

# Information Retrieval

# 信息检索

潘微科

# Outline

- 教材
- 参考教材
- 教学内容
- 相关书籍
- 相关课程
- 小结

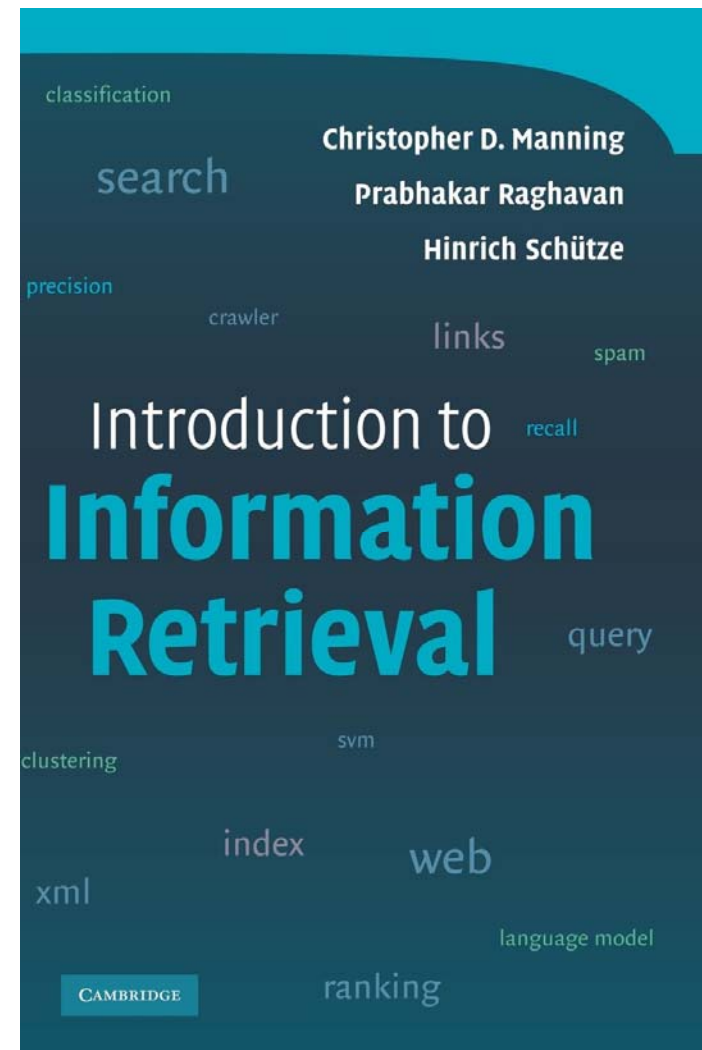
# 教材

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. **Introduction to Information Retrieval**. Cambridge University Press, 2008. ISBN:0521865719.
- 注: 作者分享了**电子版教材**等资源 <https://nlp.stanford.edu/IR-book/>

← → ↻ nlp.stanford.edu/IR-book/

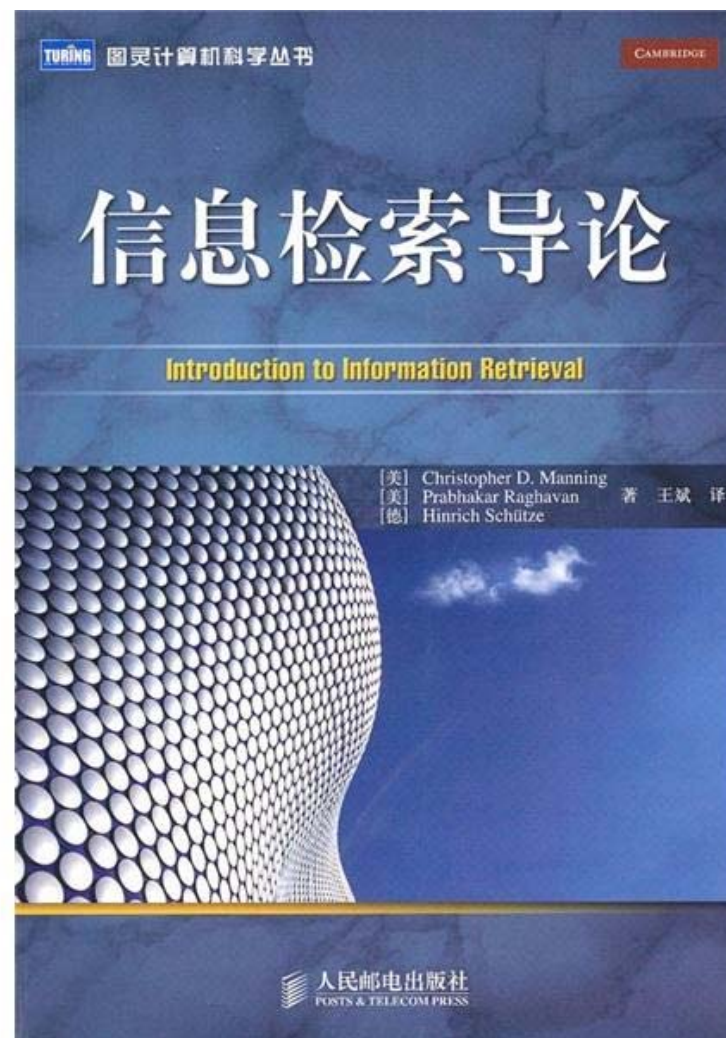
Introduction to Information Retrieval: Table of Contents

