

<<文本挖掘原理>>

图书基本信息

书名：<<文本挖掘原理>>

13位ISBN编号：9787030293060

10位ISBN编号：7030293061

出版时间：2010-11

出版时间：科学出版社

作者：程显毅,朱倩

页数：228

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<文本挖掘原理>>

内容概要

在信息时代，存储大量数据比较容易。

通过Web、企业内部网、电传新。

闻获得的文本数量在急剧增加，这导致信息过载。

然而，数据量虽然增加了，但可用的信息却在减少。

文本挖掘是一个新的令人振奋的研究领域，其试图通过综合数据挖掘、机器学习、自然语言处理、信息检索和知识管理等技术来解决信息过载问题。

文本挖掘包括文本集合的预处理(文本分类、信息抽取)、中间结果存储、中间结果分析技术(分布分析、聚类、趋势分析、关联规则抽取)和最终结果的可视化。

它与关联分析类似，为人们提供了分析海量文本数据的新工具，并且通过学习模式来指导抽取实体关系。

本书首先讨论了文本挖掘的总体结构以及文本挖掘预处理算法，然后深入地研究了文本挖掘核心操作，最后探讨真实世界中文本挖掘的主要应用和DIAL，弥补了理论和实践的脱节。

本书主要可供对文本挖掘感兴趣的本科高年级学生、研究生、研究人员和专业开发人员参考，对从事文本挖掘开发和使用文本挖掘系统的人也会有很大帮助。

<<文本挖掘原理>>

书籍目录

前言第1章 文本挖掘概述 1.1 文本挖掘的产生背景 1.2 文本 1.2.1 文本格式 1.2.2 动态文本集 1.3 文本挖掘的概念 1.4 文本挖掘的任务 1.4.1 文本挖掘预处理 1.4.2 文本模式挖掘 1.4.3 挖掘结果可视化 1.5 文本挖掘系统的通用体系结构第2章 文本表示第3章 文本挖掘预处理——文本分类第4章 文本挖掘预处理——文本聚类第5章 文本挖掘核心操作——信息抽取第6章 文本挖掘核心操作——关系抽取第7章 文本挖掘核心操作——关联分析第8章 文本挖掘结果的可视化第9章 文本挖掘的应用第10章 专门用于文本挖掘的信息抽取语言——DIAL参考文献

<<文本挖掘原理>>

编辑推荐

大量文档集内容的预处理包括特征抽取、文本分类、文本聚类等。

文本分类和文本聚类研究的成果比较多,《文本挖掘原理》只作简单介绍;而对特征生成、特征选择和特征抽取,《文本挖掘原理》进行深入地讨论。

信息抽取是文本挖掘的核心操作,目前的主要研究方向是命名实体识别、实体关系抽取和事件抽取。实体关系抽取是文本挖掘的关键任务,所以《文本挖掘原理》用大量篇幅讨论了实体关系发现技术。

由于文本挖掘强调用户交互到知识发现过程的集成性,因此,《文本挖掘原理》最后讨论了文本挖掘结果可视化方法,并介绍了一种文本挖掘语言——DIAL。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>