Web Search and Data Mining

Computer Science MSc Course

João Magalhães

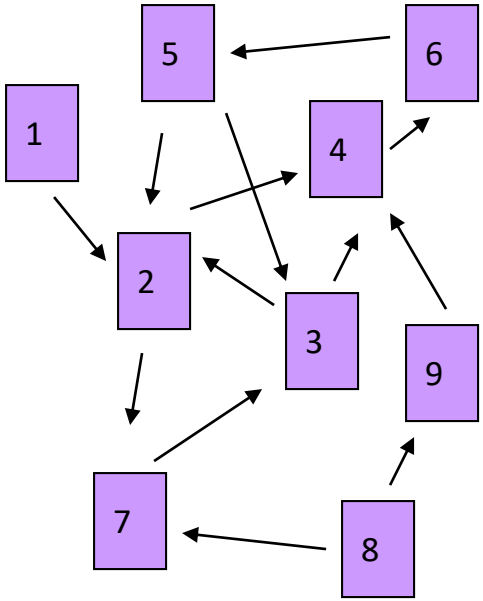
How to search Web information?

- Textual and visual data can communicate a wide variety of information that are critical for several decision processes.
- Temporal and spatial structure adds organization and usability to information.
- Non-structured data (language and vision) puts a heavy complexity burden on standard data structures.

Course plan

Web Mining and Search				
Week	#	Lecture	In-class labs	
05/mar/20	1	Introduction	1	Lab setup
12/mar/20	2	Web data representation	2	Data representations
19/mar/20	3	Web-graph analysis	3	Project
26/mar/20	4	Recommendation algorithms	4	Project
02/abr/20	5	Learning vision data representations	5	Project checkpoint
09/abr/20	6	Natural language representations	6	Project
16/abr/20	7	Case studies ECIR	7	Case studies ECIR
23/abr/20	8	Named Entities and Knowledge Graph	8	Project
30/abr/20	9	Multimodal representations	9	Project
07/mai/20	10	Locality sensitive hashing	10	Project checkpoint
14/mai/20	11	Visual Question Answering	11	Project
21/mai/20	12	Paper summary	12	Project
28/mai/20	13	Paper summary	13	Project
04/jun/20	14	Revisions	14	Project
11/jun/20		Test		Project submission

