

<<现代信息检索>>

图书基本信息

书名：<<现代信息检索>>

13位ISBN编号：9787111385998

10位ISBN编号：7111385993

出版时间：2012-10

出版时间：机械工业出版社

作者：Ricardo Baeza-Yates,Berthier Ribeiro-Neto

页数：672

译者：黄萱菁,张奇,邱锡鹏

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>

## <<现代信息检索>>

### 内容概要

《现代信息检索（原书第2版）》论述信息检索的概念和技术、这些技术在搜索引擎中的应用，及其对相关领域知识的影响等，主要内容包括：用户界面设计；经典的信息检索模型、结果质量评估和用户相关反馈；文档和查询概念及其相关技术；文档集索引和搜索技术；Web文档的爬取、检索和排序；结构化文本检索、多媒体检索和企业搜索；图书馆系统和数字图书馆等。

《现代信息检索（原书第2版）》内容广泛、细节丰富、深入浅出，可以作为高等院校信息管理与信息系统、计算机科学与技术、图书馆学、情报学、档案学等专业本科生和研究生的教材或参考书，对从事信息检索及系统分析、设计的实际工作者也有较高的参考价值。

书籍目录

- 出版者的话
- 译者序
- 第2版前言
- 第1版前言
- 第2版致谢
- 第1版致谢
- 出版商致谢
- 第1章 引言
  - 1.1 信息检索
    - 1.1.1 信息检索的早期发展
    - 1.1.2 图书馆和数字图书馆中的信息检索
    - 1.1.3 舞台中央的信息检索
  - 1.2 信息检索问题
    - 1.2.1 用户的任务
    - 1.2.2 信息检索与数据检索
  - 1.3 信息检索系统
    - 1.3.1 信息检索系统的软件架构
    - 1.3.2 检索和排序过程
  - 1.4 Web
    - 1.4.1 Web简史
    - 1.4.2 电子出版时代
    - 1.4.3 Web如何改变搜索
    - 1.4.4 Web上的实际问题
  - 1.5 本书的组织结构
    - 1.5.1 本书的重点
    - 1.5.2 本书的内容
  - 1.6 本书的教学资源网站
  - 1.7 文献讨论
- 第2章 用户搜索界面
  - 2.1 介绍
  - 2.2 人们如何搜索
    - 2.2.1 信息查找与探索式搜索
    - 2.2.2 信息搜寻的经典模型与动态模型
    - 2.2.3 导航与搜索
    - 2.2.4 对搜索过程的观察
  - 2.3 现今的搜索界面
    - 2.3.1 启动搜寻
    - 2.3.2 查询描述
    - 2.3.3 查询描述界面
    - 2.3.4 检索结果显示
    - 2.3.5 查询重构
    - 2.3.6 组织搜索结果
  - 2.4 搜索界面的可视化
    - 2.4.1 可视化布尔语法
    - 2.4.2 可视化查询结果中的查询项

## &lt;&lt;现代信息检索&gt;&gt;

- 2.4.3 可视化词语和文档间的关系
- 2.4.4 文本挖掘的可视化
- 2.5 搜索界面的设计和评价
- 2.6 趋势和研究问题
- 2.7 文献讨论
- 第3章 信息检索建模
  - 3.1 信息检索模型
    - 3.1.1 建模和排序
    - 3.1.2 信息检索模型描述
    - 3.1.3 信息检索模型的分类体系
  - 3.2 经典信息检索
    - 3.2.1 基本概念
    - 3.2.2 布尔模型
    - 3.2.3 项权重
    - 3.2.4 TF-IDF权重
    - 3.2.5 文档长度归一化
    - 3.2.6 向量模型
    - 3.2.7 概率模型
    - 3.2.8 经典模型之间的简单比较
  - 3.3 其他集合论模型
    - 3.3.1 基于集合的模型
    - 3.3.2 扩展布尔模型
    - 3.3.3 模糊集模型
  - 3.4 其他代数模型
    - 3.4.1 广义向量空间模型
    - 3.4.2 潜在语义索引模型
    - 3.4.3 神经网络模型
  - 3.5 其他概率模型
    - 3.5.1 BM25模型
    - 3.5.2 语言模型
    - 3.5.3 随机差异模型
    - 3.5.4 贝叶斯网模型
  - 3.6 其他模型
    - 3.6.1 超文本模型
    - 3.6.2 基于Web的模型
    - 3.6.3 结构化文本检索
    - 3.6.4 多媒体检索
    - 3.6.5 企业和垂直搜索
  - 3.7 趋势和研究问题
  - 3.8 文献讨论
- 第4章 检索评价
  - 4.1 介绍
  - 4.2 Cranfield范式
    - 4.2.1 历史简述
    - 4.2.2 参考集
  - 4.3 检索指标
    - 4.3.1 精度和召回率

