

## <<信息检索>>

### 图书基本信息

书名：<<信息检索>>

13位ISBN编号：9787111359906

10位ISBN编号：7111359909

出版时间：2012-1

出版时间：机械工业出版社

作者：（美）Stefan B ü ttcher,（加）Charles L.A.Clarke,（加）Gordon V. Cormack

译者：陈健,黄晋

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>

## <<信息检索>>

### 内容概要

《信息检索：实现和评价搜索引擎》从多个视角对信息检索技术进行了深入讲解，内容涵盖了信息检索系统的架构、基础技术、词条和词项、静态和动态倒排索引、查询处理、索引压缩技术、概率模型、语言模型、分类和过滤、融合和元学习、评价方法以及并行信息检索、Web检索和XML检索等具体应用。

本书以模块化的方式进行组织，理论性强，体系完整，同时强调实践。

作者以认真严谨的态度实现了书中绝大部分的主要方法，并详尽地描述了各种方法的适用环境以及取得的效果。

《信息检索：实现和评价搜索引擎》可作为高等院校信息管理与信息系统、计算机科学与技术、情报学、图书馆学以及电子商务等专业的高年级本科生和研究生的教材和参考书，对于从事信息检索与网络分析等实际工作的从业人员也具有较高的参考价值。

<<信息检索>>

作者简介

作者：(美国)布切尔 (Stefan Buttcher) (加拿大)Charles L.A.Clarke (加拿大)Gordon V.Cormack 译者：陈健黄晋 等布切尔(Stefan Buttcher)是Google公司资深网站可靠性工程师。

Charles L A.Clarke和Gordon V Cormack是滑铁卢大学David R.Cheriton计算机科学学院的计算机科学教授。

## <<信息检索>>

### 书籍目录

出版者的话

译者序

序

前言

符号

#### 第一部分 基础知识

##### 第1章 绪论

###### 1.1 什么是信息检索

###### 1.1.1 web搜索

###### 1.1.2 其他搜索应用

###### 1.1.3 其他信息检索应用

###### 1.2 信息检索系统

###### 1.2.1 信息检索系统基础架构

###### 1.2.2 文档及其更新

###### 1.2.3 性能评价

###### 1.3 使用电子文本

###### 1.3.1 文本格式

###### 1.3.2 英文文本中的分词

###### 1.3.3 词项分布

###### 1.3.4 语言模型

###### 1.4 测试集

###### 1.5 开源信息检索系统

###### 1.5.1 lucene

###### 1.5.2 indri

###### 1.5.3 wumpus

###### 1.6 延伸阅读

###### 1.7 练习

###### 1.8 参考文献

##### 第2章 基础技术

###### 2.1 倒排索引

###### 2.1.1 延伸例子：词组查找

###### 2.1.2 实现倒排索引

###### 2.1.3 文档和其他元素

###### 2.2 检索与排名

###### 2.2.1 向量空间模型

###### 2.2.2 邻近度排名

###### 2.2.3 布尔检索

###### 2.3 评价

###### 2.3.1 查全率和查准率

###### 2.3.2 排名检索的有效性指标

###### 2.3.3 创建测试集

###### 2.3.4 效率指标

###### 2.4 总结

###### 2.5 延伸阅读

###### 2.6 练习

## <<信息检索>>

- 2.7 参考文献
- 第3章 词条与词项
  - 3.1 英语
    - 3.1.1 标点与大写
    - 3.1.2 词干提取
    - 3.1.3 停词
  - 3.2 字符
  - 3.3 字符n<sup>2</sup>gram
  - 3.4 欧洲语言
  - 3.5 cjk语言
  - 3.6 延伸阅读
  - 3.7 练习
  - 3.8 参考文献
- 第二部分 索引
  - 第4章 静态倒排索引
  - 第5章 查询处理
  - 第6章 索引压缩
  - 第7章 动态倒排索引
- 第三部分 检索和排名
  - 第8章 概率检索
  - 第9章 语言模型及其相关方法
  - 第10章 分类和过滤
  - 第11章 融合和元学习
- 第四部分 评价
  - 第12章 度量有效性
  - 第13章 度量效率
- 第五部分 应用和扩展
  - 第14章 并行信息检索
  - 第15章 web搜索
  - 第16章 xml检索
- 第六部分 附录
  - 附录a 计算机性能

