be used instead of original document while preserving the overall importance of the document i.e. can summary covers all the important information of the document.
References

- Osmar R. Zaiane, "Principles of Knowledge Discovery in Data Bases"; Department of Computer Science, University of Alberta. 1999 CMPUT690
- Vishal Gupta, Jasmeen Kaur , "Effective Approaches For Extraction Of Keywords"; IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 6, November 2010
- http://www.smmry.com - "Online Automatic Text Summarization Tool"
- http://www.freesummarizer.com - "Online Automatic Text Summarization Tool"
- http://www.tools4noobs. - "Online Automatic Text Summarization Tool"
Index Terms

Computer Science
Information Sciences

Keywords

Corpus Builder; Headline Similarity; Keyword Table; Named Entities; Thematic terms