

<<个性化搜索引擎原理与技术>>

图书基本信息

书名：<<个性化搜索引擎原理与技术>>

13位ISBN编号：9787030222558

10位ISBN编号：7030222555

出版时间：2008-6

出版时间：科学出版社

作者：李树青，韩忠愿 编著

页数：125

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>

<<个性化搜索引擎原理与技术>>

前言

快速发展的现代互联网在带给人们大量信息的同时，也不可避免地产生了难以让用户快速获取有效信息的问题。

作为一种常见的Web信息资源检索工具，搜索引擎日益受到人们的关注并得到广泛的使用。

它面向任何Web用户，无需用户具有较高的专业检索知识，使用方式也较为简单。

搜索引擎已经成为人们获取Web资源的一种主要方式。

然而，现代搜索引擎也存在很多不足。

其中，最为主要的一个问题就是由于采用了全文检索的匹配方法，用户往往会得到相当多的查询结果网页，而用户一般只会访问其中感兴趣的网页，但是很多搜索引擎往往缺乏对用户个性化信息的利用，从而不能实现有针对性的个性化信息服务。

实际的情况就表现为即使是具有不同个性化信息需求的用户，在输入相同检索词语的时候也会得到相同的结果，甚至是相同的网页排列次序。

这些问题显然需要得到解决。

借鉴在电子商务网站中广泛使用的Web个性化信息推荐技术，本书提出了一个较为可行的解决方案，即在搜索引擎中使用Web个性化信息推荐技术，以实现个性化搜索引擎。

然而，传统的Web个性化信息推荐技术具有很多并不适合搜索引擎的特点。

只有结合搜索引擎工作的原理和特点，在现有的Web个性化信息推荐技术基础上加以改进，才能设计出具有个性化信息推荐能力的智能搜索引擎。

通过对基于个性化信息推荐技术的搜索引擎框架和基本技术的探讨，本书构建了一个完整的基于Web个性化信息推荐技术的搜索引擎框架结构。

这种框架结构的设计思想主要考虑了两点内容：一是尽量减少用户使用的复杂度，能够让用户在完全无需关注个性化过程的情况下，来表达自己的个性化信息需求和得到所需的个性化信息；二是尽量在现有搜索引擎技术基础上进行优化和完善，无需对现有技术和平台环境做过大的调整。

梁希侠内蒙古大学教授、博士生导师，国家突出贡献专家，全国教育系统劳动模范，享受国务院政府特殊津贴。

曾任教育部高等学校物理类专业教学指导分委员会委员，现任全国热力学与统计物理教学研究会主任。

主持国家精品课程“统计热力学”，获国家级教学名师奖，国家教委和内蒙古自治区科技进步奖、内蒙古自治区教学成果奖。

班士良理学博士，内蒙古大学教授，博士生导师，享受国务院政府特殊津贴，内蒙古自治区有突出贡献专家，劳动模范，教育部高等学校物理学类专业教学指导分委员会委员、中国物理学会理事、内蒙古自治区物理学会理事长、全国热力学与统计物理教学研究会秘书长。

曾获内蒙古自治区科技进步奖、教学成果奖、教学名师奖。

<<个性化搜索引擎原理与技术>>

内容概要

本书通过对基于个性化信息推荐技术的搜索引擎服务器日志中所具有关键词序列得到用户模式，并按照事务模式聚类的方法实现用户个性化特征的表达，最后在搜索引擎的网页索引中，利用得到的用户个性化特征改进传统的PageRank算法。

