Informatics 225
Computer Science 221

Information Retrieval

Lecture 1

These course materials borrow, with permission, from those of Prof. Cristina Videira Lopes, Prof. Alberto Krone-Martins, Addison Wesley 2008, Chris Manning, Pandu Nayak, Hinrich Schütze, Heike Adel, Sascha Rothe, Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie. Powerpoint theme by Prof. André van der Hoek.

Today

- Logistics
- Course goals
- Introduction

Logistics

- Instructor: Pramit Choudhary <pchoudha@uci.edu>

https://www.linkedin.com/in/pramitc/

https://twitter.com/MaverickPramit

Logistics

- Reader: Leming Cheng “Jay”, lemingc@uci.edu

Logistics

- The course is on Canvas

- **Lectures:** virtually via Zoom and hopefully in-person, but will be recorded and placed online in Canvas
  - Some lectures may be pre-recorded and videos posted in Canvas. Others may be in Zoom.
    - Pay attention to announcements!
  - Lectures videos will be left online during the entire quarter.
  - Slides will be online during the entire quarter.
- **Discussion Sessions:**
  - We will mostly use Ed Discussion
  - Two in Zoom
  - Two in-person (if possible)
- **Quizzes/Exam:** synchronous at the scheduled times
  - Everyone: in class! (bring your laptop)
  - Students with official remote accommodations: online, at the same time

- **Final Grade:** Projects (75%) + Quizzes (25%)

Logistics

- The course is on Canvas

- **Lectures:** virtually via Zoom and hopefully in-person, but will be recorded and placed online in Canvas
  - Some lectures may be pre-recorded and videos posted in Canvas. Others may be in Zoom.
    - Pay attention to announcements!
  - Lectures videos will be left online during the entire quarter.
  - Slides will be online during the entire quarter.
- **Discussion Sessions:**
- We will mostly use Ed Discussion
  - We will mostly use Ed Discussions
  - Two in Zoom
  - Two in-person
- **Quizzes/Exam:** synchronous at the scheduled times
  - Everyone: in class! (bring your laptop)
  - Students with official remote accommodations: online, at the same time

Quizzes might not work on phone!

- **Final Grade:** Projects (75%) + Quizzes (25%)

