

Web Search Engine Based Naming Procedure for Independent Topic

Authors : Takahiro Nishigaki, Takashi Onoda

Abstract : In recent years, the number of document data has been increasing since the spread of the Internet. Many methods have been studied for extracting topics from large document data. We proposed Independent Topic Analysis (ITA) to extract topics independent of each other from large document data such as newspaper data. ITA is a method for extracting the independent topics from the document data by using the Independent Component Analysis. The topic represented by ITA is represented by a set of words. However, the set of words is quite different from the topics the user imagines. For example, the top five words with high independence of a topic are as follows. Topic1 = {"score", "game", "lead", "quarter", "rebound"}. This Topic 1 is considered to represent the topic of "SPORTS". This topic name "SPORTS" has to be attached by the user. ITA cannot name topics. Therefore, in this research, we propose a method to obtain topics easy for people to understand by using the web search engine, topics given by the set of words given by independent topic analysis. In particular, we search a set of topical words, and the title of the homepage of the search result is taken as the topic name. And we also use the proposed method for some data and verify its effectiveness.

Keywords : independent topic analysis, topic extraction, topic naming, web search engine

Conference Title : ICWST 2019 : International Conference on Web and Semantic Technology

Conference Location : Rome, Italy

Conference Dates : March 05-06, 2019