<<信息检索>>

图书基本信息

书名：<<信息检索>>

13位ISBN编号：9787115212252

10位ISBN编号：7115212252

出版时间：2009-10

出版时间：人民邮电出版社

作者：（美）格罗斯曼，（美）弗里德　著

页数：332

<<信息检索>>

内容概要

本书是“信息检索”课程的优秀教材，书中对信息检索的概念、原理和算法进行了详细介绍，内容主要包括检索策略、检索实用工具、跨语言信息检索、查询处理、集成结构化及数据和文本、并行信息检索以及分布式信息检索等，并给出了阐述算法的大量实例。

本书有一定的深度和广度，而且所有的内容都用当前的技术阐述，是高等院校计算机及信息管理等相关专业本科生和研究生的理想教材，对信息检索领域的科研和技术人员也是很好的参考书。

作者简介

格罗斯曼（David A.Grossman），佐治亚梅森大学博士。
现在伊利诺伊理工大学计算机系任教。
曾在美国政府部门高级技术服务中心和研究发展办公室担任项目经理。
主要研究领域包括信息检索、结构化与非结构化数据集成以及数据挖掘。

书籍目录

章节摘录

3.4.1 DAmore and Mah Initial information retrieval research focused on n-grams as presented in[DAmore and Mah, 1985]. The motivation behind their work was the fact thatit is difficult to develop mathematical models for terms since the potential fora term that has not been seen before is infinite. With n-grams, only a fixednumber of n-grams can exist for a given value of n. A mathematical modelwas developed to estimate the noise in indexing and to determine appropriatedocument similarity measures. DAmore and Mahs method replaces terms with n-grams in the vector spacemodel. The only remaining issue is computing the weights for each n-gram.Instead of simply using n-gram frequencies, a scaling method is used to nor-malize the length of the document. DAmore and Mahs contention was that alarge document contains more n-grams than a small document, so it should bescaled based on its length. To compute the weights for a given n-gram, DAmore and Mah estimatedthe number of occurrences of an n-gram in a document. The first simplifyingassumption was that n-grams occur with equal likelihood and follow a binomialdistribution. Hence, it was no more likely for n-gram "ABC" to occur than"DEE" The Zipfian distribution that is widely accepted for terms is not true forn-grams. DAmore and Mah noted that n-grams are not equally likely to occur,but the removal of frequently occurring terms from the document collectionresulted in n-grams that follow a more binomial distribution than the terms. DAmore and Mah computed the expected number of occurrences of an n-gram in a particular document. This is the product of the number of n-gramsin the document （the document length） and the probability that the n-gramoccurs. The n-grams probability of occurrence is computed as the ratio ofits number of occurrences to the total number of n-grams in the document.DAmore and Mah continued their application of the binomial distribution toderive an expected variance and, subsequently。

媒体关注与评论

“本书涉及最新的研究成果，语言经得起推敲，还精心准备了大量的实例说明，适合作为研究生和本科生信息检索课程的首选教材。”　　——美国马萨诸塞大学阿默斯特校区计算机系杰出教授　W.Bruce Croft　　“推荐把本书作为计算机科学专业学生的首选教材，同时也适用于SE0专业人员和Web开发者阅读，将搜索技术，算法和启发式方法运用于他们的项目中。”　　——信息技术与服务顾问　E.Garcia博士

<<信息检索>>

编辑推荐

随着Google、百度等搜索引擎公司的崛起，信息检索已经成为令人振奋的热门研究领域。

《信息检索：算法与启发式方法(英文版·第2版)》从发展的角度描述了ad hoc信息检索，讨论了用来实现大规模数据检索的最新算法，详细介绍了推理网络和系统的效率，并且对每种方法都给出了详细可行的实例。

此外，《信息检索：算法与启发式方法(英文版·第2版)》整合了结构化和非结构化数据的处理技术，这是其他教材所不具备的。

第2版新增加了IR语言模型和跨语言检索，还讨论了许多当前的热点话题，如XML、P2P信息检索、文本查重、文档并行聚类、不同检索策略的融合、信息中间表示等。

《信息检索：算法与启发式方法(英文版·第2版)》兼顾了学科广度和主题深度，把握了最新的发展趋势，是信息检索领域的一本名著，更为许多著名高校（如美国普林斯顿大学、罗格斯大学）采用为教材。

随着Google、百度等搜索引擎公司的崛起，信息检索已经成为令人振奋的热门研究领域。

《信息检索：算法与启发式方法(英文版·第2版)》从发展的角度描述了ad hoc信息检索，讨论了用来实现大规模数据检索的最新算法，详细介绍了推理网络和系统的效率，并且对每种方法都给出了详细可行的实例。

此外，《信息检索：算法与启发式方法(英文版·第2版)》整合了结构化和非结构化数据的处理技术。这是其他教材所不具备的。

第2版新增加了IR语言模型和跨语言检索，还讨论了许多当前的热点话题，如XML、P2P信息检索、文本查重、文档并行聚类、不同检索策略的融合、信息中间表示等。

《信息检索：算法与启发式方法(英文版·第2版)》兼顾了学科广度和主题深度，把握了最新的发展趋势，是信息检索领域的一本名著，更为许多著名高校（如美国普林斯顿大学、罗格斯大学）采用为教材。

版权说明