Abstract

The idea to link texts with related content in an automated way is not new. The approach developed and presented here has two main goals: to work automatically on unstructured texts and to support a large number of parallel accesses. It is distinct from other approaches in that it determines the semantic distance between texts on the basis of an asymmetrical pre-calculated distance matrix. The relations between unstructured text objects are generated by a language independent heuristic algorithm for feature selection. The resulting bag of words is used in a query to select matching texts. The user receives recommendations to texts the content of which is related to the text that appears in front of him. Performance was analyzed on basis of the algorithms runtime complexity. Extensive real-life tests over a period of 12 months were conducted on the website of an industrial magazine in order to check the efficiency of the procedure with regard to quality of the recommendations. Results show that the presented approach nearly equals the quality of manual made recommendations by professional editors.