Abstract

Keywords are list of significant words or terms that best present the document context in brief and relate to the textual context. Extraction models are categorized into either statistical, linguistic, machine learning or a combination of these approaches. This paper introduces a model for extracting keywords based on their relatedness weight among the entire text terms. Strength of terms relationship is evaluated by semantic similarity. Document terms are assigned a weighted metric based on the likeness of their meaning content. Terms that are strongly co-related to each other are highly considered in individual terms semantic similarity. Provision of the overall terms similarity is crucial for defining relevant keywords that most expressing the text in both frequency and weighted likeness. Keywords are recursively evaluated according to their cohesion to each other and to the document context. The proposed model showed enhanced precision and recall extraction values over other approaches.

References

Keyword Extraction using Semantic Analysis

- Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. WordNet::Similarity – measuring the relatedness of concepts. In Proceedings of the Fifth Annual Meeting of the North American Chapter of the Association for Computational Linguistics, Boston, Massachusetts,
Keywords Extraction using Semantic Analysis

2004.
- BBC dataset, Machine Learning Group; http://mlg.ucd.ie/.
- Jasmeen Kaur and Vishal Gupta, "Effective Approaches for Extraction of Keywords"., International Journal of Computer Science Issues (IJCSCI), ISSN (Online): 1694-0814, Vol. 7, Issue 6, pp 144-148, November 2010

Index Terms

Computer Science
Artificial Intelligence

Keywords

Keywords Extraction  Semantic Similarity  Semantic Relatedness  Semantic Analysis  Semantic Word Sense