Abstract

Measuring the similarity between words, sentences, paragraphs and documents is an important component in various tasks such as information retrieval, document clustering, word-sense disambiguation, automatic essay scoring, short answer grading, machine translation and text summarization. This survey discusses the existing works on text similarity through partitioning them into three approaches; String-based, Corpus-based and Knowledge-based similarities. Furthermore, samples of combination between these similarities are presented.

References

- Jaro, M. A. (1989). Advances in record linkage methodology as applied to the 1985
A Survey of Text Similarity Approaches


A Survey of Text Similarity Approaches

NODALIDA '09, Odense, Denmark.

**Index Terms**

Computer Science  
Artificial Intelligence

**Keywords**

Text Similarity  
Semantic Similarity  
String-Based Similarity  
Corpus-Based Similarity  
Knowledge-Based Similarity