

<<走进搜索引擎>>

图书基本信息

书名：<<走进搜索引擎>>

13位ISBN编号：9787121049224

10位ISBN编号：7121049228

出版时间：2007-1

出版时间：电子工业出版社

作者：梁斌

页数：272

字数：258000

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>

## <<走进搜索引擎>>

### 内容概要

在网络普及的今天，人们经常在信息海洋中彷徨，在万维网迷宫般的复杂与魅力之间挣扎。直到搜索引擎这一伟大的技术产生，才使得人们犹如找到了走出迷宫的灯塔，可以非常便捷地找到自己所需要的信息。

正是因为搜索引擎离我们越来越近，所以越来越多的人期待着能够揭开她神秘的面纱。

其实搜索引擎并不是变幻莫测的大海，也不是高不可攀的山峰。

请拿起本书，它就是引领你的火炬，它就是你身边的伙伴，它将带着你走进搜索引擎。

在那里，你必将会被搜索引擎精致的设计和宏伟的架构所征服。

本书由搜索引擎开发研究领域年轻而有活力的科学家精心编写，作者将自己对搜索引擎的深刻理解和实际应用巧妙地结合，使得从未接触过搜索引擎原理的读者也能够轻松地在搜索引擎的大厦中遨游一番。

本书作为搜索引擎原理与技术的入门书籍，面向那些有志从事搜索引擎行业的青年学生、需要完整理解并优化搜索引擎的专业技术人员、搜索引擎的营销人员，以及网站的负责人等。

本书是从从事搜索引擎开发的工程技术人员难得的参考书，也可作为大中专院校相关专业的教学辅导书。

## 作者简介

梁斌毕业于南京大学，获得软件工程硕士学位，曾经发表过多篇论文，获得1项国家专利，作者主要的兴趣方向包括数据挖掘、Web挖掘、搜索引擎和软件工程等，目前在清华大学信息科学与技术国家实验室从事搜索引擎相关研究工作。

## &lt;&lt;走进搜索引擎&gt;&gt;

## 书籍目录

第一章 引言	第一节 什么是搜索引擎	第二节 搜索引擎的发展简史	搜索引擎的发展历史	第三
节 搜索引擎大事快览	第四节 国内著名搜索引擎	百度 (www.baidu.com)	中搜	
(www.zhongsou.com)	天网 (e.pku.edu.cn)	搜狗 (www.sogou.com)	参考文献	第二章 搜
引擎概貌	第一节 搜索引擎的主要需求	查得快	查得全	查得准
引擎的大系统	搜索引擎的体系结构	第三章 搜索引擎的下载系统	第一节 爬虫的发展历史	第二
世界上第一个爬虫	爬虫的发展历程	第二节 万维网及其网页分析	蝴蝶结型的万维网	万维网
的直径	万维网的规模及变化特征	网页的特征	第三节 有关爬虫的基本概念	爬虫
子站点	URL	Backlinks	第四节 网页抓取原理	telnet和wget
不重复抓取策略	网页抓取优先策略	网页重访策略	Robots协议	其他应该注意的
貌性问题	抓取提速策略 (合作抓取策略)	第五节 网页库	第六节 下载系统回顾及未来发展	
参考文献	第四章 搜索引擎的分析系统	第一节 知识准备	HTML语言	锚文本 (anchor text)
结构化数据 (Semi-structured data)	第二节 信息抽取及网页信息结构化	网页结构化的目标	建	
立HTML标签树	通过投票方法得到正文	网页结构化过程回顾	第三节 网页查重	网页查重技术发展
历史	网页查重实现方法	第四节 中文分词	什么是中文分词	通过字典实现分词
分词	第五节 PageRank	PageRank的来由	PageRank的基本想法	PageRank的计算公式
方法	第六节 分析系统结构图	参考文献	第五章 搜索引擎的索引系统	第一节 知识准备
倒排索引、倒排表、临时倒排文件、最终倒排文件	其他概念	第二节 全文检索	全文检索	第三节
档编号	编号的本质	文档编号的方法	游程编码	第四节 倒排索引
索引)	倒排索引	第五节 数据规模的估计	齐普夫法则	布尔检索模型下的索引规模估计
涉及存储规模的一些计算	正排表与倒排表的合并	多个临时倒排文件的归并	倒排索引分布式存储	
倒排文件缓存	倒排索引词典统计信息的计算	第七节 倒排索引文件的创建过程	创建倒排表	计算
计信息	参考文献	第六章 搜索引擎的查询系统	第一节 知识准备	什么是信息熵
检索词和查询词的区别	自动文本摘要 (Automatic Text Summarization)	第二节 网页信息检索	早期的检	
索模型	向量空间模型 (Vector Space Models)	关键词权重的量化方法TF/IDF	搜索引擎采用的检索	
模型	多文档列表求交计算	检索结果排序	堆排序	第三节 中文自动摘要
摘要的含义和实现	第四节 生成搜索结果页	生成搜索结果页	第五节 搜索结果页的缓存	搜索结果
的缓存	第六节 推测用户查询意图	查询分类	推测信息类、事物类的查询意图	第七节 查询系统的
前热点和发展方向	查询系统的当前热点	参考文献	第七章 搜索引擎的其他话题	第一节 搜索引擎问
答	为什么搜索引擎的搜索速度这么快	为什么搜索引擎能够返回那么多的查询结果	为什么搜索引擎	
总能返回最想要的结果	搜索引擎如何大规模存储网页的	什么是SEO	什么是元搜索引擎	搜索引擎
为的作弊行为是哪些	如何进一步学习和了解搜索引擎发展的最新成果	第二节 搜索引擎未来的发展		
新兴的搜索产品	搜索技术的未来	参考文献	附录A 搜索引擎系统结构全视图	

#### 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>