

AUTOMATIC SUBJECT INDEXING : A BOOTSTRAPPING APPROACH THROUGH TEXT MINING OPERATIONS

ANI THOMAS, M. K. KOWAR AND SANJAY SHARMA

Abstract

The authors are motivated by the concept of bootstrapping semantics of a given text to build an automatic subject-index available in electronic format, pertaining to diversified areas of interests and topics. An attempt is to display similar concepts or words uniquely, patched with respective page nos. depicting multiple referencing locations. No data needs to be keyed in on the text editor or front-end interface. The table of contents with chapter heading & subheadings will act as the only input lexicon, no keywords taken from a controlled vocabulary such as a thesaurus or a domain oriented ontology trigger a document in order to describe its subject, that too at the levels within paragraphs, sections and chapters. Finally, the Hopfield net is imbued with a parallel search; its spreading activation process converges the final output in the form of strongly related concepts among the different chapters of the book, so as to formalize a hierarchical subject index at the back-of-the-book.

Keywords: Automatic Subject indexing, Text Mining, Dependency triples, n-grams, Hopfield network.