<<信息检索导论>>

图书基本信息

书名：<<信息检索导论>>

13位ISBN编号：9787115218247

10位ISBN编号：7115218242

出版时间：2010-1

出版时间：人民邮电出版社

作者：（美）曼宁，（美）拉哈万，（德）舒策　著

页数：482

字数：605000

前言

As recently as the 1990s, studies showed that most people preferred getting information from other people rather than from information retrieval OR） systems. Of course, in that time period, most people also used human travel agents to book their travel. However, during the last decade, relentless opti- mization of information retrieval effectiveness has driven web search engines to new quality levels at which most people are satisfied most of the time, and web search has become a standard and often preferred source of information finding. For example, the 2004 Pew Internet Survey （Fallows 2004） found that "92% of Internet users say the Internet is a good place to go for getting everyday information." To the surprise of many, the feld of information re- trieval has moved from being a primarily academic discipline to being the basis underlying most peoples preferred means of information access. This book presents the scientific underpinnings of this field, at a level accessible to graduate students as well as advanced undergraduates. Information retrieval did not begin with the Web. In response to various challenges of providing information access, the field of IR evolved to give principled approaches to searching various forms of content. The field be- gan with scientific publications and library records but soon spread to other forms of content, particularly those of information professionals, such as journalists, lawyers, and doctors. Much of the scientific research on IR has occurred in these contexts, and much of the continued practice of IR deals with providing access to unstructured information in various corporate and governmental domains, and this work forms much of the foundation of our book.

<<信息检索导论>>

内容概要

本书是信息检索的教材，旨在从计算机科学的视角提供一种现代的信息检索方法。书中从基本概念讲解网络搜索以及文本分类和文本聚类等，对收集、索引和搜索文档系统的设计和实现的方方面面、评估系统的方法、机器学习方法在文本收集中的应用等给出了最新的讲解。

书中所有重要的思想都是用示例进行解释，图文并茂。

本书非常适合作为计算机科学及相关专业的高年级本科生和研究生的“信息检索”课程的入门教材，当然也同样适合研究人员和专业人士阅读。

作者简介

Christopher D.Manning，斯坦福大学语言学博士，现任斯坦福大学计算机科学和语言学副教授，主要研究方向是统计自然语言处理、信息提取与表示、文本理解和文本挖掘等。

Prabhakar Raghavan，加州大学伯克利分校博士，现任Yahoo！实验室主任，斯坦福大学计算机科学系顾问教授，是ACM和IEEE会士。主要研究兴趣是文本及Web数据挖掘、算法设计等。此前，他曾任Verity公司CTO，并在旧M研究院担任过管理工作。

Hinrich Schuze斯坦福大学博士，现任斯图加特大学自然语言处理研究所理论计算语言学主任。他在美国硅谷工作过多年，曾在施乐Palo Alto研究中心供职，担任过Outride公司（后被Google公司收购）副总裁，做过Novation生物科技公司CTO和Enkata公司首席科学家。

<<信息检索导论>>

书籍目录

章节摘录

An example information retrieval problem A fat book that many people own is Shakespeares Collected Works.Suppose you wanted to determine which plays of Shakespeare contain the words Brutus AND Caesar AND NOT Calpurnia.One way to do that is to start at the beginning and to read through all the text, noting for each play whether it contains Brutus and Caesar and excluding it from consideration if it contains Calpurnia.The simplest form of document retrieval is for a computer to do this sort of linear scan through documents.This process is commonly referred to as grepping through text, after the Unix command g r e p, which performs this process.Grepping through text can be a very effective process, especially given the speed of modem computers , and often allows useful possibilities for wildcard pattern matching through the use of regular expressions.With modem computers.for simple querying of modest collections (the size of Shakespeares Collected Works is a bit under one million words of text in total) , you really need nothing more. But for many purposes, you do need more : 1.To process large document collections quickly.The amount of online data has grown at least as quickly as the speed of computers, and we would now like to be able to search collections that total in the order of biHions to trillions of words. 2.To allow more flexible matching operations.For example, it is impractical to perform the query Romans NEAR countrymen with g r e p, where NEAR might be defined as within 5 words or within the same sentence ?

3.To allow ranked retrieval.In many cases, you want the best answer to an information need among many documents that contain certain words. The way to avoid linearly scanning the texts for each query is to index the documents in advance.Let us stick with Shakespeares Collected Works, and use it to introduce the basics of the Boolean retrieval model.Suppose we record foreachdocument—here aplayofShakespeare ' s —whetheritcontainseach word out of all the words Shakespeare used (Shakespeare used about 32, 000 different words) .The result is a binary term—document incidence matrix,as in Figure 1.1.Terms are the indexed units (further discussed in Section 2.2) ;they are usuany words,and for the moment you can think of them as wordsf but the information retrieval literature normally speaks of terms because some of them, such as perhaps I-9 or Hong Kong are not usuaHy thought of as words.

媒体关注与评论

“如何排定SVM、XML、DNS和LSI的顺序？
什么是信息检索中的垃圾信息、隐藏页和门页？
MapReduce和其他一些并行运算方法是如何实现由兆字节（MB）到百万兆字节（PB）的飞跃的？
这些问题在本书中您都能找到答案，本书首次将构建Web搜索引擎的复杂过程以一种清晰的全景方式展现给读者。
”——Peter Norving，Google公司研究主管 “本书将信息检索这个举足轻重而又发展迅猛的领域进行了全面、新颖、准确的介绍，我们非常需要这样一本教科书。
”——Raymond J.Mooney，得克萨斯大学奥斯汀分校教授 “此书内容新颖，选材独特，对信息检索的基础知识和发展方向进行了生动的描述。
”——Jon Kleinberg，康奈尔大学教授

<<信息检索导论>>

编辑推荐

《信息检索导论(英文版)》从计算机科学领域的角度出发，介绍了信息检索的基础知识，并对当前信息检索的发展做了回顾，重点介绍了搜索引擎的核心技术，如文档分类和文档聚类问题，以及机器学习和数值计算方法。
书中所有重要的思想都用示例进行了解释，生动形象，引人入胜，实现了理论与实战的完美结合。
《信息检索导论(英文版)》的三位作者均是信息检索领域的顶级专家，两位来自学术教育界，一位来自硅谷业界，使《信息检索导论(英文版)》既具备深厚的理论基础，又代表了尖端科技水准。因此，该书甫一出版，即被奉为该领域的权威著作，备受瞩目，目前已被众多世界名校采用为信息检索课程的教材。

版权说明