

数据库、信息处理

运用有向图进行中文分词研究

张培颖

中国石油大学(华东) 计算机与通信工程学院, 山东 东营 257061

收稿日期 2008-4-30 修回日期 2008-7-23 网络版发布日期 接受日期

摘要 首先说明了分词在中文信息处理中的作用, 然后介绍了分词系统中的关键技术。提出了一种基于有向图的中文分词算法, 该算法首先构造中文分词有向图, 然后计算中文分词有向图中所有可能的切分路径, 最后利用了最少分词原则、汉字之间的互信息和词语的频率等信息给中文分词有向图中的每条切分路径打分, 分数最高的路径就对应正确的切分结果。开放测试结果表明分词精确率可达90%以上。

关键词 [中文分词](#) [有向图](#) [中文分词有向图](#) [切分路径](#) [互信息](#)

分类号

Method of Chinese word segmentation based on directed graph

ZHANG Pei-ying

College of Computer & Communication Engineering, University of Petroleum (East China), Dongying, Shandong 257061, China

Abstract

Chinese word segmentation is the first step for any Chinese information processing and hinders seriously its development. This paper introduces the critical technologies in the segmentation systems. It proposes a refinement of the segmentation algorithm based on the directed graph, this algorithm first constructs the Chinese segmentation directed graph, and calculates the weight of every segmentation path, last evaluates every segmentation path based on the principle of least segmentation, the mutual info of characters and the frequency of words, the highest scores on the path corresponding the correct segmentation results. Open-Test results show that the accuracy rate is more than 90%.

Key words [Chinese segmentation](#) [directed graph](#) [Chinese segmentation directed graph](#) [segmentation path](#) [mutual information](#)

DOI: 10.3778/j.issn.1002-8331.2009.22.040

通讯作者 张培颖 smartfrom1024@yahoo.com.cn

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(357KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“中文分词” 的相关文章](#)
- ▶ [本文作者相关文章](#)
- [张培颖](#)