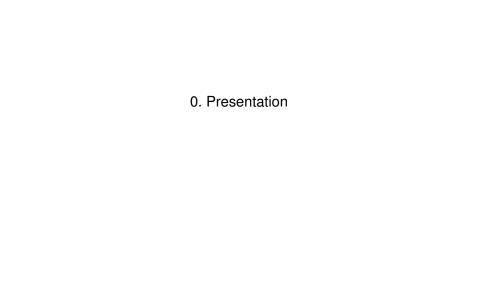
IRRS: Information Retrieval and Recommender Systems

FIB, Master in Data Science

Slides by Marta Arias, José Luis Balcázar, Ramon Ferrer-i-Cancho, Ricard Gavaldá Department of Computer Science, UPC

Fall 2024

http://www.cs.upc.edu/~ir-miri



Instructors

- ► Ramon Ferrer-i-Cancho (lectures + exercices + lab)
 - rferrericancho@cs.upc.edu
 - Omega 220, 93 413 4028

Class Logistics

Theory + problems: weekly.

- Mondays, 17–19 (theory).
- Wednesdays, 15–16 (problems).
- Problem list proposed regularly
- To be handed 1 week later

Labs / tutoring, 2 hours every 2 weeks **approx** (tentative plan).

- ▶ Wednesdays, 16–18.
- Some extra work, 2 hours on average per session Report to be handed 2 weeks later
- ► A total of 5 sessions, on dates¹: shedule in progress

¹Dates may be subjected to change; if so it will be appropriately announced in advance through Racó

Evaluation I

- Exercises: Collected in exercise sessions. 25 %
- Lab work: Weighted average of report grades. 25 %
- ► Exam (January): 30 %
- Presentation of a paper (January): 20 %

Check here date of Exam and Presentations

https://www.fib.upc.edu/en/studies/masters/master-data-science/exams.

Lab work and exercises will be scored on a 4-point scale:

- 0 not really tried
- 1 tried, but with major flaws
- 2 main ideas correct, but incomplete or some insight missing
- 3 basically ok (normally here if some silly numerical error not systematic errors).

Evaluation II

About exercise and lab assignments

- To be solved in teams of two people.
- ► The partner for exercises cannot be repeated for exercises. The partner for lab cannot be repeated for lab. The partner for exercises and the partner for lab may coincide.
- In case that the number of students available is odd, teams of three people are not allowed (thus one student will work alone).
- Post solution through the "Racó" (one team member is enough, but please state authors clearly in your delivered document)

Contents I

First half:

- Core Information Retrieval:
 - Introduction: Concept. The IR process
 - Information Retrieval Models
 - Indexing and Searching, Implementation
 - Information Retrieval Evaluation, Feedback Models
- Web Search:
 - Link analysis: Page Rank
 - Crawling the web
 - Architecture of a Web search system

Contents II

Second half:

- The "Big Data" Slogan
 - Architecture of large-scale web search systems
 - The Map-Reduce paradigm
 - Introduction to NoSQL databases
 - The Apache ecosystem for web search.
- Social Network Analysis:
 - Characterizing of real complex networks
 - Communities, influence, information diffusion
- Clustering and Locality Sensitive Hashing
- Recommender Systems

Bibliography

- R. Baeza-Yates, B. Ribeiro-Neto: Modern Information Retrieval (2nd ed.). Addison Wesley, 2010.
- ► I.H. Witten, A. Moffat, T. Bell: Managing Gigabytes. Morgan Kaufmann, 1999.
- C.D. Manning, P. Raghavan, H. Schütze: Introduction to Information Retrieval. Cambridge 2008.
- Z. Markov, D.T. Larose: Data Mining the Web. Wiley, 2007.
- Russell, Matthew, Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Site. O'Reilly, 2011
- There's a whole web out there