

《搜索引擎》

图书基本信息

书名：《搜索引擎》

13位ISBN编号：9787030146335

10位ISBN编号：7030146336

出版时间：2005-4

出版社：科学出版社发行部

作者：李晓明

页数：247

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu111.com

《搜索引擎》

内容概要

《搜索引擎:原理技术与系统》系统地介绍了互联网搜索引擎的工作原理、实现技术及其系统构建方案。《搜索引擎:原理技术与系统》分三篇共13章内容,从基本工作原理概述,到一个小型简单搜索引擎具体细节的实现,进而详细讨论了大规模分布式搜索引擎系统的设计要点及其关键技术;最后介绍了面向主题和个性化的web信息服务,阐述了中文网页自动分类等技术及其应用。《搜索引擎:原理技术与系统》层次分明,由浅入深;既有深入的理论分析,也有大量的实验数据,具有学习和实用双重意义。

作者简介

李晓明：天网搜索引擎领域负责人

闫宏飞 王继民：天网搜索引擎项目负责人

《搜索引擎》

书籍目录

前言第一章 引论 第一节 搜索引擎的概念 第二节 搜索引擎的发展历史 第三节 一些著名的搜索引擎上篇 Web搜索引擎基本原理和技术 第二章 Web搜索引擎工作原理和体系结构 第一节 基本要求 第二节 网页搜集 第三节 预处理 第四节 查询服务 第五节 体系结构 第三章 Web信息的搜集 第一节 引言 一、超文本传输协议 二、一个小型搜索引擎系统 第二节 网页搜集 一、定义URL类和Page类 二、与服务器建立连接 三、发送请求和接收数据 四、网页信息存储的天网格式 第三节 多道搜集程序并行工作 一、多线程并发工作 二、控制对一个站点并发搜集线程的数目 第四节 如何避免网页的重复搜集 一、记录未访问、已访问URL和网页内容摘要信息 二、域名与IP的对应问题 第五节 如何首先搜集重要的网页 第六节 搜集信息的类型 第七节 本章小结 第四章 对搜集信息的预处理 第一节 信息预处理的系统结构- 第二节 索引网页库 第三节 中文自动分词 第四节 分析网页和建立倒排文件 第五节 本章小结 第五章 信息查询服务 第一节 查询服务的系统结构 第二节 检索的定义 第三节 查询服务的实现 一、结果集合的形成 二、查询结果显示 第四节 本章小结中篇 对质量和性能的追求 第六章 可扩展搜集子系统 第一节 天网系统概述和集中式搜集系统结构 一、天网系统结构 二、集中式搜集系统 第二节 利用并行处理技术高效搜集网页的一种方案 一、节点间URL的划分策略 二、关于性能的讨论 三、性能测试和评价 四、系统的动态可配置性设计 第三节 本章小结 第七章 网页净化与消重 第一节 网页净化与元数据提取 一、引言 二、Docview模型 三、网页的表示 四、提取Docview模型要素的方法 五、模型应用及实验研究 第二节 网页消重算法 一、消重算法..... 第八章 高性能检索子系统 第九章 用户行为的特征及缓存的应用 第十章 相关排序与系统质量评估下篇 面向主题和个性化的Wed信息服务 第十一章 中文网页自动分类技术 第十二章 搜索引擎个性化查询服务 第十三章 面向主题的信息搜集与应用参考文献附录 术语后记