	chapter
Front matter (incl. table of notations)	<a href="#">pdf</a>
01 Boolean retrieval	<a href="#">pdf</a> <a href="#">html</a>
02 The term vocabulary & postings lists	<a href="#">pdf</a> <a href="#">html</a>
03 Dictionaries and tolerant retrieval	<a href="#">pdf</a> <a href="#">html</a>
04 Index construction	<a href="#">pdf</a> <a href="#">html</a>
05 Index compression	<a href="#">pdf</a> <a href="#">html</a>
06 Scoring, term weighting & the vector space model	<a href="#">pdf</a> <a href="#">html</a>
07 Computing scores in a complete search system	<a href="#">pdf</a> <a href="#">html</a>
08 Evaluation in information retrieval	<a href="#">pdf</a> <a href="#">html</a>
09 Relevance feedback & query expansion	<a href="#">pdf</a> <a href="#">html</a>
10 XML retrieval	<a href="#">pdf</a> <a href="#">html</a>
11 Probabilistic information retrieval	<a href="#">pdf</a> <a href="#">html</a>
12 Language models for information retrieval	<a href="#">pdf</a> <a href="#">html</a>
13 Text classification & Naive Bayes	<a href="#">pdf</a> <a href="#">html</a>
14 Vector space classification	<a href="#">pdf</a> <a href="#">html</a>
15 Support vector machines & machine learning on documents	<a href="#">pdf</a> <a href="#">html</a>
16 Flat clustering	<a href="#">pdf</a> <a href="#">html</a>
17 Hierarchical clustering	<a href="#">pdf</a> <a href="#">html</a>
18 Matrix decompositions & latent semantic indexing	<a href="#">pdf</a> <a href="#">html</a>
19 Web search basics	<a href="#">pdf</a> <a href="#">html</a>
20 Web crawling and indexes	<a href="#">pdf</a> <a href="#">html</a>
21 Link analysis	<a href="#">pdf</a> <a href="#">html</a>
Bibliography & Index	<a href="#">pdf</a>
bibtex file	<a href="#">bib</a>



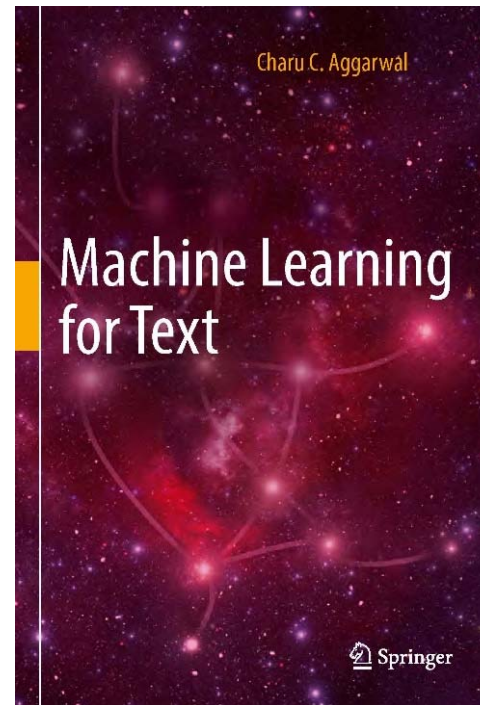
# 参考教材(1/2)

- 王斌(译). 信息检索导论. 人民邮电出版社. 2010.9.  
ISBN:9787115234247.
- 注: 是英文教材的中译本.



