

## Graph-Based Semantical Extractive Text Analysis

**Authors :** Mina Samizadeh

**Abstract :** In the past few decades, there has been an explosion in the amount of available data produced from various sources with different topics. The availability of this enormous data necessitates us to adopt effective computational tools to explore the data. This leads to an intense growing interest in the research community to develop computational methods focused on processing this text data. A line of study focused on condensing the text so that we are able to get a higher level of understanding in a shorter time. The two important tasks to do this are keyword extraction and text summarization. In keyword extraction, we are interested in finding the key important words from a text. This makes us familiar with the general topic of a text. In text summarization, we are interested in producing a short-length text which includes important information about the document. The TextRank algorithm, an unsupervised learning method that is an extension of the PageRank (algorithm which is the base algorithm of Google search engine for searching pages and ranking them), has shown its efficacy in large-scale text mining, especially for text summarization and keyword extraction. This algorithm can automatically extract the important parts of a text (keywords or sentences) and declare them as a result. However, this algorithm neglects the semantic similarity between the different parts. In this work, we improved the results of the TextRank algorithm by incorporating the semantic similarity between parts of the text. Aside from keyword extraction and text summarization, we develop a topic clustering algorithm based on our framework, which can be used individually or as a part of generating the summary to overcome coverage problems.

**Keywords :** keyword extraction, n-gram extraction, text summarization, topic clustering, semantic analysis

**Conference Title :** ICEMNL 2023 : International Conference on Empirical Methods in Natural Language Processing

**Conference Location :** Kuala Lumpur, Malaysia

**Conference Dates :** February 06-07, 2023