

Information Retrieval

Matthias Hagen

Martin Potthast

Benno Stein

Contents

- I. Introduction
- II. Indexing
- III. Retrieval Models
- IV. User Interface
- V. Evaluation
- VI. IR Applications

Learning Objectives

- ❑ Understand a retrieval system's components and explain their interactions
- ❑ Understand and compare text preprocessing options
- ❑ Understand and compare options for retrieval system indexes
- ❑ Acquire a solid background in the theory of retrieval models
- ❑ Understand and apply the principles and practice of retrieval evaluation
- ❑ Implement and evaluate retrieval systems in practice

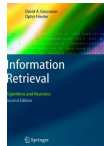
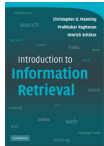
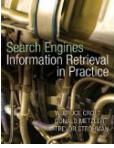
Related Fields

1. Statistics [paradigms, models]
2. Mathematics
3. Data Mining [methods, algorithms]
4. Machine Learning
5. Natural Language Processing
6. Knowledge Processing
7. Search Engines [applications]
8. Recommender Systems
9. Decision Support Systems

Literature

Information Retrieval:

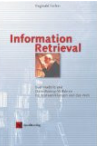
- ❑ W.B. Croft, D. Metzler, T. Strohman.
Search Engines: Information Retrieval in Practice
Pearson 2009. ciir.cs.umass.edu/downloads/SEIRiP.pdf
- ❑ R. Baeza-Yates, B. Ribeiro-Neto.
Modern Information Retrieval: The Concepts and Technology behind Search
Pearson 2011.
- ❑ S. Büttcher, C.L.A. Clarke, G.V. Cormack.
Information Retrieval: Implementing and Evaluating Search Engines
MIT Press 2010.
- ❑ C.D. Manning, P. Raghavan, H. Schütze.
Introduction to Information Retrieval
Cambridge University Press 2008. nlp.stanford.edu/ir-book/
- ❑ P. Ingwersen, K. Järvelin.
The Turn: Integration of Information Seeking and Retrieval in Context
Springer 2005.
- ❑ D.A. Grossman, O. Frieder.
Information Retrieval: Algorithms and Heuristics
Springer 2004.



Literature

Information Retrieval: (continued)

- ❑ R. Ferber.
Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web
dpunkt 2003. information-retrieval.de/irb/ir.html
- ❑ I.H. Witten, A. Moffat, T.C. Bell.
Managing Gigabytes: Compressing and Indexing Documents and Images
Morgan Kaufmann 1999.
- ❑ G. Salton, M.J. McGill.
Introduction to Modern Information Retrieval
McGraw-Hill 1983.
- ❑ C.J. van Rijsbergen.
Information Retrieval
Butterworths 1979. www.dcs.gla.ac.uk/keith/preface.html



Data Mining:

- ❑ S. Chakrabarti.
Mining the Web: Discovering Knowledge from Hypertext Data
Morgan Kaufmann 2003.



Literature

“Core” information retrieval conferences: [[The Information Retrieval Anthology](#)]

- ❑ SIGIR, ECIR, AIRS, ICTIR, CHIIR, DESIRES, TREC, CLEF, NTCIR, FIRE
- ❑ WWW, WSDM, CIKM, RecSys, SPIRE

Conferences from related fields:

- ❑ ACL, EMNLP, NAACL, COLING, CoNLL, AACL, EACL
- ❑ JCDL, TPDL
- ❑ ASIST, iConference
- ❑ CHI, CUI, IUI, UMAP
- ❑ KDD, SDM, ICDM, PAKDD
- ❑ NeurIPS, ICML, ECML
- ❑ AAI, IJCAI, ECAI
- ❑ WebScience, Hypertext
- ❑ ICWSM, ASONAM

Journals:

- ❑ ACM SIGIR Forum, TOIS, TIST, and TWEB
- ❑ Elsevier Information Processing and Management (IPM)
- ❑ Springer Information Retrieval Journal (IRJ)
- ❑ Wiley Journal of the American Society for Information Science and Technology (JASIST)

Monograph series:

- ❑ Foundations and Trends in Information Retrieval (now publisher)
- ❑ The Information Retrieval Series (Springer)
- ❑ Synthesis Lectures on Information Concepts, Retrieval, and Services (Morgan & Claypool)

Software

Industry (Open Source):

- ❑ The Apache Software Foundation
Lucene: Ultra-fast Search Library and Server
lucene.apache.org
- ❑ The Apache Software Foundation
Solr
lucene.apache.org/solr/
- ❑ Elasticsearch BV, The Netherlands
Elasticsearch
elastic.co/elasticsearch
- ❑ Vespa, Xapian, Sphinx, ...

Services:

- ❑ Common Crawl Foundation, USA
Common Crawl
commoncrawl.org
- ❑ Bauhaus-Universität Weimar, Friedrich-Schiller-Universität Jena, and Leipzig University, Germany
ChatNoir Search
chatnoir.eu

Software

Research:

- ❑ Iadh Ounis, Giani Amati, Craig Macdonald, et al.
Terrier IR Platform
terrier.org, github.com/terrier-org/pyterrier
- ❑ Bruce Croft, Jamie Callan, et al.
The Lemur Project
lemurproject.org
- ❑ Peilin Yang, Hui Fang, and Jimmy Lin
Anserini
anserini.io, pyserini.io
- ❑ Andrew Yates, Jimmy Lin, et al.
Capreolus
capreolus.ai
- ❑ Jimmy Lin et al.
PyGaggle
pygaggle.ai
- ❑ Nils Reimers and Iryna Gurevych
Sentence Bert
sbert.net
- ❑ [Matchmaker](#), [MatchZoo](#), [OpenMatch](#), [OpenNIR](#), [Tevatron](#), [Zettair](#), . . .

Software

Research:

- ❑ Alexander Bondarenko, Maik Fröbe, Jan Heinrich Reimers, Matthias Hagen, et al.
ir_axioms
github.com/webis-de/ir_axioms/
- ❑ Sean MacAvaney, Andrew Yates, Sergey Feldman, et al.
ir_datasets
ir-datasets.com
- ❑ Sean MacAvaney, Craig Macdonald, Benjamin Piwowarski, and Eugene Yang
ir-measures
ir-measur.es, github.com/terrierteam/ir_measures
- ❑ NIST
trec_eval
github.com/usnistgov/trec_eval
- ❑ Joao Palotti, Harrisen Scells, and Guido Zuccon
TrecTools
github.com/joaopalotti/trectools
- ❑ Antonio Mallia, Michal Siedlaczek, Joel Mackenzie, and Torsten Suel
PISA
github.com/pisa-engine/pisa