## &lt;&lt;现代信息检索&gt;&gt;

- 4.3.2 单值总结 : P@n , MAP , MRR , F
  - 4.3.3 面向用户的指标
  - 4.3.4 折扣累积增益
  - 4.3.5 二元偏好
  - 4.3.6 排序相关性测度
  - 4.4 参考文档集
    - 4.4.1 TREC参考集
    - 4.4.2 其他参考集
    - 4.4.3 其他小规模测试文档集
  - 4.5 基于用户的评价
    - 4.5.1 实验室中的人工实验
    - 4.5.2 并排面板
    - 4.5.3 A/B测试
    - 4.5.4 众包
    - 4.5.5 使用点击数据的评价
  - 4.6 实践说明
  - 4.7 趋势和研究问题
  - 4.8 文献讨论
- 第5章 相关反馈与查询扩展
- 5.1 介绍
  - 5.2 反馈方法的框架
  - 5.3 显式相关反馈
    - 5.3.1 向量模型的相关反馈 : Rocchio方法
    - 5.3.2 概率模型的相关反馈
    - 5.3.3 相关反馈的评价
  - 5.4 基于点击的显式反馈
    - 5.4.1 眼动追踪和相关性评价
    - 5.4.2 用户行为
    - 5.4.3 点击作为用户偏好的指标
  - 5.5 通过局部分析的隐式反馈
    - 5.5.1 通过局部聚类的隐式反馈
    - 5.5.2 通过局部上下文分析的隐式反馈
  - 5.6 通过全局分析的隐式反馈
    - 5.6.1 基于相似度同义词典的查询扩展
    - 5.6.2 基于统计同义词典的查询扩展
  - 5.7 趋势和研究问题
  - 5.8 文献讨论
- 第6章 文档 : 语言及属性
- 6.1 介绍
  - 6.2 元数据
  - 6.3 文档格式
    - 6.3.1 文本
    - 6.3.2 多媒体
    - 6.3.3 图形和虚拟现实
  - 6.4 标记语言
    - 6.4.1 SGML
    - 6.4.2 HTML

## &lt;&lt;现代信息检索&gt;&gt;

- 6.4.3 XML
- 6.4.4 RDF
- 6.4.5 HyTime
- 6.5 文本属性
  - 6.5.1 信息论
  - 6.5.2 自然语言建模
  - 6.5.3 文本相似度
- 6.6 文档预处理
  - 6.6.1 文本的词汇分析
  - 6.6.2 去除禁用词
  - 6.6.3 词干提取
  - 6.6.4 关键词选择
  - 6.6.5 同义词典
- 6.7 组织文档
  - 6.7.1 分类体系法
  - 6.7.2 分众分类法
- 6.8 文本压缩
  - 6.8.1 基本概念
  - 6.8.2 统计方法
  - 6.8.3 统计方法：建模
  - 6.8.4 统计方法：编码
  - 6.8.5 字典方法
  - 6.8.6 压缩预处理
  - 6.8.7 文本压缩技术的比较
  - 6.8.8 结构化文本压缩
- 6.9 趋势和研究问题
- 6.10 文献讨论
- 第7章 查询：语言及属性
  - 7.1 查询语言
    - 7.1.1 基于关键词的查询
    - 7.1.2 非关键词查询
    - 7.1.3 结构化查询
    - 7.1.4 查询协议
  - 7.2 查询属性
    - 7.2.1 Web查询的特征
    - 7.2.2 用户搜索行为
    - 7.2.3 查询意图
    - 7.2.4 查询主题
    - 7.2.5 查询会话与任务
    - 7.2.6 查询难度
  - 7.3 趋势和研究问题
  - 7.4 文献讨论
- 第8章 文本分类
  - 8.1 介绍
  - 8.2 文本分类的特性描述
    - 8.2.1 机器学习
    - 8.2.2 文本分类问题

