

搜索引擎



[搜索引擎 下载链接1](#)

著者:李晓明

出版者:科学出版社发行部

出版时间:2005-4

装帧:平装

isbn:9787030146335

《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。

作者介绍：

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

目录：

[搜索引擎 下载链接1](#)

标签

搜索引擎

信息检索

搜索引擎原理

互联网

计算机

计算机科学

Search

IR

评论

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

: G252.7/4466

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）

初学者之心，偏实践，轻理论，Tse的代码是可以读的。有些过时。

paper collection

囧。。。这书zou不是写给咱看的。。。

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。

翻过而已，好书，权威

可能是关于SE的唯一一本中文书了，里边讲常用技术一锅端了，但是要深入还是参考相应的论文，还有一点比较好的是此书有代码（TSE）支持，做到了知行合一。

泛读一遍

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

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。

配合tse源代码一起看，很受用啊~~

深入浅出，我学搜索的入门书，可以当作天网代码的文档。

内容有点陈旧，但还是比较丰富易懂的。

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。
◦

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

国内罕见的真材实料技术书

不错的书～能继续写新书就好了～

通过阅读它，可以学点搜索引擎原理，有道二面之前买的

搜索引擎 下载链接1

书评

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...

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，但是在国外这本书应该是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

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...

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...

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

谷歌搜索引擎优化seo？能不能自动优化？

万水千山总是情，你在那里头，我在这头，默默的祝福，深深的思念，就让清风捎去我的问候，一切都未曾改变，你永远是我的牵挂！

走过山山水水，路过春夏秋冬，克服千险万阻，为你搜集一件无忧衫，前身是吉祥如意，后背是平安喜气，袖子是快...

是阅读该书及TSE源代码非常好的参考资料，可以作为想从零了解搜索引擎的朋友的入门资料。

可以参考某人的csdn博客中的笔记：<http://blog.csdn.net/column/details/inside-tse.html>

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...

[搜索引擎 下载链接1](#)