

<<搜索引擎原理与实践>>

图书基本信息

书名：<<搜索引擎原理与实践>>

13位ISBN编号：9787563518616

10位ISBN编号：7563518614

出版时间：2008-11

出版单位：北京邮电大学出版社有限公司

作者：袁津生，蔡岳 编著

页数：327

字数：529000

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

网络的发展彻底改变了人们的思维、习惯与生活。

一方面，它使我们更容易获取各种各样的信息，而另一方面，要想在数十亿网页的网络信息中精确地找到自己需要的信息简直就如“大海捞针”一般。

那么在巨大的网络信息世界里，怎么样才能找到我们需要的数据呢？

这就要靠搜索引擎。

面对浩如烟海的网络资源，搜索引擎就好像是航船的指南针，引领着人们在网络中冲浪。

目前，搜索引擎已经成为信息检索最有效的工具。

据统计，搜索引擎已经成为仅次于电子邮件的第二大网络应用服务，是用户获取信息的首要途径，在美国有超过84%的网民经常使用搜索引擎。

## <<搜索引擎原理与实践>>

### 内容概要

随着搜索引擎技术的发展和不断完善，越来越多的人开始对搜索引擎原理和技术进行研究，越来越多的人喜欢上了搜索引擎。

本书从教学的角度出发，全面地阐述了搜索引擎的原理和实践，包括搜索引擎的基本原理与技术、搜索引擎的数据结构、搜索引擎的爬虫、多媒体信息检索技术以及搜索引擎开发技术。

本书适合高等院校计算机科学与技术专业及相关专业的高年级学生和研究生阅读参考，也适合相关领域的工程技术人员参阅。

## 书籍目录

第1章 搜索引擎概述 1.1 搜索引擎的概念 1.2 搜索引擎的历史 1.3 搜索引擎的分类 1.4 搜索引擎的关键技术 1.5 当代主要搜索引擎介绍 1.5.1 谷歌搜索 1.5.2 雅虎搜索 1.5.3 百度搜索 1.5.4 北大天网搜索 1.6 搜索引擎的发展 1.7 小结 思考题第2章 搜索引擎的体系结构和工作原理 2.1 搜索引擎的体系结构 2.1.1 搜索器 2.1.2 索引器 2.1.3 检索器 2.1.4 用户接口 2.2 搜索引擎的工作原理 2.2.1 网页搜集 2.2.2 网页处理 2.2.3 查询服务 2.3 元搜索引擎 2.3.1 元搜索引擎的基本构成 2.3.2 元搜索引擎的分类 2.3.3 常用元搜索引擎介绍 2.3.4 元搜索引擎的特点 2.3.5 主要技术指标 2.4 小结 思考题第3章 信息处理技术 3.1 检索模型 3.1.1 经典模型 3.1.2 代数模型 3.2 文本处理 3.2.1 词法分析 3.2.2 分词技术 3.2.3 无用词汇删除 3.2.4 词干提取 3.2.5 索引词选择 3.2.6 词典 3.3 文本压缩 3.3.1 基本概念 3.3.2 统计方法 3.3.3 字典方法 3.3.4 倒排文档压缩 3.4 Web信息处理 3.4.1 Web信息的特点 3.4.2 Web信息的表现方式 3.4.3 Web信息系统结构 3.5 小结 思考题第4章 信息检索技术 4.1 顺排检索 4.1.1 表展开法 4.1.2 逻辑树展开法 4.1.3 BF算法 4.1.4 KMP算法 4.1.5 BM算法 4.2 倒排检索 4.2.1 倒排检索 4.2.2 倒排文档 4.2.3 逆波兰表达式.....第5章 信息检索评价第6章 网络搜索引擎技术第7章 多媒体检索概述第8章 基于内容的多媒体信息检索技术第9章 搜索引擎开发技术参考文献

章节摘录

第1章 搜索引擎概述在浩瀚的网络资源中，搜索引擎（Search Engine）是一种网上信息检索工具，它能帮助用户迅速而全面地找到所需要的信息。

我们可以这样对搜索引擎进行定义：搜索引擎是一种能够通过因特网接受用户的查询指令，并向用户提供符合其查询要求的信息资源网址的系统。

多数网上用户使用搜索引擎来获得所需信息，据CNNIC的统计，用搜索引擎搜索仅次于电子邮件的应用。

目前网上比较有影响的中文搜索工具有：Google、百度（Baidu）、北大天网、爱问（iask）、雅虎（Yahoo）、搜狗（Sogou）等搜索引擎。

英文的有：Yahoo、AltaVista、Excite、Infoseek、Lycos、Aol等。

另外还有专用搜索引擎，例如，专门搜索歌曲和音乐的；专门搜索电子邮件地址、电话与地址及公众信息的；专门搜索各种文件的FTP搜索引擎等。

本章主要介绍搜索引擎的概念、搜索引擎的发展史、搜索引擎的分类以及一些著名的搜索引擎。

1.1 搜索引擎的概念搜索引擎是指根据一定的策略、运用特定的计算机程序搜集互联网上的信息，在对信息进行组织和处理后，为用户提供检索服务的系统。

搜索引擎并不真正搜索互联网，它搜索的实际上是预先整理好的网页索引数据库。

真正意义上的搜索引擎，通常指的是收集了互联网上几千万到几十亿个网页并对网页中的每一个词（即关键词）进行索引，建立索引数据库的全文搜索引擎。

当用户查找某个关键词的时候，所有在页面内容中包含了该关键词的网页都将作为搜索结果被搜出来。

在经过复杂的算法进行排序后，这些结果将按照与搜索关键词的相关度高低，依次排列。

<<搜索引擎原理与实践>>

编辑推荐

《搜索引擎原理与实践》由北京邮电大学出版社出版。

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>