## &lt;&lt;现代信息检索&gt;&gt;

- 8.2.3 文本分类算法
- 8.3 无监督算法
  - 8.3.1 聚类
  - 8.3.2 朴素文本分类
- 8.4 监督算法
  - 8.4.1 决策树
  - 8.4.2 k近邻分类器
  - 8.4.3 Rocchio分类器
  - 8.4.4 概率朴素贝叶斯文档分类
  - 8.4.5 支持向量机分类器
  - 8.4.6 集成分类器
  - 8.4.7 关于监督算法的结束语
- 8.5 特征选择或降维
  - 8.5.1 项-类别出现列联表
  - 8.5.2 索引项文档频率
  - 8.5.3 TF-IDF权重
  - 8.5.4 互信息
  - 8.5.5 信息增益
  - 8.5.6 卡方检验
  - 8.5.7 特征选择的作用
- 8.6 评价指标
  - 8.6.1 列联表
  - 8.6.2 准确率和错误率
  - 8.6.3 精度和召回率
  - 8.6.4 F测度和F
  - 8.6.5 交叉检验
  - 8.6.6 标准文档集
- 8.7 类别组织--构建分类体系
- 8.8 趋势和研究问题
- 8.9 文献讨论
- 第9章 索引和搜索
  - 9.1 介绍
  - 9.2 倒排索引
    - 9.2.1 基本概念
    - 9.2.2 完全倒排索引
    - 9.2.3 搜索
    - 9.2.4 排序
    - 9.2.5 构建
    - 9.2.6 压缩的倒排索引
    - 9.2.7 结构化查询
  - 9.3 签名文件
  - 9.4 后缀树和后缀数组
    - 9.4.1 结构: trie树和后缀树
    - 9.4.2 简单字符串搜索
    - 9.4.3 复杂模式的搜索
    - 9.4.4 构建
    - 9.4.5 压缩的后缀数组

## &lt;&lt;现代信息检索&gt;&gt;

## 9.5 序列搜索

- 9.5.1 简单字符串：Horspool
- 9.5.2 复杂模式：自动机和位并行
- 9.5.3 更快的位并行算法
- 9.5.4 正则表达式
- 9.5.5 多重模式
- 9.5.6 近似搜索
- 9.5.7 搜索压缩文本

