<<Web搜索>>

图书基本信息

书名：<<Web搜索>>

13位ISBN编号：9787040278170

10位ISBN编号：7040278170

出版时间：2009-8

出版时间：高等教育出版社

作者：郭军 编

页数：294

前言

当今时代，如何从源源不断、无边无际的海量Web数据中搜索信息已经成为一个对社会的政治、经济、文化、安全等具有全方位影响的重大课题。
在这一背景下，以信息检索、过滤和推荐为主要内容的Web搜索引起了全球学术界、产业界以及各国政府的极大关注。
商用搜索引擎巨头迅速崛起，强有力地带动了社会经济的发展。
相关的学术研究异常活跃，为自然科学和社会科学的多个领域的研究注入了活力。

Web搜索是一种高度智能化的信息处理技术。
在目前已经形成的理论和技术体系中，融合了模式识别、自然语言处理、机器学习、数据挖掘等多个学科的成果，综合性和交叉性十分突出。
此外，海量信息处理、Web网页自动获取及分析、网页索引、网页链接分析、社会网络挖掘等内容更是具有独特性和新颖性。
这门技术也因此走入了大学的课堂，并迅速受到了广大学生的青睐。
目前，国内外IT背景较强的大学都至少在研究生层次上开设了相关的课程。

相对于这种旺盛的教学需求，Web搜索的教材建设明显滞后，特别是中文教材非常稀缺。
即使是外文教材也在系统性和前沿性等方面落后于技术的发展现状。
因此，编写出版紧跟最新技术进展的Web搜索的大学教材有十分紧迫的需求。

作者长期从事模式识别和网络技术领域的研究和教学工作，近年来对Web搜索产生了浓厚的兴趣，带领一支十多人的教师团队指导上百名研究生对该领域进行了多方面的深入研究。
通过研究工作的不断积累，对Web搜索的技术体系和主要内涵有了比较深刻的认识和理解，感到值得将其梳理和总结为一部主要面向研究生教学的教材，为解当前的燃眉之急贡献一份力量。

本书将Web上的信息检索、过滤和推荐等技术定义为Web搜索，使其具有比较宽泛的内涵。
这样做的好处是将Web检索、过滤、推荐等既联系紧密又相互区分的技术统一在一个体系中，便于进行系统地学习和研究。
这是本书的一个显著特色。

本书紧跟技术的最新进展，讨论和介绍重要的研究成果，以及不断涌现的挑战。
在写法上以Web搜索所包含的主要任务和核心问题为纲、以典型理论模型为例介绍研究的进展，分析对比不同方法在不同方面的优劣，并着力指出它们的局限、当前的研究重点和发展趋势。
这一点与通常的教材一般只对成熟的理论进行系统总结相比有很大的不同。

<<Web搜索>>

内容概要

《Web搜索》内容包括导论、文本检索、图像检索、音频检索、信息过滤、信息推荐以及发展前沿。
对Web搜索的基本概念进行定义，阐述其科学价值和研究状况，根据Web搜索所涵盖的检索、过滤以及推荐技术，论述其中的核心问题、基本概念和基本方法，并介绍Web搜索若干新的研究方向。
《Web搜索》的最大特点是将Web上的信息检索、过滤和推荐等技术定义为Web搜索，使其具有比较宽泛的内涵。
将Web检索、过滤和推荐统一在一个体系中，既符合这三项技术发展的现状和趋势，又便于读者进行系统的学习和研究。
另外，《Web搜索》紧跟近年来的最新研究进展，具有显著的先进性和独特性。
《Web搜索》可以作为信息、通信、计算机类研究生或高年级本科生的教材和教学参考书，也可作为专业技术人员的阅读和培训资料。

作者简介

郭军，教授，现任北京邮电大学信息与通信工程学院院长，日本东北学院大学博士、博士生导师。
主要学术兼职包括国家自然科学基金委员会信息科学部学科评审组成员、北京市科学技术奖评审专家组成员、北京市计算机与控制学科高级职称评审组副组长、中国人工智能学会理事、中国自动识别协会专家组成员、中文信息处理学会理事等。
主要社会兼职包括北京市政协委员、北京市高级知识分子联谊会理事、中共中央统战部信息员等。
现主要从事Web搜索、模式识别、网络管理等方面的研究。
在SCIENCE、IEEE Trans. on PAMI、IEICE Trans、ICPR、IOOV、SIGIR等模式识别、计算机视觉以及信息检索领域国际顶级刊物和会议上发表了多篇论文。
出版著作6部，其中《网络管理》一书被评为首批（2004年）北京市精品教材。

<<Web搜索>>

书籍目录

章节摘录

Web搜索广阔的应用领域、巨大的社会经济作用以及高度的技术挑战性使其充满了科学研究价值。

第一，Web搜索所研究的是一个崭新的科学问题，即如何在无边的动态的Web信息中寻找最符合用户需求的信息。

这个问题不仅在尺度上空前巨大，而且约束条件非常不确定。

因为系统通常难以了解用户真正的信息需求。

用户总是希望以最简单的提问或最便捷的操作，如输入少量关键字的方式来表达自己的请求，因而系统得到的指示是十分笼统和模糊的。

我们应该认识到，Web搜索在计算规模和约束的不确定性方面已经将人类的科学研究带到了一个新高度。

第二，Web搜索既要考虑信息的客观性，又要考虑信息的主观性。

所谓信息的客观性，是指信息的数据形式在Web中是客观存在的，不论面对哪个主体（用户），承载信息的数据都是相同的。

而信息的主观性是指同样的数据给用户提供的信息（量）是不同的。

一篇介绍摄影常识的文章对初学者来说可能“很有信息量”，而对一个摄影师来说信息量几乎为零。

在Web搜索中，上述客观性因素和主观性因素都会影响搜索结果的正确性（质量）。

这种特点在普通的自然科学研究中是很少见的，因此引起了人们更大的研究兴趣。

第三，Web搜索强有力地带动了相关学科，特别是智能学科的发展。

智能学科中的自然语言理解、模式识别、机器学习、数据挖掘等在Web搜索中找到了巨大的发展空间，近年来已经形成了空前高涨的研究热潮。

例如文本分类、多媒体识别、海量数据挖掘、在线增量机器学习、在线分类和聚类、信息抽取、信息摘要、命名实体识别等研究都紧密地与Web搜索联系了起来。

商用搜索引擎的智能化趋势也正是在这些研究的基础上形成的。

甚至可以预期Web搜索将成为一个大面积涵盖智能学科的新兴独立学科。

媒体关注与评论

本书最主要的特色是将信息“检索”、“过滤”和“推荐”一并考虑，具有前瞻性意义；另外一个特色是在讨论搜索的时候，不仅考虑了传统上为主的文本，也一并考虑了当前越来越重要的图像视频和语音的检索问题，很好地结合了作者的研究工作领域特长。

——李晓明

随着网络技术的发展和普及，Web搜索技术也变得越来越重要。现在的互联网上，信息——包括文本、图像、视频和语音信息——可以说已经非常丰富，关键是让用户方便快捷地找到这些信息。这正是本书所探讨的主要内容。

——马少平

版权说明