Information Retrieval

Matthias Hagen   Martin Potthast   Benno Stein

Contents

Objectives

❑ understand information systems, and search engines in particular

❑ acquire solid background on the theory of retrieval models

❑ implement and apply retrieval algorithms in practice

❑ comprehend practical considerations, enabling comparative evaluation

❑ be able to educate oneself

Related Fields

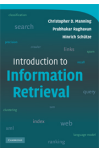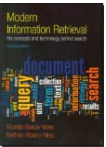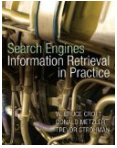1. Statistics   [paradigms, models]

2. Mathematics

3. Data Mining   [methods, algorithms]

4. Machine Learning

5. Natural Language Processing

6. Knowledge Processing

7. Search Engines   [applications]

8. Business Intelligence

9. Decision Support Systems
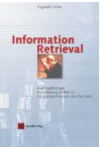
Literature

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:

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

Conferences from related fields:

- ❑ ACL, COLING, CoNLL, EMNLP, NAACL
- ❑ KDD, SDM, ICDM, PAKDD
- ❑ ASIST, iConference
- ❑ ICML, ECML, NeurIPS
- ❑ AAAI, IJCAI, ECAI
- ❑ JCDL, TPDL

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/products/elasticsearch

- ❑ Vespa, Xapian, Sphinx, . . .

Research:

- ❑ I. Ounis, G. Amati, C. Macdonald, and others
  Terrier IR Platform
  terrier.org

- ❑ B. Croft, J. Callan, and others
  The Lemur Project
  lemurproject.org

- ❑ Anserini, Wumpus, Zettair, . . .

Software

Services:

- ❑ Common Crawl Foundation, USA
  Common Crawl
  commoncrawl.org

- ❑ Bauhaus-Universität Weimar, MLU Halle-Wittenberg, and Leipzig University, Germany
  ChatNoir Search
  chatnoir.eu