精彩短评

- 1、囧。。。这书zou不是写给咱看的。。。
- 2、研究生入学时买的，就是内容有点旧，还是很好的一本书。（买的时候看不懂，后来因为各种事，就压箱底了，才翻出来）
- 3、这本书的名字应该叫《天网搜索的实现及成果》，书中内容将天网的核心技术倒是都讲了
- 4、这是一本非常好搜索引擎的入门书籍！
- 5、图书馆借了好几次，最终决定买一本。很有收藏价值的书，相信以为一直会用到
- 6、对于入门来说是不错的，虽然原理老了点，但是大体方向还是不变的。
- 7、比较完整地介绍了实现方法
- 8、2天内收到货 速度快
- 9、爬虫部分的知识很受用
- 10、搜索引擎算法
- 11、深入浅出，我学搜索的入门书，可以当作天网代码的文档。
- 12、刚拿到手,匆匆翻阅了一下,感觉内容很实在,应该是一本好书,真正阅读的时候希望不要令人失望
- 13、天网该讲的都讲了，就是有点陷入琐碎
- 14、不错，只是有些内容希望写的再详细一些！
- 15、适合学习搜索引擎的同学阅读，要是java版本就更好了
- 16、入门读物
- 17、在别人的博客中看见的，不错
- 18、理论联系实际，对深入了解搜索引擎的工作原理非常有用，遗憾的是对技术细节讨论不够充分
- 19、非常值得看的一本搜索引擎基础书籍。看了对搜索有比较多的认识
- 20、讲的非常好 我一看就明白了
- 21、给先生买的，对工作有帮助，呵呵
- 22、不错的一本书，适合入门了解搜索引擎
- 23、适合入门看，可以对搜索引擎有个整体了解
- 24、搜索引擎较为全面的入门书。
- 25、技术原理类的书 朋友推荐的 值得一看，这种工具类的书，只能在办公室书桌旁看，需要用心去思考和理解。
- 26、对搜索引擎的研究综述和基本框架谈论的较为透彻，作为科研论文的参考文献是不错的。第6、7章在今日看来也不失指导价值，作者在圈内算是有资历的人了，写出的东西当然有意义。搜索引擎领域的仁人志士理应买一本看看。美中不足的是出版年份较早，有些内容过时了，希冀能有新版出现！
- 27、to do
- 28、基础知识，不错的一本书，推荐一下。
- 29、因为第二版有点贵，就选择了第一版。做seo的话还是需要多熟悉一下基础的内容。
- 30、听seo三人行小涵推荐的，应该算是不错的了。里面很多公司，主要想学下搜索引擎的原理，想理解多一点百度。相信会有比较深刻的搜索引擎原理认识
- 31、叫搜索引擎，大多数时间只是在描述爬虫
- 32、系统理论 用词简单有深度
- 33、内容有点陈旧，但还是比较丰富易懂的。
- 34、满分好书 这个书比这就是搜索引擎好
- 35、算是原原本本将当年的天网还原了出来
最精华的部分，其实我觉得是其中的reference，在每个topic中都列出了一堆的可以参考的文献，对于搜索引擎的入门来说，提供了大家在浩瀚的资料里寻找的便利。
- 36、算法随便看看就行，水太深，书本讲的不够详细。
- 37、不错的书~能继续写新书就好了~
- 38、不适合初学者啊 好难明啊
- 39、本书介绍的较为详细，但不足的是没有自己编写的具体代码，用的是别人的开源项目，对于做项目的人来书，可以作为参考书，但是不能作为一本可以作为模板的设计辅助教材