# 参考教材(2/2)

- Charu C. Aggarwal. **Machine Learning for Text**. Springer, 2018. ISBN: 9783319735306.
- 注:
  - Chapter 9: **Information Retrieval and Search Engines**
  - 黎琳,潘微科,明仲 (译). **文本机器学习**. 机械工业出版社. 2020.5. ISSN:9787111648055.

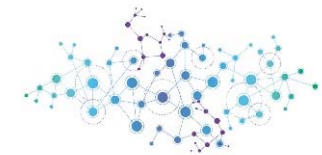


 Springer

## 文本机器学习

Machine Learning for Text

[美] 查鲁·C.阿加瓦尔 (Charu C. Aggarwal) 著  
黎琳 潘微科 明仲 译



 机械工业出版社  
CHINA MACHINE PRESS

# 教学内容(1/5)

- 第1章 布尔检索 (Boolean retrieval)
- 第2章 词项词典及倒排记录表 (The term vocabulary & postings lists)
- 第3章 词典及容错式检索 (Dictionaries and tolerant retrieval)
- 第4章 索引构建 (Index construction)
- 第5章 索引压缩 (Index compression)

# 教学内容(2/5)

- 第6章 文档评分、词项权重计算及向量空间模型 (Scoring, term weighting & the vector space model)
- 第7章 一个完整搜索系统中的评分计算 (Computing scores in a complete search system)
- 第8章 信息检索的评价 (Evaluation in information retrieval)

# 教学内容(3/5)

- 第9章 相关反馈及查询扩展 (Relevance feedback & query expansion)
- 第10章 XML检索 (XML retrieval) 注:本章不作要求
- 第11章 概率检索模型 (Probabilistic information retrieval)
- 第12章 基于语言建模的信息检索模型 (Language models for information retrieval)



# 教学内容(4/5)

- 第13章 文本分类及朴素贝叶斯方法 (Text classification & Naive Bayes)
- 第14章 基于向量空间模型的文本分类 (Vector space classification) 注: 本章不作要求
- 第15章 支持向量机及文档机器学习方法 (Support vector machines & machine learning on documents) 注: 本章不作要求
- 第16章 扁平聚类 (Flat clustering) 注: 本章不作要求
- 第17章 层次聚类 (Hierarchical clustering) 注: 本章不作要求
- 第18章 矩阵分解及隐性语义索引 (Matrix decompositions & latent semantic indexing)

# 教学内容(5/5)

- 第19章 Web搜索基础 (Web search basics)
- 第20章 Web采集及索引 (Web crawling and indexes)
- 第21章 链接分析 (Link analysis)

## 相关书籍(1/8)

- Ricardo Baeza-Yates, Berthier Ribeiro-Neto. **Modern Information Retrieval: The Concepts and Technology Behind Search (Second Edition)**. Addison-Wesley Professional, 2011. ISBN-10:0321416910, ISBN-13:978-0321416919. 注:国内有影印版(机械工业出版社, ISBN:9787111331742).
- 黄萱菁,张奇,邱锡鹏 (译). 现代信息检索(原书第**2**版). 机械工业出版社, 2012.10. ISBN:9787111385998.

## 相关书籍(2/8)

- W. Bruce Croft, Donald Metzler and Trevor Strohman. **Search Engines: Information Retrieval in Practice**. Pearson, 2009.2. ISBN-10:0136072240, ISBN-13:978-0136072249. 注:国内有影印版(机械工业出版社,ISBN:9787111282471). 有免费电子版<http://ciir.cs.umass.edu/irbook/> 其他资源<http://www.search-engines-book.com/>
- 刘挺,秦兵,张宇,车万翔(译). 信息检索实践. 机械工业出版社, 2010.6. ISBN:9787111288084.

## 相关书籍(3/8)

- Stefan Büttcher, Charles L. A. Clarke, Gordon V. Cormack. **Information Retrieval: Implementing and Evaluating Search Engines**. MIT Press, 2010. ISBN-10:0262026511, ISBN-13:978-0262026512.
- 陈健,黄晋(译). 信息检索:实现和评价搜索引擎. 机械工业出版社, 2012.1. ISBN:9787111359906.

## 相关书籍(4/8)

- Ian H. Witten, Alistair Moffat, Timothy C. Bell. **Managing Gigabytes: Compressing and Indexing Documents and Images (Second Edition)**. ISBN:9781558605701.
- 梁斌, 杨青(译). 管理海量数据:压缩、索引和查询. 电子工业出版社. 2014.1. ISBN:9787121219337.

