数据库、信号与信息处理

概念与文档的语义相似度计算

宋 玲[1], 郭家义[2], 张冬梅[1], 汤晓兵[1], 高 楠[1]

1.山东建筑大学 计算机科学与技术学院，济南 250101
2.北京市信息资源管理中心，北京 100082

摘要　　将本体作为背景知识引入到概念之间相似度和文档之间相似度的计算中。通过图模型表示本体中概念以及概念之间的语义关系，用来将一个概念和一个文档扩展为一个语义模糊集，并计算模糊集合之间的相似度。文档相似度的计算是在概念相似度计算的基础之上。在概念相似度的计算过程中引入了语义相似度矩阵以及基于共信息理论的模糊相似度方法。
关键词　　概念相似度　文档相似度　本体　文档聚类
分类号

Semantic similarity computation of concepts and documents

SONG Ling[1],GUO Jia-yi[2],ZHANG Dong-mei[1],TANG Xiao-bing[1],GAO Nan[1]

1.School of Computer Science & Technology，Shandong Jianzhu University，Jinan 250101，China
2.Beijing Information Resource Management Center，Beijing 100082，China

Abstract

　A novel method that integrates core ontology as background knowledge into the process of computing similarity of concepts and documents is proposed.Ontology is represented as a graph-based model that reflects semantic relationship between concepts，with which a concept and a document are extended to a semantic fuzzy set.Then fuzzy similarity between two fuzzy sets is computed.Documents comparison is based on concepts comparison.A semantic similarity matrix that exploits semantic relation of the ontology is defined，and fuzzy similarity measure based on shared information content is proposed in the processing of concepts comparison.

Key words　concept similarity　document similarity　ontology　documents clustering

通讯作者　宋 玲　song_ling@sdjzu.edu.cn