

数据库、信号与信息处理

语义相似的PageRank改进算法

何明¹, 周军¹, 李树友²

1. 辽宁工业大学 电子与信息工程学院, 辽宁 锦州 121000

2. 辽宁工业大学 数理科学系, 辽宁 锦州 121000

收稿日期 2008-11-3 修回日期 2009-1-19 网络版发布日期 2009-9-28 接受日期

摘要 PageRank算法是一种用于网页排序的算法, 它利用网页间的相互引用关系评价网页的重要性。但由于它只考虑网页与网页之间的链接结构, 忽略了网页与主题的相关性, 容易造成主题漂移现象。在分析了原PageRank算法基础上, 给出了一种基于语义相似度的PageRank改进算法。该算法能够按照网页结构和网页主要内容计算出网页的PageRank值, 既不会增加算法的时空复杂度, 又极大地减少了“主题漂移”现象, 从而提高查询效率和质量。

关键词 [PageRank算法](#) [主题漂移](#) [语义相似度](#)

分类号 [TP391.3](#)

New semantic similarity PageRank algorithm

HE Ming¹, ZHOU Jun¹, LI Shu-you²

1. College of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, Liaoning 121000, China

2. Department of Mathematical Science, Liaoning University of Technology, Jinzhou, Liaoning 121000, China

Abstract

The PageRank algorithm is used in ranking web pages. It estimates the pages' authority by taking into account the link web structure. However, it only considers the structure of webs' link, without any topic about webs, so this algorithm leads to topic-drift. After analyzing the original PageRank algorithm several times, an improved PageRank algorithm based on semantic similarity is proposed. This new PageRank algorithm can calculate the PageRank's value according to the structure and content about web, it can avoid the topic-drift problem and improve the quality of web search effectively without adding any other extra time and space complexity degree.

Key words [PageRank algorithm](#) [topic-drift](#) [semantic similarity](#)

DOI: 10.3778/j.issn.1002-8331.2009.27.042

通讯作者 何明 heming0405@163.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(571KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [参考文献](#)

服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ 本刊中 包含“[PageRank算法](#)”的
[相关文章](#)

▶ 本文作者相关文章

· [何明](#)

· [周军](#)

· [李树友](#)