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

Slides by Marta Arias, José Balcázar, Ricard Gavaldá
Department of Computer Science, UPC

Fall 2017
http://www.cs.upc.edu/~ir-miri
0. Presentation
Instructors

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

- Marta Arias (lab sessions)
  - marias@cs.upc.edu
  - Omega 322, 93 413 7858
Class Logistics

Theory + problems: weekly.

- Tuesdays, 11–12, A6105
- Friday 9–11, A6105
- Problem list proposed weekly
- To be handed 1 week later

Labs / tutoring, 2 hours every 2 weeks **approx**.

- Tuesdays, 9–11, C6S301
- Some extra work, 2 hours in average per session Report to be handed 2 weeks later
Evaluation I

- Exercises: Collected in exercise sessions. 25% (scored as 0,1,2 or 3!)
- Lab work: Weighted average of report grades. 25%
- Exam, January, date to be announced: 30%
- Presentation of a paper, January: 20%
Evaluation II

- About exercise and lab assignments:
- To be solved in teams of two people
- You should team with at least 3 different people during the course, i.e., rotate partners as much as you can
- We can make reasoned exceptions, but tell us in advance
- Both members of a team must post their (identical) solutions at Racó
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
Second half:

- Social Network Analysis:
  - Characterizing of real complex networks
  - Communities, influence, information diffusion

- Clustering and Locality Sensitive Hashing

- 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.

- Recommender Systems
References

- 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