ABSTRACT

HILMA ZIKRA. Automatic Text Summarization System Using PageRank Algorithm. Under direction of SRI NURDIATI and AHMAD RIDHA.

With the rapid growth of World Wide Web, a huge amount of information is available and accessible online. People do not have the time to read everything, and yet they have to make critical decisions based on whatever information is available. Automatic text summarization is a technique to assist people in digesting the vast amount of information online. Recently, a number of graph-based approaches, such as PageRank, LexRank, and TextRank, have been suggested for text summarization. Automatic text summarization is a computerized process of distilling the most important information of a source (or sources) for making a brief version of text(s). This research implements graph-based summarization algorithm and similarity using graph-based ranking concept for ranking sentences. The process produces the output in the form of extractive summary that consists of high ranking sentences. The graph-based method applied is PageRank combined with cosine and content overlap similarity. Evaluation of the summaries uses three human judges and their judgement is compared using kappa measure. The result of our experiment shows that the accuracy of PageRank and cosine similarity is better than PageRank with content overlap similarity. Similarity to title produced the best result than similarity with no title.

Keywords: Automatic Text Summarization, PageRank, Similarity, Sentences Extraction.