Abstract

The amount of textual information available on the web is estimated by terra bytes. Therefore constructing a software program to summarize web pages or electronic documents would be a useful technique. Such technique would speed up of reading, information accessing and decision making process. This paper investigates a graph based centrality algorithm on Arabic text summarization problem (ATS). The graph based algorithm depends on extracting the most important sentences in a documents or a set of documents (cluster). The algorithm starts computing the similarity between two sentences and evaluating the centrality of each sentence in a cluster based on centrality graph. Then the algorithm extracts the most important sentences in the cluster to include them in a summary. The algorithm is implemented and evaluated by human participants and by an automatic metrics. Arabic NEWSWIRE-a corpus is used as a data set in the algorithm evaluation. The result was very promising.

References
- Thakkar, K. and Shrawankar, U. &quot;Test Model for Text Categorization and Text Summarization&quot;. In the International Journal on Computer Science and Engineering (IJCSE), vol. 3. No. 4 Apr 2011, India.
- Violeta, S. &quot;A Collocation-Driven Approach to Text Summarization&quot;. In the TALN 2011 Montpellier, 27 juin – 1er juillet 2011.
- Gülçin, Ö., Alpaslan, F. and Çiçekli, ?. &quot;text summarization using latent semantic analysis&quot;: Master thesis, Middle East Technical University, February 2011
- Saggion, H. &quot;Using SUMMA for Language Independent Summarization at TAC 2011&quot;: In the proceeding of the TAC 2011 Workshop November, 2011, National Institute of Standards and Technology Gaithersburg, Maryland USA.
- Delort, J. and Alfonseca, E. &quot;Description of the Google update summarizer at TAC-2011&quot;: In the proceeding of the TAC 2011 Workshop November, 2011, National Institute of Standards and Technology Gaithersburg, Maryland USA.
- El-Haj, Mahmoud, Kruschwitz, Chris Fox &quot;University of Essex at the TAC 2011 Multilingual Summarization Pilot&quot;.


Index Terms

Computer Science
Information Sciences

Keywords
Text Summarization
Text Mining and Centrality Concept