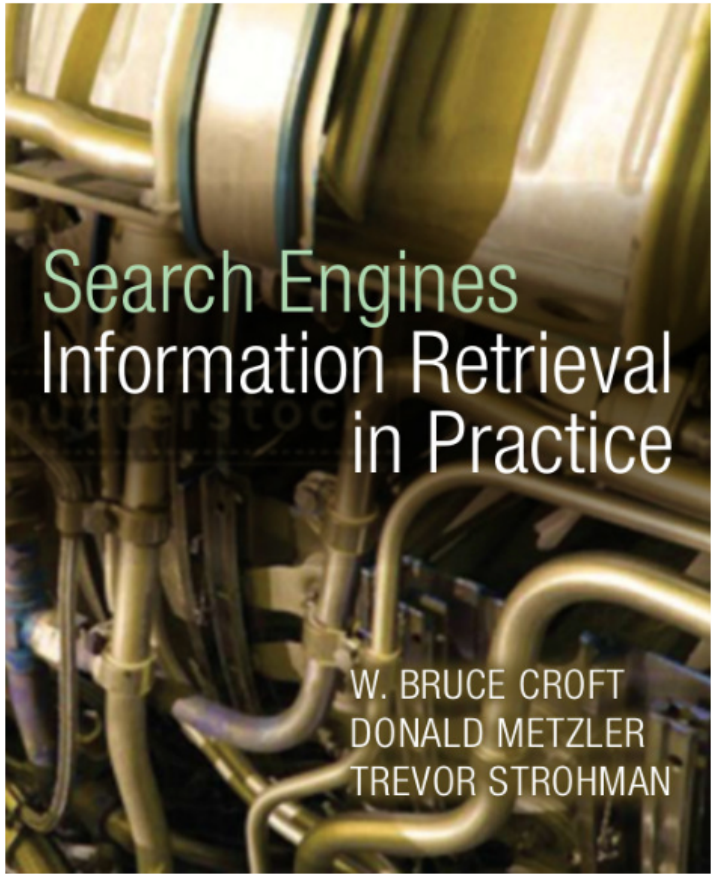
Required book

Search Engines Information Re

ciir.cs.umass.edu/irbook/

DOWNLOAD HERE

PDF



BOOK DESCRIPTION

This book provides an overview of the important issues in information retrieval, and how those issues affect the design and implementation of search engines. Not every topic is covered at the same level of detail. The focus is on some of the most important alternatives to implementing search engine components and the information retrieval models underlying them. The target audience for the book is advanced undergraduates in computer science, although it is also a useful introduction for graduate students.

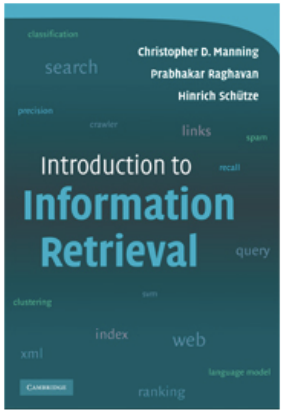
This version of the book is being made available for free download. It has been edited to correct the minor errors noted in the 5 years since the book's publication. The authors, meanwhile, are working on a second edition.

Additional book

Introduction to Information Ret

nlp.stanford.edu/IR-book/information-retrieval-book.html

Introduction to Information Retrieval



This is the companion website for the following book.

Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.

You can order this book at CUP, at your local bookstore or on the internet. The best search term to use is the ISBN: 0521865719.

The book aims to provide a modern approach to information retrieval from a computer science perspective. It is based on a course we have been teaching in various forms at Stanford University, the University of Stuttgart and the University of Munich.

We'd be pleased to get feedback about how this book works out as a textbook, what is missing, or covered in too much detail, or what is simply wrong. Please send any feedback or comments to: informationretrieval (at) yahoogroups (dot) com

Online resources

Apart from small differences (mainly concerning copy editing and figures), the online editions should have the same content as the print edition.

The following materials are available online. The date of last update is given in parentheses.

Goals

- For you to know the fundamentals of information retrieval

Goals

- For you to know the fundamentals of information retrieval
- For you to have at your disposal some techniques to find information spread among 10^n documents

Goals

- For you to know the fundamentals of information retrieval
- For you to have at your disposal some techniques to find information spread among 10^n documents
- For you to practice Web crawling and **to build your own simple search engine from scratch**

This course answers three primary questions

- What is information retrieval?
- How to build some of the most important blocks of information retrieval systems?
- How to perform information retrieval in the web?

Structure of the course

- Online asynchronous lectures
- Discussion: use the help of the TAs to complete your assignments, to clarify doubts before quizzes and the final exam
- Three assignments
  - 75% of the total grade
- Four graded quizzes
  - 25% of the total grade
  - The lowest grade will be dropped – so **no make up quiz** if you get sick/cannot attend/forget/etc.
- Typically, I don't curve the grades.
  - But reserve the right to do it if necessary.

Basic tenor of the class

- **Creativity**
- Some discussion of ideas between the students in Ed Discussion
  - Perhaps with disagreements
  - But always with constructive criticism
  - Sometimes there is more than one right answer…
    - **But there are many wrong answers!**

Additional remark

Special needs: *Students who may need an accommodation due to a disability should contact the UCI Disability Services Center at (949) 824-7494 as soon as possible to explore the possible range of accommodations.*

We encourage all students having difficulty, whether or not due to a disability, to consult privately with the instructor at any time.

Questions?

Please, write any questions in Ed Discussion!

Our definition of information retrieval

The activity of searching and extracting information from a collection of information resources.

In **this class** : textual information from text documents

Quiz 0 – On Canvas today!

You have until 12th Jan,2022, to complete Quiz 0.

Quiz 0 has no wrong answers,
if you provide reasonable answers.

It values zero points but may be used to calibrate the course,
so answering it reasonably and honestly is in your best interest.