Abstract

Document summarization is an important step while clustering the large no. of digital documents data base. Documents are clustered in accordance with their contents using the document text summary. The document summarization involves the knowledge corpus scheme comprising of corpus coverage, sentence coverage and term coverage weight. Further, three new weights are introduced as super sentence coverage weight, super corpus coverage weight and super term coverage weight. Super coverage weight is based on synonyms of the key words. The quality of document summary improves and diversified when synonyms of key words are also given due weightage in the process of text processing. The evaluation for the document summary quality is based on inner content metrics precision, recall, F-measure method.

References

- Li Chengcheng, "Automatic Text Summarization Based On Rhetorical Structure..."
Theory”, 978-1-4244-7237-6 2010 IEEE.
- Egitim Fakültesi, Mehmet Akif Ersoy,” Quality of written summary texts: An analysis in the context of gender and school variables”, 1877-0428 © 2010 Published by Elsevier Ltd.
- Kushal Bafna, Durga Toshniwal,” Feature Based Summarization of Customers’ reviews; Reviews of Online Products”, 2013 The Authors. Published by Elsevier B. V.
- Tiedan Zhu, Kan Li,” The Similarity Measure Based on LDA for Automatic Summarization”, 2011 Published by Elsevier Ltd

Index Terms

Computer Science
Information Science

Keywords
Super Sentence Coverage Weight  Super Corpus Coverage Weight  Super Term Coverage Weight
Document Summarization

Knowledge Corpus

Synonyms