



Sistemas & Telemática
ISSN: 1692-5238
EditorSyT@icesi.edu.co
Universidad ICESI
Colombia

Cabeza Ruiz, Robin
Text segmentation by language
Sistemas & Telemática, vol. 14, núm. 38, 2016, pp. 63-72
Universidad ICESI
Cali, Colombia

Available in: <http://www.redalyc.org/articulo.oa?id=411547493004>

Abstract

There are two approaches for text segmentation by language: first, assuming that language changes happen in the “border” between sentences (never within a sentence); second, assuming that language changes can happen anyplace in the text. This work presents methods for both types of text’s segmentation by languages. On the first proposal, the text is initially segmented by sentence, then the language of each sentence is obtained; the second proposal is an adaptation of hidden Markov model to this task. Both cases, according to results obtained in experimental proofs, exceed the state of art.

Keywords

Hidden Markov model, text segmentation by language, natural language processing.

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org



Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative