IR: Information Retrieval
FIB, Master in Innovation and Research in Informatics

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

Fall 2021
http://www.cs.upc.edu/~ir-miri
0. Presentation
Instructors

- Ramon Ferrer-i-Cancho (lectures + exercises + lab)
  - rferrericancho@cs.upc.edu
  - Omega S124, 93 413 4028
Class Logistics

Theory + problems: weekly.
- Wednesdays, 11–13, A6103
- Wednesdays, 13–14, A6103
- Problem list proposed regularly
- To be handed 1 week later

Labs / tutoring, 2 hours every 2 weeks approx (tentative plan).
- Mondays, 8–10, C6S306
- Some extra work, 2 hours on average per session Report to be handed 2 weeks later
- A total of 6 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, date to be announced: 30%
- Presentation of a paper (late December or early January, to be determined): 20%

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.
▶ You can pair with the same person for the whole course but then the couple for exercises and the couple for lab cannot be the same.
▶ 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

- 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