Abstract

With the documents increasing amount available in local or Web repositories, the comparison methods have to analyze large documents sets with different types and terminologies to obtain a response with minimum documents and with as much useful content to the user. For large documents sets where each document can contain many pages, it is impossible to compute the similarity using the entire document, to require creating solutions to analyze a few meaningful terms, in summary form. This article presents TextSSimily, a method that compares documents semantically considering only short text for comparison (text summary), using semantics to improve the set of responses and summaries to improve time to obtain results for large sets of documents.

References

- D. Sankoff and J. Kruskal, Time Warps, String Edits, and Macromolecules: The Theory

**Index Terms**

Computer Science  
Information Sciences

**Keywords**

Semantic Similarity  
Comparison by Similarity  
Short Text Comparison