

An experimental comparison of unsupervised keyphrase extraction techniques for extracting significant information from scientific research articles

Talha Bin Sarwar, Noorhuzaimi Mohd Noor

Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia
Pahang, Pekan, Malaysia

ABSTRACT

The automatic extraction of key information from an article that expresses all of the document's main elements is referred to as keyphrase extraction. The number of scientific research articles each year is growing. Finding a research article on relevant topics or summarizing a particular research article using important information has become time-consuming by going through the entire article. Therefore, the textual information processing task involves the automatic keyphrase extraction from a document that expresses all of the document's main elements. This article aims to make an experimental comparison of different unsupervised keyphrase extraction approaches, namely statistical-based, graph-based, and tree-based. The experiment is conducted upon 120 research articles from different subject areas of the computer science. The comparison between different techniques is made by calculating the precision, recall, and F1-score. The overall performance of the experimental result shows that KP-Miner, a statistical-based technique, outperforms all the other graph-based and tree-based techniques. Among the other techniques, the tree-based technique TeKET performs better after KPMiner. The statistical-based and tree-based approach performs better than the graph-based approach.

KEYWORDS

Unsupervised keyphrase extraction; Automatic keyphrase extraction; Statistical-based technique; Tree-based technique; Graph-based technique; Comparison analysis

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2019/ICT04-/UMP/02/1 (University Reference No. RDU1901109) and Universiti Malaysia Pahang for additional financial support.