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.


- Saggion, H. "Using SUMMA for Language Independent Summarization at TAC 2011". In the proceedings of the TAC 2011 Workshop November, 2011, National Institute of Standards and Technology Gaithersburg, Maryland USA.

- Delort, J. and Alfonseca, E. "Description of the Google update summarizer at TAC-2011". In the proceedings of the TAC 2011 Workshop November, 2011, National Institute of Standards and Technology Gaithersburg, Maryland USA.


- El-Haj, Mahmoud, Kruschwitz, Chris Fox "University of Essex at the TAC 2011 Multilingual Summarization Pilot".


Index Terms

Computer Science

Information Sciences

Keywords

Text Summarization  Text Mining and Centrality Concept