## 相关书籍(5/8)

- Christopher D. Manning and Hinrich Schütze. **Foundations of Statistical Natural Language Processing**. MIT Press, Cambridge, MA, 1999.5. ISBN:9787505399211.
- 苑春法,李庆中,王昀,李伟,曹德芳 等(译). 统计自然语言处理基础. 电子工业出版社, 2005.1. ISBN:9787505399211.

## 相关书籍(6/8)

- David A.Grossman, Ophir Frieder. **Information Retrieval Algorithms and Heuristics (Second Edition)**. Springer, 2004. ISBN-10:1402030045, ISBN-13:978-1402030048. 注:国内有影印版(人民邮电出版社,ISBN:9787115212252).



## 相关书籍(7/8)

- Erik Hatcher, Otis Gospodnetic and Mike McCandless. **Lucene in Action (Second Edition)**. Manning Publications, 2010.7. ISBN-10:1933988177, ISBN-13:978-1933988177.
- 牛长流,肖宇(译). **Lucene实战(第2版)**. 人民邮电出版社, 2011.6. ISBN:9787115251770.

## 相关书籍(8/8)

- 成龙. **Lucene**搜索引擎开发进阶实战. 机械工业出版社, 2015.1. ISBN:9787111488422.
- 罗刚等. 解密搜索引擎技术实战: **Lucene & Java**精华版(第2版). 电子工业出版社, 2014.5. ISBN:9787121217326.
- ...

# 相关课程(1/8)

- **Information Retrieval** (University of Munich, by Hinrich Schütze).
  - <http://www.cis.lmu.de/~hs/teach/14s/ir/>
  - 本课程所采用的讲义是在Prof. Hinrich Schütze提供的Latex讲义的基础上修改的
  - Hinrich Schütze是**Introduction to Information Retrieval**的第三作者

## 相关课程(2/8)

- **Information Retrieval and Web Search** (Stanford University, by Christopher D. Manning and Pandu Nayak)
  - <http://www.stanford.edu/class/cs276/>
  - Christopher D. Manning是**Introduction to Information Retrieval**的第一作者

## 相关课程(3/8)

- **Information Retrieval and Web Search** (The University of Texas at Austin, by Raymond J. Mooney)
  - <https://www.cs.utexas.edu/~mooney/ir-course/>

## 相关课程(4/8)

- **Information Retrieval** (University of California, Berkeley, by Ray Larson)
  - <http://courses.ischool.berkeley.edu/i240/s13/>

## 相关课程(5/8)

- **Information Retrieval and Web Agents** (Johns Hopkins University, by David Yarowsky)
  - <http://www.cs.jhu.edu/~yarowsky/cs466.html>

# 相关课程(6/8)

- 信息检索 (中国科学院大学 计算机科学与技术学院, by 何苯教授)
  - <https://scceucas.ac.cn/index.php/zh-CN/jxpy/1148-1paivrduetz0n5311>
  - 教材: 王斌(译). 信息检索导论. 人民邮电出版社. 2010.9.
  - 参考教材: Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. **Introduction to Information Retrieval**. Cambridge University Press, 2008.
  - 教学内容: 信息检索概论、检索系统结构、检索系统评价、信息检索模型、相关反馈、基于上下文的信息检索模型、网络搜索、网络爬虫



## 相关课程(7/8)

- 信息检索 (清华大学 计算机科学与技术系 by 孙茂松教授)

## 相关课程(8/8)

- 信息检索 (武汉大学 黄如花教授) at 爱课程 (中国大学MOOC)
  - <https://www.icourse163.org/course/WHU-29001>
  - 仅部分相关
- **Web**信息检索与数据抓取 (北京信息职业技术学院 郑淑晖副教授、伊新讲师) at 智慧职教
  - [https://www.icve.com.cn/portal\\_new/courseinfo/courseinfo.html?courseid=gw3calmnqb9csho7fnqryq](https://www.icve.com.cn/portal_new/courseinfo/courseinfo.html?courseid=gw3calmnqb9csho7fnqryq)
  - 仅部分相关
- 信息检索与推荐系统 (武汉理工大学 李琳) at 学堂在线
  - <https://www.xuetangx.com/course/whutP0854004163/10331420>
  - 仅部分相关

# 小结

- 教材
- 参考教材
- 教学内容
- 相关书籍
- 相关课程