## 9.6 多维索引

## 9.7 趋势和研究问题

## 9.8 文献讨论

## 第10章 并行与分布式信息检索

## 10.1 介绍

## 10.2 分布式信息检索系统的分类

## 10.3 数据划分

- 10.3.1 文档集划分
- 10.3.2 文档集选择
- 10.3.3 倒排索引划分
- 10.3.4 划分其他索引

## 10.4 并行信息检索

## 10.4.1 介绍

## 10.4.2 在MIMD架构上的并行信息检索

## 10.4.3 在SIMD架构上的并行信息检索

## 10.5 基于集群的信息检索

## 10.6 分布式信息检索

## 10.6.1 介绍

## 10.6.2 索引

## 10.6.3 查询处理

## 10.6.4 Web问题

## 10.7 联合搜索

## 10.8 在对等网络中的检索

## 10.9 趋势和研究问题

## 10.10 文献讨论

## 第11章 Web检索

## 11.1 介绍

## 11.2 一个有挑战性的问题

## 11.3 Web

- 11.3.1 特性
- 11.3.2 Web图的结构
- 11.3.3 对Web建模
- 11.3.4 链接分析

## 11.4 搜索引擎架构

- 11.4.1 基本架构
- 11.4.2 基于集群的架构
- 11.4.3 缓存
- 11.4.4 多级索引
- 11.4.5 分布式架构



## &lt;&lt;现代信息检索&gt;&gt;

- 11.5 搜索引擎排序
    - 11.5.1 排序信号
    - 11.5.2 基于链接的排序
    - 11.5.3 简单的排序函数
    - 11.5.4 排序学习
    - 11.5.5 学习排序函数
    - 11.5.6 质量评价
    - 11.5.7 Web垃圾
  - 11.6 管理Web数据
    - 11.6.1 为文档分配标识符
    - 11.6.2 元数据
    - 11.6.3 压缩Web图
    - 11.6.4 处理重复数据
  - 11.7 搜索引擎用户交互
    - 11.7.1 搜索矩形范式
    - 11.7.2 搜索引擎结果页面
    - 11.7.3 培养用户
  - 11.8 浏览
    - 11.8.1 扁平浏览
    - 11.8.2 结构导向的浏览和Web目录
  - 11.9 浏览之外
    - 11.9.1 超文本和Web
    - 11.9.2 搜索与浏览相结合
    - 11.9.3 Web查询语言
    - 11.9.4 动态搜索
  - 11.10 相关问题
    - 11.10.1 计算广告学
    - 11.10.2 Web挖掘
    - 11.10.3 元搜索
  - 11.11 趋势和研究问题
    - 11.11.1 静态文本数据之外
    - 11.11.2 目前的挑战
  - 11.12 文献讨论
- 第12章 Web爬取
- 12.1 介绍
  - 12.2 网络爬虫的应用
    - 12.2.1 通用Web搜索
    - 12.2.2 聚焦爬取
    - 12.2.3 Web刻画
    - 12.2.4 镜像
    - 12.2.5 网站分析
  - 12.3 爬虫的分类体系
  - 12.4 架构和实现
    - 12.4.1 爬虫架构
    - 12.4.2 实际问题
    - 12.4.3 并行爬取
  - 12.5 调度算法

## &lt;&lt;现代信息检索&gt;&gt;

- 12.5.1 选择策略
- 12.5.2 重访问策略
- 12.5.3 友好策略
- 12.5.4 组合策略
- 12.6 评价
  - 12.6.1 评价网络使用
  - 12.6.2 评价长期调度
- 12.7 趋势和研究问题
  - 12.7.1 爬取“暗网”
  - 12.7.2 在网站帮助下的爬取
  - 12.7.3 分布式爬取
- 12.8 文献讨论
- 第13章 结构化文本检索
  - 13.1 介绍
  - 13.2 结构化能力
    - 13.2.1 显式和隐式结构对比
    - 13.2.2 静态与动态结构对比
    - 13.2.3 单一层次结构与多层次结构对比
  - 13.3 早期文本检索模型
    - 13.3.1 基于非覆盖列表的模型
    - 13.3.2 基于相邻结点的模型
    - 13.3.3 结构化文本结果排序
  - 13.4 XML检索
    - 13.4.1 XML检索中的挑战
    - 13.4.2 索引策略
    - 13.4.3 排序策略
    - 13.4.4 去除重叠
  - 13.5 XML检索评价
    - 13.5.1 文档集
    - 13.5.2 主题
    - 13.5.3 检索任务
    - 13.5.4 相关性
    - 13.5.5 测度
  - 13.6 查询语言
    - 13.6.1 特性
    - 13.6.2 XML查询语言分类
    - 13.6.3 XML查询语言样例
  - 13.7 趋势和研究问题
  - 13.8 文献讨论
- 第14章 多媒体信息检索
  - 14.1 介绍
    - 14.1.1 什么是多媒体
    - 14.1.2 多媒体检索
    - 14.1.3 文本检索与多媒体检索的对比
  - 14.2 挑战
    - 14.2.1 语义鸿沟
    - 14.2.2 特征歧义性