《搜索引擎》

- 40、真正的理解了搜索的内容，比那些抄算法的强多了
- 41、力顶老板的书啊！！
- 42、搜索引擎经典著作，讲到的原理都很实用。值得一看。
- 43、看了前半部分的基础知识。把整个搜索引擎涉及的流程和需要解决的问题阐述的比较清楚简洁，非常适合从整理上了解一个搜索引擎需要解决哪些问题以及一些基本的解决方案。
- 44、封面不怎么喜欢，但是内容还行
- 45、比较专业，正在研究中
- 46、这本书可以帮助读者很快的了解整个搜索引擎的框架结构
- 47、这本书系统地介绍了搜索引擎的实现方法及原理，是了解搜索引擎知识的一本好书
- 48、讲的比较通俗易懂，在工程性方面更有指导意义。
- 49、可以对搜索的概念有个笼统的认识，有少许实例，06~08年读过一些。可以泛读参考。
- 50、描述了是做什么，讲了一些算法，和遇到的实际应用问题，具体来做还是要很多实践的
- 51、第一次读这类书，收获不小，对我有一定深度！
- 52、国内搜索引擎的书本来就少，这本书算是国内出版的最好的一本。书上内容结合了北大天网的实例来讲解说明，算是有理有据的，并非泛泛而谈。要知道天网搜索引擎算是国内的Numberone了。百度里面都是从天网招过去的，酷讯的老板干脆就是原来开发天网的。由此可知这本书的含金量是很高的。只是版本有点老，如果作者能重新出版，加入一些新的内容就更好了。
- 53、国内的学se比较好的书籍了。主要通过天网se介绍了se的架构和原理
- 54、对天网有深刻的探索，比较系统，便于只是融合
- 55、很不错的入门书
- 56、国内关于搜索引擎方面的书籍，本身就少，此书是非常好的搜索引擎方面的书籍，他不仅仅介绍理论，还有一个小的模型。关于中篇的介绍，要是在有实际的解决案例就好了
- 57、刚开始学搜索引擎，这本书蛮不错的，很适合初学者。
- 58、史上最经典的搜索引擎基础学习资料，详实，对学习SEO有相当大的帮助！
- 59、相对于其它几本来说，这本书还是不错的。。。可以一看
- 60、泛读一遍
- 61、推荐SEO从业人员看看类似的书,不要天马行空的瞎猜.
- 62、但还是有点贵,这本书不是很厚啊,材料不足那么贵吧
- 63、内容一气呵成，非常流畅，由浅入深，适合初学者
- 64、当当的送货速度真的很快...
- 65、学习seo首先要先系统的学习搜索引擎。知道了搜索引擎的工作原理才能让seo工作展开的更顺利，效果更明显。这本书虽然老了点，但是还是值得我们借鉴，看看就知道了，我也是从seo三人行那里得知的这本书。
- 66、第一次用,很方便,价钱也说的过去!很好的互联网产品!赞
- 67、paper collection
- 68、学习搜索引擎的原理和方法的好书，经典好书，值得推荐，作者还有系列相关书籍！
- 69、买来还没有看，应该还可以
- 70、好，但是可能要有基础的人才看得懂
- 71、质量、速度都还不错。
- 72、很基础，看起来内容也比较旧了，基本是论文改出来的，适合简单了解。
- 73、中文搜索，不啰嗦，简练的一书。
- 74、配合tse源代码一起看，很受用啊~~
- 75、书好，我喜欢，希望大家购买！
- 76、非常不错的一本搜索引擎教程，很值得一看！
- 77、以前翻过，爬虫、文本清洗、分词、存储、检索
- 78、还不错的入门书，百度前身，天网实验室的人写的，有些技术有点老了
- 79、这个书，还是值得买的，如果你对搜索引擎不是很了解的话
- 80、初学者之心，偏实践，轻理论，Tse的代码是可以读的。有些过时。
- 81、没读完。。以后需要时候继续读

《搜索引擎》

- 82、骗吃大喝
- 83、很实际,很现实,很实在的一本书
- 84、通过阅读它，可以学点搜索引擎原理，有道二面之前买的
- 85、虽然很多东西现在已经不是最新的了，但作为搜索引擎入门教材还是很不错的。
- 86、顶！很难得的一本好书
- 87、感觉写的不错，才看到第六章。。。
- 88、教科书
- 89、对于初学，给5星。读上篇，结合TSE代码还是很有趣。
- 90、国内罕见的真材实料技术书
- 91、买了好几本书，质量都不错，当当一如既往的好
- 92、相当专业，在网上看人家推荐的，还可以
- 93、其实俺不是很喜欢这种学究性的东西，但理论学习的意义真的很重要，这本书读起来清爽自在，可能因为比较薄有关，但作者说话也轻快。比较烦的是两个方面，一是中间的代码太丑了，代码风格差劲，排版更是丑；二是凡是点到为止，甚至没点到就止。算是起步吧。
- 94、：
G252.7/4466
- 95、老师给过了，就好
- 96、翻过而已，好书，权威
- 97、了解中文搜索引擎原理必备书籍
- 98、这本书早几个月就买好了，放在家里一直没看，对书没什么评价，不过送货速度还蛮好的。
- 99、内容不怎么生动，拿到这本书让我想起大学的古板的教科书。
- 100、还没看，争取尽快看看
- 101、入门还是不错的

