Distributed Information Systems
Fall Semester – 2022
CS-423

Time and Place
Lecture: Thursday 10:15 – 12:00, CM3
https://epfl.zoom.us/j/66237387610

Exercise: Thursday 12:15-13:00, CM3

Karl Aberer

Distributed Information Systems Laboratory

Welcome back to a normal semester!

We will be working in hybrid mode

Program of today:

- Hour 1: Everything you need to know on the organization
- Hour 2-3: An overview of the basic concepts related to distributed information systems

Goals of the Course

Understand what is a "**Distributed Information System**"?

- e.g. Web Search Engines, Online Social Networks, etc.

Know which are **key tasks** relevant for DIS?

- e.g. retrieval, mining, recommending, information extraction, data integration etc.

Master **common techniques** used to solve these problems

- e.g. vector space model, graph mining, word embeddings etc.

Pre-existing knowledge not required

Knowledge in databases and machine learning helpful

Focus of the Course

Master important **Models and Algorithms** for representing and processing information:

Data Science

Conceptual foundations to practically use tools and platforms for Data Science

- Complementary to *Applied Data Analysis* by Bob West

Other Related Courses

In synergy with

- Applied Data Analysis

Complementary to

- Introduction to database systems
- Database systems

Some overlaps possible with

- Introduction to machine learning
- Machine learning
- Introduction to natural language processing
- Internet analytics
- …

Which master's program are you from?

1. Computer Science
2. Communications
3. Data Science
4. Cybersecurity
5. Digital Humanities
6. Life Science
7. Electrical Engineering
8. Environmental Science
9. Others

Did you take Applied Data Analysis?

1. Yes
2. No

The Course - Lecture

Standard online ex cathedra lecture

- Lecture streamed via Zoom
  - https://epfl.zoom.us/j/66237387610
  - Zoom QA tool to ask questions
  - Will be answered privately by assistants, or by the lecturer, depending on the questions
- Zoom Quizzes (anonymous)
- Zoom Chat to collect feedbacks

Video recordings

https://tube.switch.ch/channels/PH1KLlGUsX

Materials

Web platform: Moodle

- General announcements will be published on Moodle
- Course notes and exercises will be published on the Web in advance: https://lsir.github.io/DIS/

Exercises

Weekly exercises

- 2-3 problems to solve

Most problems will be (simple) programming exercises

- Uses Python
- Focus on understanding the techniques (not programming skills etc)

Exercises and exam questions from previous years will be made available as well

Exercise Platform

We will be using **Zoom** for communicating with assistants during exercises

Ed Forum to ask questions offline:
https://edstem.org/eu/courses/90/discussion/

- Both among students and with assistants

Continuous control

During the semester

- Midterm programming exercise
- 1 Quiz

Will be graded

Grading

Final Exam: 75%

- Questions similar to the question in exercises and quizzes
- will assume you attended the lecture
- will assume you did the exercises
- examples from earlier years (exercises, exams) provided for preparation

Exam Support

Your computer will be admitted to the exam

- You will have Internet access
- But: communication not allowed (messaging, social platform etc.)
- You can use your notes (paper of electronically, all lecture materials)

Are you planning to join the lecture live or virtually?

1. I join live today, and plan to continue live
2. I join live today, but plan to join virtually
3. I join virtually today, and plan to continue virtually
4. I join virtually today, but plan to join live

Schedule

Week	Date	Cont. Eval.	Area	Topic
1	22 September 2022		**Introduction**	Distributed Information Systems - An Overview
2	29 September 2022		**Information Retrieval**	Information Retrieval Basics
3	06 October 2022			Probabilistic Retrieval and Relevance Feedback
4	13 October 2022			Indexing and Distributed Retrieval
5	20 October 2022			Embedding Models
6	27 October 2022	Prog. midterm		Link-based ranking
7	03 November 2022		**Data Mining**	Graph Mining
8	10 November 2022			Document Classification
9	17 November 2022			Recommender Systems
10	24 November 2022			Association Rules
11	01 December 2022		**From Documents to Knowledge**	Semantic Web
12	08 December 2022			Information Extraction
13	15 December 2022	Quiz		Knowledge Inference
14	22 December 2022			Reserve

Lecturer

Karl Aberer
Head of LSIR

EPFL - I&C - LSIR
BC108
station 14
CH-1015 Lausanne

+41 21 693.46.73
karl-aberer@epfl.ch

Organizational Info

Moodle

- http://moodle.epfl.ch/course/view.php?id=4051

Lecturers

- Prof. Karl Aberer — karl.aberer@epfl.ch — BC 108

Assistants

- Romanou Angelika — angelika.romanou@epfl.ch
- Negar Foroutan — negar.foroutan@epfl.ch
- Mohammadreza Banaei — mohammadreza.banaei@epfl.ch
- Mokhtarian Ehsan — ehsan.mokhtarian@epfl.ch