## <<信息检索>>

### 章节摘录

版权页：插图：1.1 什么是信息检索信息检索被认为是对大规模电子文本和其他人类语言数据进行表示、搜索和处理的技术。

信息检索系统和服务现在已经非常普遍了，成千上万的人每天都使用它们来方便地进行商务、教育和娱乐。

Google、Bing等Web搜索引擎，是目前为止最普遍和大量使用信息检索服务的形式，提供获取最新技术信息、搜索人和组织、总结新闻和事件以及简化比较购物的途径。

电子图书馆系统帮助医学界和学术界的研究人员了解他们研究领域内最新的期刊文章和会议报告。

消费者使用本地搜索服务来找到提供所需产品和服务的零售商。

在大型公司中，企业搜索系统作为电子邮件、备忘录、技术报告和其他业务文档的存储库，通过保存这些文档和提供相应的手段获得文档蕴涵的知识来提供企业记忆。

桌面搜索系统则允许用户搜索他们的个人电子邮件、文档和文件。

1.1.1 Web搜索对Web搜索引擎的一般用户而言，通常希望只要在一个文本框里输入一个简短的查询——几个简单的词，然后点击一下搜索按钮，马上就可以得到问题的精确答案。

在这简单直观的界面后面是一组计算机集群，包括成千上万台协同工作的机器，用来产生最有可能满足查询中所包含信息的网页排名列表。

这些机器要识别包含查询词的网页集合，计算每个网页的得分，消除重复和多余的页面，生成余下页面的摘要，最后将摘要和链接返回给用户以便浏览。

## <<信息检索>>

### 媒体关注与评论

“这本书是越来越多的信息检索著作中的一本好书。

”——Donald H. Krafft, 计算机评论 “学术巨匠齐聚一堂编撰了一部信息检索的优秀教材。

Stefan Buttcher.Charles L.A.Clarke和Gordon V.Cormack以合计超过50年的研究经验, 组成了横跨三代的信息检索研究泰斗组合.....这本书是所有信息检索研究者和从业人员的必读教材!

”——摘自Amit Singhal撰写的序言

## &lt;&lt;信息检索&gt;&gt;

## 编辑推荐

《信息检索:实现和评价搜索引擎》：信息检索奠定了现代搜索引擎的基石。

《信息检索:实现和评价搜索引擎》介绍了现代搜索技术的核心主题，包括算法、数据结构、索引、检索和评价，重点在于实现和实验，每一章都有练习和对学生的建议。

Wumpus（《信息检索:实现和评价搜索引擎》其中一位作者开发的一个多用户开源信息检索系统，可以在网上下载）提供了模型实现，可作为学生练习的一个基础，《信息检索:实现和评价搜索引擎》采用的模块化结构使教师可以将此书用于不同水平的研究生课程中，包括从数据库系统角度教授的课程、专注于理论的传统信息检索课程和关于Web检索基础的课程。

对信息检索的基础进行介绍之后，《信息检索:实现和评价搜索引擎》分别在相应的部分介绍了3个重要主题——索引、检索和评价。

《信息检索:实现和评价搜索引擎》的最后一部分借用并扩展了前面部分的基本内容，考虑了以下具体应用：并行搜索引擎、Web搜索和XML检索。

除了用于课堂教学，《信息检索:实现和评价搜索引擎》对计算机科学、计算机工程和软件工程的专业人员来说也具有很好的参考价值。

<<信息检索>>

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>