- 1、适合搜索引擎入门时阅读：内容还算比较全面，涉及到SE的各个方面，但很多技术的确有点老了，毕竟这本书出的比较早建议配合TSE 代码阅读
- 2、国内的著作，特别是冠以学术的东西，不论是可读性还是内容的质量都很糟糕，但这本却是例外。300多页的内容把搜索引擎的原理讲的很清晰，此书成于2005年，搜索引擎领域的发展发生了极大的变化，但是基本的原理还是想通的，需要解决的问题还是一致。比如分词，检索还有存储，书中给出的例子通俗易懂，远比直接看xx引擎的API收获大多了，如果对搜索有兴趣却苦于无法入门的朋友，力荐本书！
- 3、因为以后要从事搜索开发的工作，所以公司推荐了这本书。书挺薄的，前后一个月看完吧，总体感觉还行。这本书把搜索引擎相关的各项技术基本都做了介绍，比较全面，算是为数不多、质量不错的入门书籍。说说缺点吧，这本书应该是北大n多论文拼出来的，有一种前后不太连贯的感觉；因为是论文，讲得都有点太偏学术性了，涉及实践操作的东西比较少；还有就是有一些细节没有交代清楚，既然是启蒙的书，应该详细些吧。这本书出版得有些早了，不少技术都过时了，希望作者能与时俱进，再出新版，应该不错！
- 4、北大天网实验室出的一本书，主要结合了天网的实践，并有一套称为TSE的C++代码。全书分为三部分。除了第三部分涉及更多的高级问题，理论性较强，书中描述也不太详细之外，前两部分对于非专业人士了解IR系统的“原理，技术与系统”很有帮助。该书对网页抓取，文本分析，索引建立，搜索效率等方面都做出了分析，提供了一个功能详尽的web文本信息检索框架，阅读时可能需要查找少量的论文或其他信息。
- 5、主要是由北大李晓明那个实验室所发表的论文组成，很多地方偏学术，但是在国内这本书应该是最最好的搜索引擎方面的书籍，推荐大家作为搜索引擎入门的书籍，要了解最新的搜索引擎技术还是要多读SIGIR,WWW等会议的相关论文。读完这本书，可以进一步学习<Pattern Recognition and Machine Learning> <http://www.douban.com/subject/2061116/?i=0>
- 6、最近埋头苦看各种搜索引擎原理的书籍，当然我是一个入门者，所以从入门者的角度来说几句吧~首先我的背景是给老外打工，所以几乎都是英文，挑选这本书仅仅是偶然，其实我想找的是另外一本《信息检索实践》，在误点的情况下下载了本书，然后读完了，通读一遍的感觉是里面还不是很深入。个人感觉搜索引擎的重点有三个：划词，搜索，抓取。划词是最难的，个人感觉是最难的，因为不论从中文还是英文来说，各种的动词数词和名词都是比较难以掌控，简单点举个例子“google”是一个新生词，很多人用google代表搜索，用百度同样也可以代表搜索，在这种情况下，划词是最困难的.....而电脑搜索和图书馆搜索的最大不一样在于图书馆是既定的，但是电脑是无限扩张。但是这本书里面其实并没有很详细的说这种东西，这种东西如果说，其实也不只有300页就能讲完了吧。排序，其实这玩意就是数学，加权神马的谁最高谁排第一，问题是怎么算这个值。这段部分其实作者也结合了各种划词的原理，来讨论各种词的属性之类，不过还不是很详细。搜索是最接近程序员的，从一开始的SQL的各种select语句开始，就开始接触到了搜索，不过搜索引擎的搜索是以文档为基础的，而非以数据库。本书里面提到一个新的东西叫BigTable，我从来没有接触过，但是个人感觉和Linux的系统很相似，BigTable是一种基于文档结构而开发的带有数据检索功能的文档系统（怎么这么绕？）好吧，反正就是很好玩啦。最近在接触的是NOSQL的数据库，也是大型的文档型数据库，正在尝试两者互相融入，因为数据流量并非只有100000这么少，日常的IO工作保守估计也会很大量，暂时还没做测试，所以不表。最后一点就是抓取，花了大量篇幅来说抓取工作，从URL的角度，从页面链接的角度。对于这个我比较外行，所以我根本不知道我该说什么.....因为什么都很重要，人工智能也不过如此吧。最后的最后，欢迎交流，请勿吐槽，随笔于TUE图书馆。

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:www.tushu111.com