Abstract

In the present work we have developed an algorithm based on maximum entropy and stop word removal modules, which works with almost 99% accuracy and have established supremacy over the existing paragraph breaker software developed by Text Mining Group, School of Computer Science, Manchester University, United Kingdom.

Reference

  Proceeding of the NAACL HLT2009:Association for computational linguistics, (short
  papers),241-244.Boulder,Colorado.
- Kiss T. and Strunk J.(2006) Unsupervised multilingual sentence boundary detection,
  Computational linguistics, 32(4), 485-525.
  264-271, Seattle, WA.
  Disambiguation. Computation Linguistics, 23(2), 241-269.
- Siminski Krzysztof (2007) Sentence boundary verification in Polish text, Computer
- Weijian, Xuan, Watson, Stanley J. and Meng Fan (2007) Tagging sentence bares to
  Biomedical literature, Computational linguistics and Intelligent text processing, Lect. Notes in

Index Terms

Computer Science  
Computational Linguistics

Key words

Sentence Boundary  
Information retrieval

Evaluation