通过上述工作，本书构建了一个完整的基于Web个性化信息推荐技术的搜索引擎框架结构。

本书可作为计算机专业的本科生和研究生的参考用书，也可供有关技术人员参考。

<<个性化搜索引擎原理与技术>>

书籍目录

前言第1章 绪论 1.1 Web信息检索 1.1.1 Web信息检索模型 1.1.2 向量空间模型 1.1.3 搜索引擎
1.1.4 搜索引擎工作原理 1.1.5 相关度排序技术 1.2 Web挖掘 1.2.1 Web挖掘的概念 1.2.2 Web挖掘的
类型 1.2.3 Web挖掘的研究进展 1.3 web个性化信息推荐服务 1.3.1 概念 1.3.2 Web个性化信息推荐
服务的种类划分 1.3.3 Web个性化信息推荐服务的发展第2章 网页权重分析技术 2.1 网页质量分析技
术 2.1.1 结合网页质量分析的Web信息检索模式 2.1.2 网页质量测度方法 2.2 网页相关度分析技术
2.2.1 标准PageRank算法 2.2.2 PageRank的优化计算第3章 个性化搜索引擎 3.1 概念与特点 3.1.1 现代
搜索引擎系统存在的问题 3.1.2 个性化搜索引擎的含义 3.1.3 现阶段个性化搜索引擎的不足 3.2 基本
类型 3.2.1 基于个性化信息采集的个性化搜索引擎 3.2.2 基于查询改进的个性化搜索引擎 3.2.3 基于
个性化网页权重的个性化搜索引擎第4章 用户个性化模式的获取和表达 4.1 基于关键词序列的用户模
式识别 4.1.1 概述 4.1.2 数据准备 4.1.3 用户识别 4.1.4 事务模式识别 4.2 基于用户事务模式聚类
的Web信息个性化表达 4.2.1 用户事务模式的类别构造 4.2.2 基于频繁路径的用户事务模式类别构造
4.2.3 降维处理问题第5章 基于关键词序列的个性化网页权重方法 5.1 方法概述 5.2 基于修改网页权重
值的个性化PageRank 5.3 基于添加修正参数的个性化PageRank 5.3.1 使用事务聚类模式的个性
化PageRank方法 5.3.2 使用主题化事务聚类模式的个性化PageRank方法第6章 系统原型的实现 6.1 系
统的开发方式 6.2 数据结构 6.3 存储过程 6.4 系统的功能模块 6.4.1 爬虫模块 6.4.2 Web网页分析模
块 6.4.3 日志分析模块 6.4.4 用户接口模块 6.5 结果分析 6.5.1 关键词访问序列的获取情况 6.5.2 用
户事务模式的获取情况 6.5.3 用户事务模式的聚类情况 6.5.4 个性化PageRank值的计算情况 6.6 系统
框架评价参考文献

<<个性化搜索引擎原理与技术>>

章节摘录

插图：第1章 绪论本章主要介绍了与个性化搜索引擎相关的基本技术内容，主要包括Web信息检索、Web挖掘和Web个性化信息推荐服务等。

1.1 Web信息检索随着互联网技术的快速发展，网络信息资源呈现一种爆炸式的增长态势，比如在2005年，Google搜索引擎能够遍历到的Web网页数量就已经达到近60亿。

这些网络信息资源在给人们带来丰富知识和极大便利的同时，也暴露出一些亟待解决的问题。

其中，最主要的问题表现在这种信息资源的增长速度远远超出了人们能够处理它们的能力，动辄千万吉的信息量让用户实际上难以获取所需的有效信息，更难以对收集来的海量信息进行分析和获取知识。

奈斯比特在《大趋势》一书中准确形容了人们目前所处的困境，即信息是丰富的，而知识是贫乏的。针对上述特点，人们在传统信息检索系统的基础上开发出新的Web信息检索系统，典型的系统就是搜索引擎。

自从1994年问世以来，搜索引擎逐渐成为人们获取网络信息资源的主要方式，相关搜索引擎网站也是Web用户使用网络时首选的访问站点。

另外，它和免费电子邮箱、网络实时通信软件构成了当今门户网站用来吸引用户访问的三种主要方式。

现在，相关搜索引擎厂商日益成为促进互联网产业发展的重要力量。

1.1.1 Web信息检索模型广义的信息检索是指信息用户为处理和解决各种问题而查找、识别、获取相关的事实、数据、文献的活动及过程，而狭义的信息检索主要是指信息用户在计算机信息检索系统上进行的信息查询行为。

具体的计算机检索行为包含脱机批处理检索、联机检索、光盘检索和网络化联机检索。

Web信息检索是一种网络化的联机检索，它的检索对象就是互联网上的Web资源。

<<个性化搜索引擎原理与技术>>

编辑推荐

《个性化搜索引擎原理与技术》可作为计算机专业的本科生和研究生的参考用书，也可供有关技术人员参考。

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>