## &lt;&lt;现代信息检索&gt;&gt;

- 14.2.3 机器生成的数据
- 14.3 基于内容的图像检索
  - 14.3.1 基于颜色的检索
  - 14.3.2 纹理
  - 14.3.3 显著点
- 14.4 声音和音乐检索
  - 14.4.1 指纹识别
  - 14.4.2 语音识别
  - 14.4.3 说话人识别
  - 14.4.4 语音文档检索
  - 14.4.5 音频基础知识
- 14.5 检索和浏览视频
  - 14.5.1 视频摘要
  - 14.5.2 静态摘要
  - 14.5.3 图像拼接与跳跃剧照
  - 14.5.4 动态摘要
  - 14.5.5 交互式摘要
  - 14.5.6 视觉与听觉浏览对比
  - 14.5.7 摘要评价
- 14.6 融合模型：合并所有信息
  - 14.6.1 人脸命名
  - 14.6.2 图像命名
  - 14.6.3 音频命名
  - 14.6.4 结合音频与视频的音-视频语音识别
  - 14.6.5 结合音频和视频的多媒体处理
- 14.7 分割
  - 14.7.1 视频分割样例
  - 14.7.2 视频分割方案
  - 14.7.3 利用边缘的视频分割
  - 14.7.4 语音分割
  - 14.7.5 分割评价
- 14.8 压缩和MPEG标准
  - 14.8.1 强度和采样
  - 14.8.2 颜色
  - 14.8.3 有损压缩
  - 14.8.4 无损压缩
  - 14.8.5 时间冗余
  - 14.8.6 运动预测
  - 14.8.7 MPEG标准
- 14.9 趋势和研究问题
- 14.10 文献讨论
- 第15章 企业搜索
  - 15.1 介绍
    - 15.1.1 企业搜索的特点和应用
    - 15.1.2 企业搜索软件
    - 15.1.3 工作场所搜索
  - 15.2 企业搜索任务

## &lt;&lt;现代信息检索&gt;&gt;

- 15.2.1 搜索支持任务的例子
- 15.2.2 搜索类型
- 15.2.3 研究企业搜索
- 15.3 企业搜索系统的结构
  - 15.3.1 收集
  - 15.3.2 提取
  - 15.3.3 索引
  - 15.3.4 文本注释的索引
  - 15.3.5 查询处理
  - 15.3.6 搜索结果的展示
  - 15.3.7 安全模型
  - 15.3.8 联合/元搜索
- 15.4 企业搜索评价
  - 15.4.1 企业搜索的公开测试集
  - 15.4.2 企业搜索内部评价
  - 15.4.3 企业搜索调试
  - 15.4.4 所能期待的是什么
- 15.5 不满意的可能原因
- 15.6 情境化和个性化
  - 15.6.1 情境化的控制和工具
  - 15.6.2 情境化：本地、企业或全球
  - 15.6.3 轮廓的隐私
  - 15.6.4 定义、建立和维护轮廓
  - 15.6.5 用户建模
  - 15.6.6 隐式评价
  - 15.6.7 信息过滤
  - 15.6.8 社会化推荐系统
- 15.7 趋势和研究问题
- 15.8 文献讨论
- 第16章 图书馆系统
  - 16.1 图书馆的信息环境
  - 16.2 联机公共检索目录
    - 16.2.1 OPAC和书目记录
    - 16.2.2 来自ILS的信息检索
    - 16.2.3 混合图书馆的整合
    - 16.2.4 OPAC和最终用户
    - 16.2.5 ILS：供应商和产品
  - 16.3 信息检索系统与文档数据库
    - 16.3.1 书目和全文数据库
    - 16.3.2 数据库记录的内容
    - 16.3.3 联机产业：数据库供应商
    - 16.3.4 来自文档数据库的信息检索
  - 16.4 组织机构内部的信息检索
  - 16.5 趋势和研究问题
  - 16.6 文献讨论
- 第17章 数字图书馆
  - 17.1 介绍

<<现代信息检索>>

- 17.2 定义数字图书馆
- 17.3 通用架构
- 17.4 基本概念
  - 17.4.1 数字对象和馆藏
  - 17.4.2 元数据和目录
  - 17.4.3 资源库/档案库
  - 17.4.4 服务
- 17.5 社会经济问题
  - 17.5.1 社会问题
  - 17.5.2 经济问题
- 17.6 软件系统
  - 17.6.1 Greenstone
  - 17.6.2 Eprints
  - 17.6.3 DSpace
  - 17.6.4 Fedora
  - 17.6.5 ODL
  - 17.6.6 5S套件
- 17.7 数字图书馆案例研究
  - 17.7.1 联网学位论文数字图书馆
  - 17.7.2 国家科学数字图书馆
  - 17.7.3 ETANA-DL考古数字图书馆
- 17.8 趋势和研究问题
  - 17.8.1 评价
  - 17.8.2 集成
  - 17.8.3 其他研究挑战
- 17.9 文献讨论
- 附录A 开源搜索引擎
- 附录B 作者简介
- 参考文献
- 索引

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>