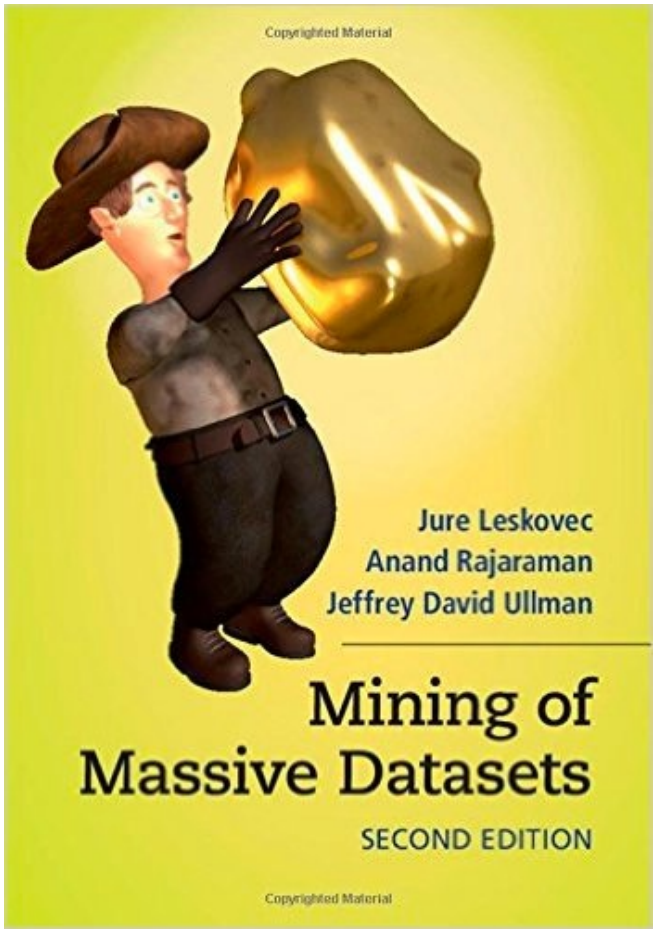
References

Parts of the course are based on the following text books

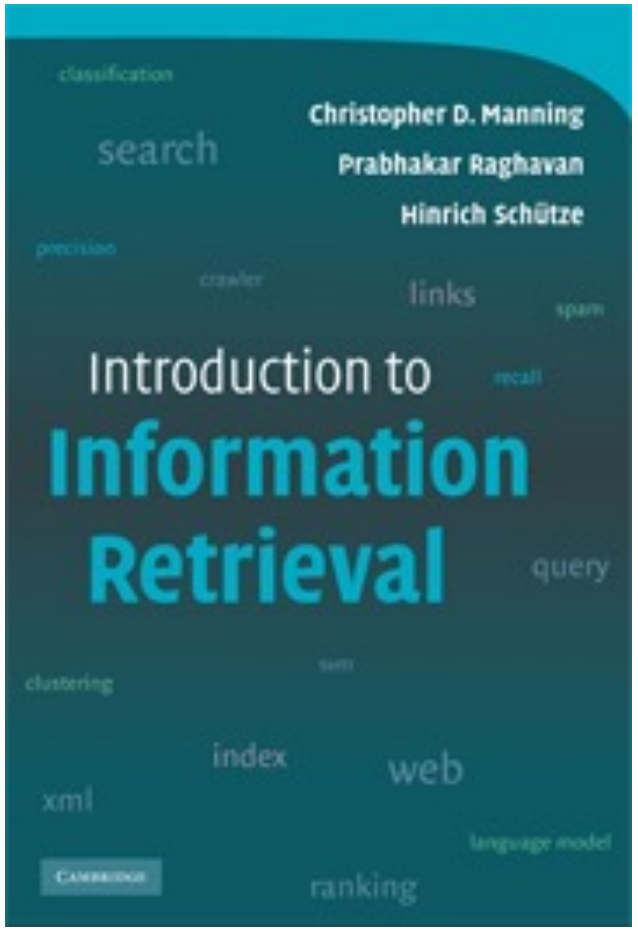
- Ricardo Baeza-Yates, Berthier Ribeiro-Neto, Modern Information Retrieval (Acm Press Series), Addison Wesley, 1999.
- Jiawei Han, Data Mining: concepts and techniques, Morgan Kaufman, 2000.
- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
- J Leskovec, A Rajaraman, JD Ullman, Mining of Massive Datasets, 2014.

Further references to the literature will be given during the lecture

Free books



mmds.org



http://nlp.stanford.edu/IR-book/

Exam Date