Supporting Topic Map Creation Using Data Mining Techniques

Witold Abramowicz, Tomasz Kaczmarek, Marek Kowalkiewicz

Abstract

There is an increasing interest in automating creation of semantic structures, especially topic maps, by taking advantage of existing, structured information resources. This article gives a preview of the most popular method – based on RDF triples, and suggests a way information sources. The method can be applied in information systems development domain when analysing vast unstructured data repositories in preparation for system design, or when migrating large amounts of unstructured data from legacy systems. There are Term Crawling (TC) and Clustering Hierarchy Projection (CHP), which are applied to build a topic map based on free text documents from local repositories and those downloaded from the Internet. The methods originate from data mining techniques for knowledge techniques, has been implemented. The preliminary results that have been achieved on the test collection are presented in concluding sections of the article.

Full Text: