

<<网络信息分类>>

图书基本信息

书名：<<网络信息分类>>

13位ISBN编号：9787030237637

10位ISBN编号：7030237633

出版时间：2008-12

出版时间：科学出版社

作者：施国良

页数：152

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>

前言

随着网络信息的迅猛发展，庞大的网络信息资源和人们特定的信息需求之间形成了巨大的矛盾。在此背景下，本书专门讨论了分面分类法在网络信息组织中应用的理论与实践问题。信息量的迅速增长与人们对信息的特定需求是人们学习、工作和生活中的一对矛盾。只有研究网络信息资源的现状与特征，找到网络信息资源组织的规律，才能从根本上解决这一矛盾，为人们方便、快捷地利用网络信息提供一个理想的途径。

正是在这样的背景下，本书首先从搜索引擎的不足之处出发，对分类法、分面分类法、网络信息组织、数据库等相关领域进行文献调查；其次运用实验方法对实物、服务和网络文献三个类型的网络信息进行分面分析；再次选择其中的实物类型进行概念设计、逻辑设计和物理设计；最后建立一个简单的分面分类网络应用模型，并将这个模型初步应用于上海宝山钢铁股份有限公司（以下简称宝钢公司），旨在建立一个基于工艺和流程的企业知识组织系统。

这个系统的底层正是分面数据库。

设计系统模型参数为：开源服务器Apache 2.0.50，开源数据库5.0，服务器端脚本为PHP 5.2.4，浏览器端使用AJAX技术实现无刷新更新数据功能，从而基本上适应了分面分类法的应用原理。

本书的观点是：搜索引擎的局限性决定了仅仅依靠特性检索方式无法解决用户个性化的信息需求问题；枚举式分类法用于指导实物排架的功能特性决定了其无法根本解决网络信息激增所带来的网络信息组织问题；分面分类法并不能简单地等同于阮冈纳赞的冒号分类法，网络上应用分面分类法必须遵循严谨的分面分析过程；网络既为分面分类法的复苏提供了广阔的应用舞台，也为分面分类法再次兴起提供了强大的技术支持。

本书的结构为：第1章，主要从网络信息资源的现状出发，分析目前网络信息组织与检索存在的问题，从而引出全书的研究问题；第2章，主要对目前网络信息组织尤其是网络信息分类的理论与实践进行评述，让读者对与本书有关的研究有一个简要的了解；第3章，主要在理论上探讨用分面分类法对网络信息进行组织的必要性；第4章，主要在理论上探讨用分面分类法对网络信息进行组织的可能性；第5~7章，主要通过实验的方法将分面分类法具体应用于网络信息组织；第8章，作为案例，对宝钢公司做了基于业务和流程的分面分类系统模型的介绍；第9章，主要讨论分面分类法网络应用的一些拓展问题；第10章，主要是概括全书的结论，并粗略地勾勒了未来的一些研究方向。

<<网络信息分类>>

内容概要

随着网络信息的迅猛发展，庞大的网络信息资源和人们特定的信息需求之间形成了巨大的矛盾。在此背景下，本书专门讨论了分面分类法在网络信息组织中应用的理论与实践问题。

首先，本书分析了网络信息组织面临的问题；其次，专门论述了分面分类法的原理、特征和独特的性能及其用于网络信息组织的长处、必要性和可能性；再次，详细讨论了分面分类法应用于网络信息组织的基本技术特点、过程和方法，并结合实例说明其可操作性，使读者既对网络信息组织的特殊性有一个完整的了解，又对分面分类法的具体应用有了感性的认识；最后，将理论、技术与实践相结合，用实验的方法将分面分类法应用于网络信息组织，并结合一个案例做了具体的说明。

本书集原理和应用于一体，语言深入浅出，通俗易懂，并配有必要的图表，具有较强的可读性。

本书适合作为高等院校图书情报类和信息管理类专业各层次学生的教学参考书和补充读物，也可作为各类信息资源管理部门（包括政府部门和科研机构）工作人员、众多网络公司工作人员以及广大计算机与网络爱好者的参考书。

书籍目录

前言第1章 绪论 1.1 网络信息资源的现状与特征 1.2 网络信息资源组织的问题 1.3 搜索引擎的现状与问题 1.4 本书的框架结构与写作目的第2章 网络信息分类研究进展评述 2.1 关于分面分类研究 2.2 关于网络信息组织的研究 2.3 与网络信息分类有关的技术问题的研究第3章 网络信息分类的基本原理 3.1 分类问题的起源 3.2 从人类基本的思维方式看族性检索的重要性 3.3 对分类过程、性质与目的的探讨有助于认识分类的基本原理 3.4 对分类法种类的探讨有助于认识网络信息分类的基本原理 3.5 枚举式分类法对网络的尝试 3.6 网络信息分类的特殊性 3.7 小结第4章 分面分类法与网络信息组织 4.1 分面分类法的特征 4.2 网络信息组织应用分面分类法的标准 4.3 网络信息分面分类法的编制 4.4 小结第5章 网络分面分类系统概念模型设计 5.1 概述 5.2 实物类 5.3 服务类 5.4 电子文献类第6章 网络分面分类系统逻辑模型设计 6.1 逻辑模型概述 6.2 逻辑模型的设计第7章 网络分面分类系统物理模型设计 7.1 概述 7.2 数据结构的设计 7.3 访问方法的设计 7.4 数据存放位置的设计 7.5 系统配置的设计 7.6 用户界面设计第8章 案例研究——宝钢公司分面分类系统设计 8.1 案例背景 8.2 分面本体的设计 8.3 宝钢公司信息知识系统的物理设计与开发 8.4 小结第9章 网络信息分类的拓展 9.1 分面分类法网络应用中的两个问题 9.2 数据库模型的选择 9.3 用户界面设计问题第10章 结束语 10.1 研究结论 10.2 分类研究的局限性 10.3 未来的研究方向主要参考文献附录A 缩略语与全称对照表附录B 部分XML数据库代码 B1 DTD模式 B2 XSD1 B3 数据表(部分)

章节摘录

互联网出现后，搜索引擎是一种重要的网络信息检索工具，它让用户键入关键词，然后到数据库中去匹配。

不同于学科目录，搜索引擎不用人工索引员来编制索引，而是通过软件来自动生成包含网页的数据库。

概括说来，搜索引擎的工作原理包括三个部分：一是用某种程序，如“网络爬虫”（crawler），来收集众多的网页内容；二是以某种利于高效检索的方式（如标引）组织这些网页，形成数据库；三是接受查询，并用某种排序软件进行排序，并输出结果。

第一，爬虫负责跟踪网络，以广度优先或深度优先的方法从Web上下载页面，按照链接从一个网站到另一个网站。

不同的搜索引擎有不同的爬虫，有的遍历所有网站，有的则根据自己的标准选择一些流行的网站遍历。

前者返回的结果量很大，而后者会返回更加相关的结果，速度也更快。

第二，爬虫返回的每一个页面都存放在一个数据库中，对下载页面的内容进行分析以用于索引，具体包括分词、过滤、转换等工作；然后将文档表示为一种便于检索的方式并存储在索引数据库中，一般采用的方法有矢量空间模型、倒排文档、概率模型等；并通过自动抽词和字顺排列编制好索引。

索引是每一个有效词的列表，并有一个相应的指针指向它在数据库中的位置。

第三，实现用户查询关键词和目标文档匹配度的计算，根据计算结果所有符合查询要求的页面URL按照相关度递减的顺序排列，并返回给用户；用户接口为用户提供一个输入查询请求，定制查询结果的Web页面并将查询结果格式化后返回给浏览器。

同样，不同的搜索引擎会遵循不同的原则：有的标引爬虫返回页面中每一个单个的词，有的只标引标题或短语。

搜索引擎的第三个要素是搜索软件。

该软件将用户键入的提问关键词与索引进行比较，发现匹配的结果并按照相关度进行排序。

相关度排序标准依不同的搜索引擎而不同。

不同的搜索引擎所使用的爬虫和标引方法会导致不同的结果。

这就是为什么在不同的搜索引擎中键入相同的关键词会得到不同的结果。

而且，由于这样的操作方式，使得搜索引擎会返回较分类目录更全面更专业的结果。

这是因为它们搜索整个网页而非网站的顶层页面，使用的索引也是自动生成的，而非向分类目录那样使用预先生成的索引。

编辑推荐

《网络信息分类：原理与应用》的第1章，主要从网络信息资源的现状出发，分析目前网络信息组织与检索存在的问题，从而引出全书的研究问题；第2章，主要对目前网络信息组织尤其是网络信息分类的理论与实践进行评述，让读者对与《网络信息分类：原理与应用》有关的研究有一个简要的了解；第3章，主要在理论上探讨用分面分类法对网络信息进行组织的必要性；第4章，主要在理论上探讨用分面分类法对网络信息进行组织的可能性；第5~7章，主要通过实验的方法将分面分类法具体应用于网络信息组织；第8章，作为案例，对宝钢公司做了基于业务和流程的分面分类系统模型的介绍；第9章，主要讨论分面分类法网络应用的一些拓展问题；第10章，主要是概括全书的结论，并粗略地勾勒了未来的一些研究方向。

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>