

数据库、信号与信息处理

基于大规模真实文本的数值知识元挖掘研究

肖洪, 薛德军

中国学术期刊(光盘版)电子杂志社, 北京 100084

收稿日期 2007-11-27 修回日期 2008-2-3 网络版发布日期 2008-10-18 接受日期

摘要 探讨了从海量文献中挖掘知识元的背景, 并详述了从海量年鉴文本中抽取宏观数值知识元的基本流程和各主要环节的算法, 并重点对数值知识元主体抽取的效果进行了分析, 实验结果表明在特定领域内知识元挖掘要达到实用水平是可行的。

关键词 [真实文本](#) [文本挖掘](#) [数值知识元](#) [自动编辑](#)

分类号

Numeric knowledge element mining based on large-scale realistic corpora

XIAO Hong, XUE De-jun

China Academic Journal (CD) Publishing House, Beijing 100084, China

Abstract

This paper discusses the necessity of the knowledge element mining, then describes in detail the base process and algorithms of main steps of extraction numeric knowledge elements from China yearbook full-text database, and then specially analyzes the extraction quality.

Key words [realistic corpora](#) [text mining](#) [numeric knowledge element](#) [automated editing](#)

DOI: 10.3778/j.issn.1002-8331.2008.30.046

通讯作者 肖洪 xiaohong@cnki.net

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(717KB\)](#)
- ▶ [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“真实文本”的相关文章](#)
- ▶ 本文作者相关文章
 - [肖洪](#)
 - [薛德军](#)