Vol. 3, No. 2, Page 87-95

M2VSM: Extension of Vector Space Model for Text Mining by Introducing Meta Keyword

Yasufumi Takama[1], Toru Ishibashi[2], Takashi Okada[2]
[1]Tokyo Metropolitan University, Tokyo, Japan
[2] Tokyo Metropolitan Institute of Technology, Tokyo, Japan

Received Date: 31 March 2009; Accepted Date: 12 June 2009

Abstract
This paper proposes an extended vector space model (VSM), which is called M2VSM (Meta keyword-based Modified VSM). When conventional VSM is applied to document clustering, it is difficult to adjust the granularity of cluster in terms of topic. In order to solve the problem, M2VSM considers meta keywords such as adjectives and adverbs, as additional value of indexing terms. The similarity between documents is calculated by considering the matching of meta keywords for each index term, which makes it possible to cluster documents with various granularities in terms of topic. Experimental results show that clustering results by M2VSM match the results by test subjects in both rough and detailed clustering. A prototype text mining system is implemented based on M2VSM. This paper also shows the result of analyzing management visions of Japanese companies with using the text mining system.