Distributed Information Systems
Fall Semester – 2021
CS-423

Time and Place
Lecture: Monday 13:15-15:00, BCH 2201
https://epfl.zoom.us/j/96988744528

Exercise: Monday 15:15-16:00, BCH 2201

Karl Aberer

Distributed Information Systems Laboratory

Welcome to (another) Special Semester!

Everything is different

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 main concepts that will be covered during the semester

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 rule 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 masters 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

Live lecture will be transmitted via webinar
https://epfl.zoom.us/j/96988744528

- Standard online ex cathedra lecture
- Use QA tool to ask questions
  - Will be answered privately by assistants, or by the lecturer, depending on the questions
- Quizzes using Zoom (anonymous)

Video recording
https://tube.switch.ch/channels/45c71cb4

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

In addition, we will use **Piazza** for posing questions and discussions

- Both among students and with assistants

“Continuous control”

Due to the current situation not clear whether graded continuous control is possible

But

- Midterm programming exercise
- 2 Quizzes

Will allow to test your skills

Grading

Final Exam: 100%

- 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. Live
2. Virtually

Schedule

Week	Date	Cont. Eval.	Area	Topic
1	20 September 2021			*Holiday*
2	27 September 2021		**Introduction**	Distributed Information Systems - An Overview
3	04 October 2021		**Information Retrieval**	Basic Text Retrieval Models
4	11 October 2021			Probabilistic Retrieval and Relevance Feedback
5	18 October 2021	Prog. midterm		Indexing and Distributed retrieval
6	25 October 2021			Word Embeddings
7	26 October 2021			Link-based ranking
8	02 November 2021		**Data Mining**	Graph Mining
9	09 November 2021	Quiz		Document classification
10	16 November 2021			Recommender Systems
11	23 November 2021			Association Rules
12	30 November 2021		**From Documents to Knowledge**	Semantic Web
13	07 December 2021	Quiz		Entity and Information Extraction
14	14 December 2021			Inference for Knowledge Graphs

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
- Tugrulcan Elmas — tugrulcan.elmas@epfl.ch
- Mokhtarian Ehsan — ehsan.mokhtarian@epfl.ch
- Farhadkhani Sadegh — sadegh.farhadkhani@epfl.ch

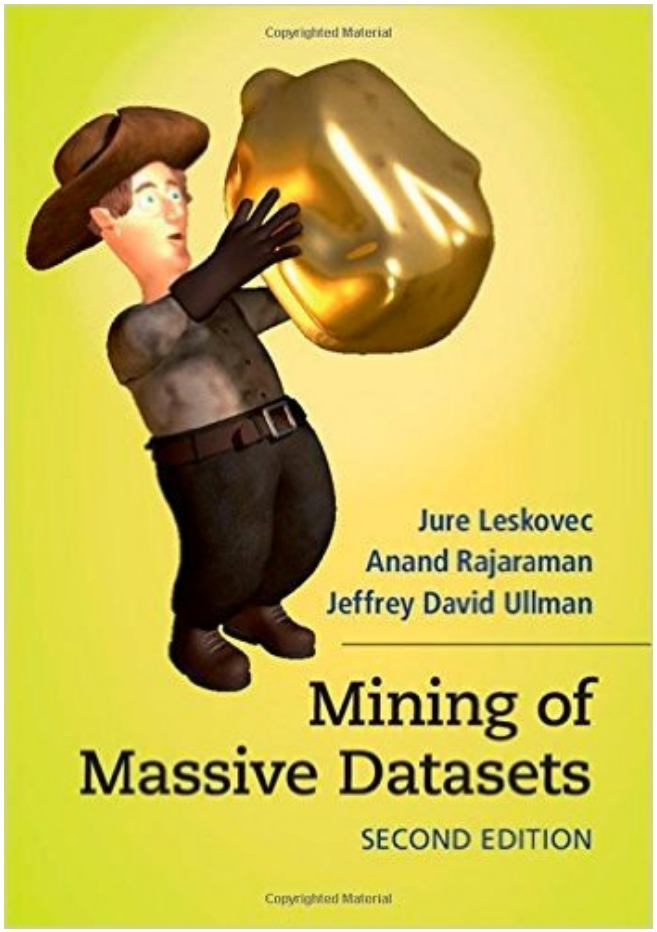
References

Parts of the course are based on the following text books

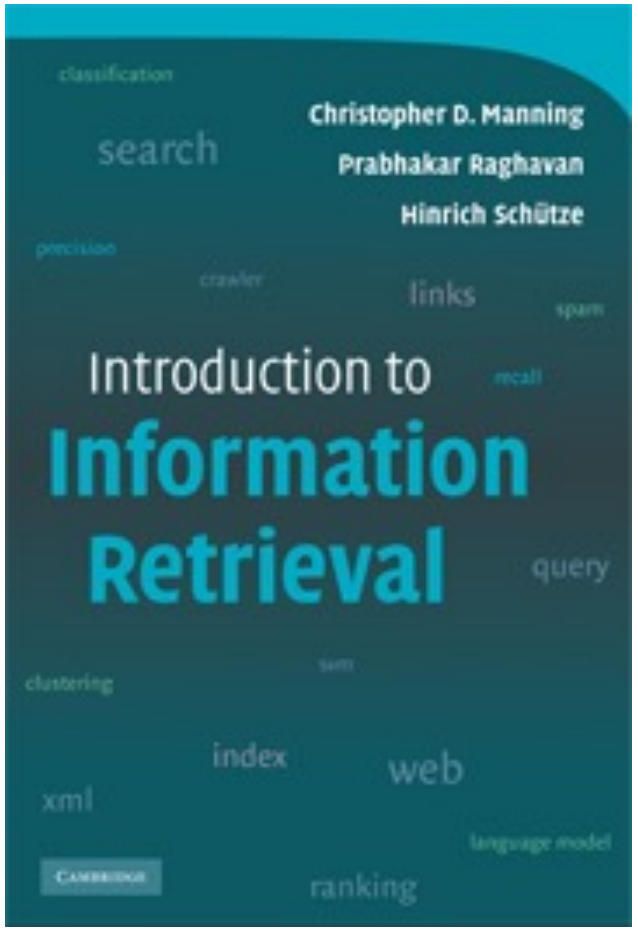
- Ricardo Baeza-Yates, Berthier Ribeiro-Neto, Modern Information Retrieval (Acm Press Series), Addison Wesley, 1999.
- Jiawei Han, Data Mining: concepts and techniques, Morgan Kaufman, 2000.
- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge 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