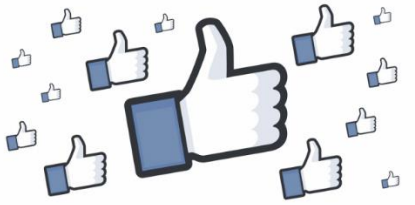
Web data



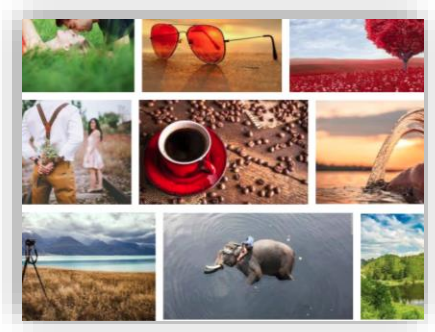
Links

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim…

Text

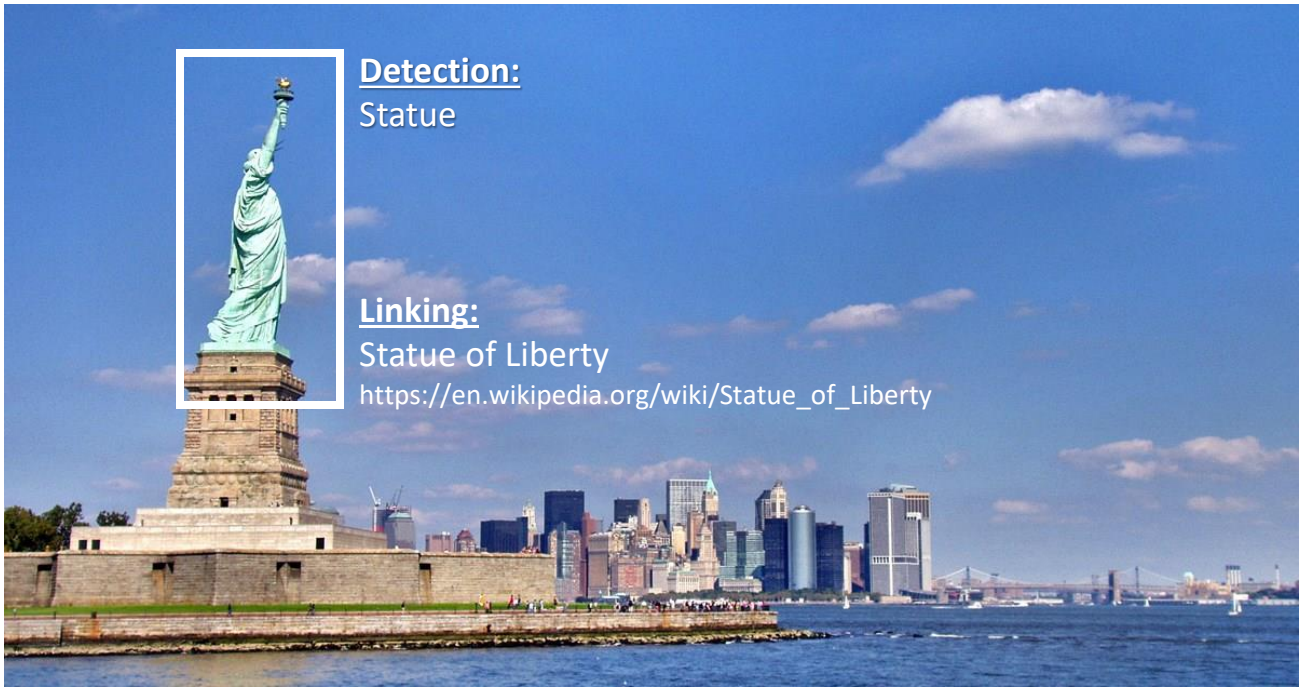


Preferences



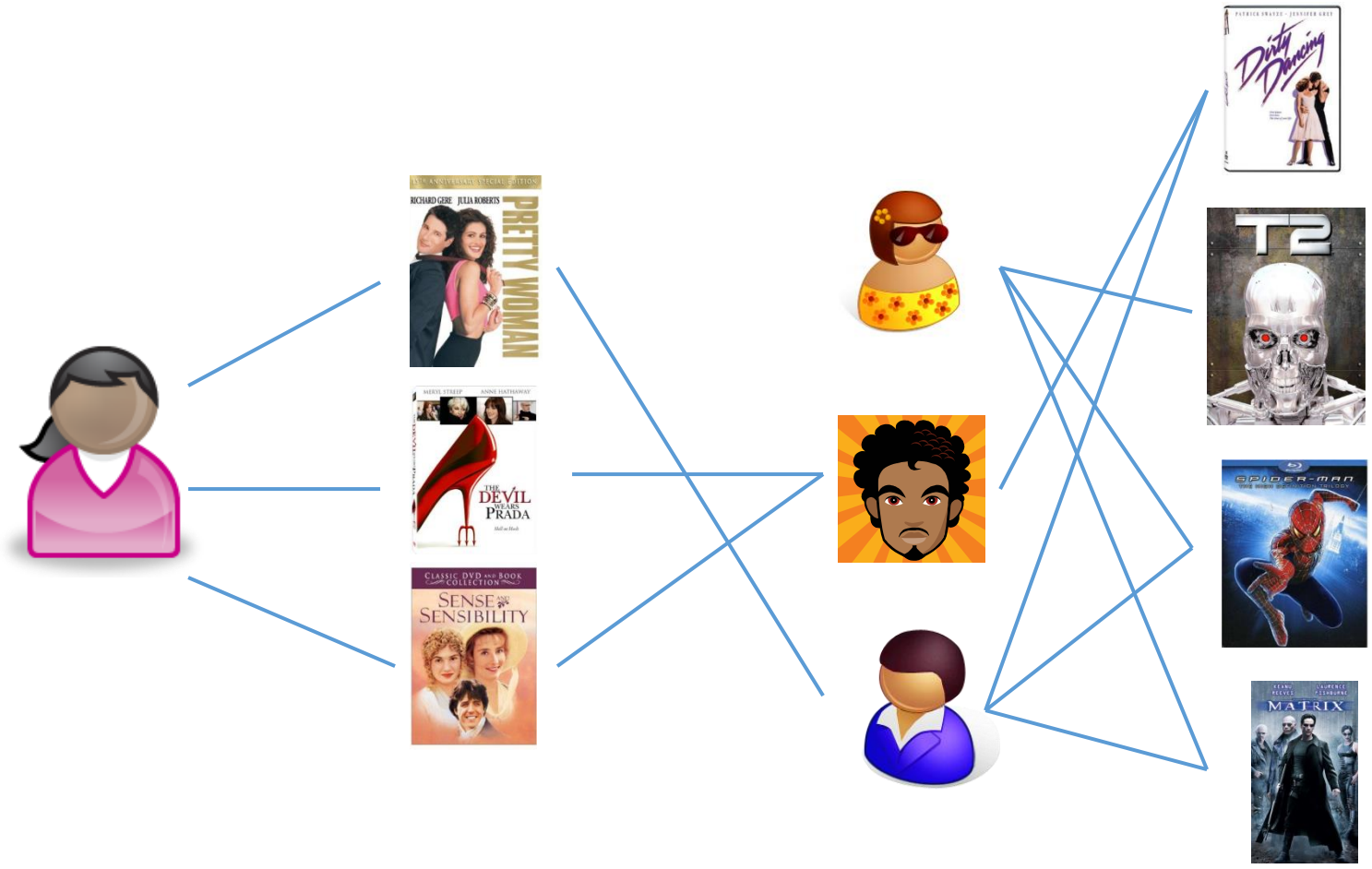
Images/videos

Classification, detection, linking

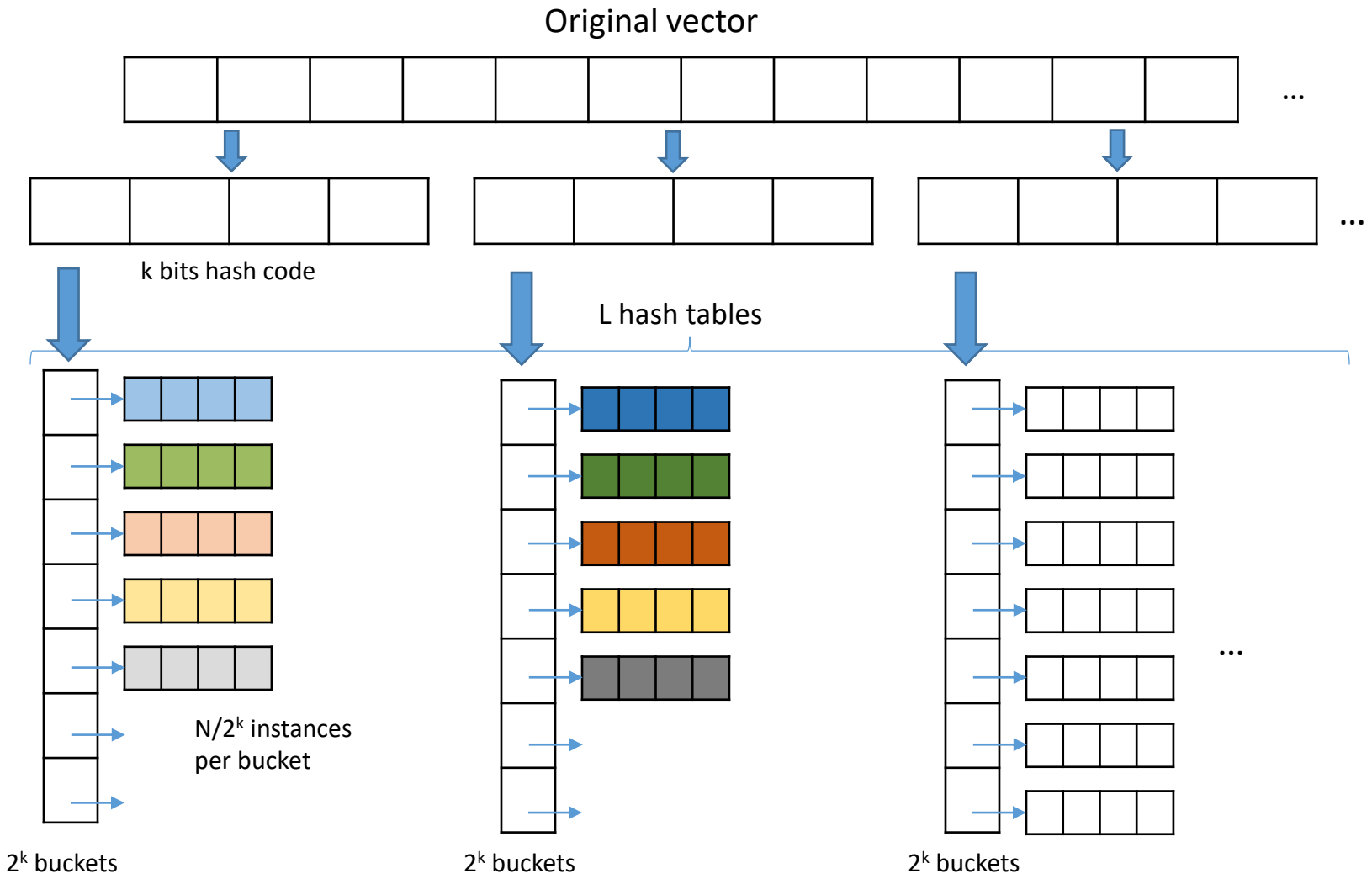


Classification:
Sea side
Statue
City
Sky

Collaborative filtering



Large-scale search data structures



Funny reflex...

Share This

ADD NOTE SEND TO GROUP ADD TO SET BLOG THIS ALL SIZES ORDER PRINTS ROTATE EDIT PHOTO DELETE

Uploaded on August 3, 2006
by jmc_mag

jmc_mag's photostream

Best ones (Set)

You are at the first photo.

11 items

browse

Tags

- Arizona
- Reflection
- Desert

Add a tag

Comments

alexei_322 pro says:

thumbs up! great shot
Posted 31 months ago. (permalink | delete)

Add your comment

Additional Information

All rights reserved (edit)
Anyone can see this photo (edit)

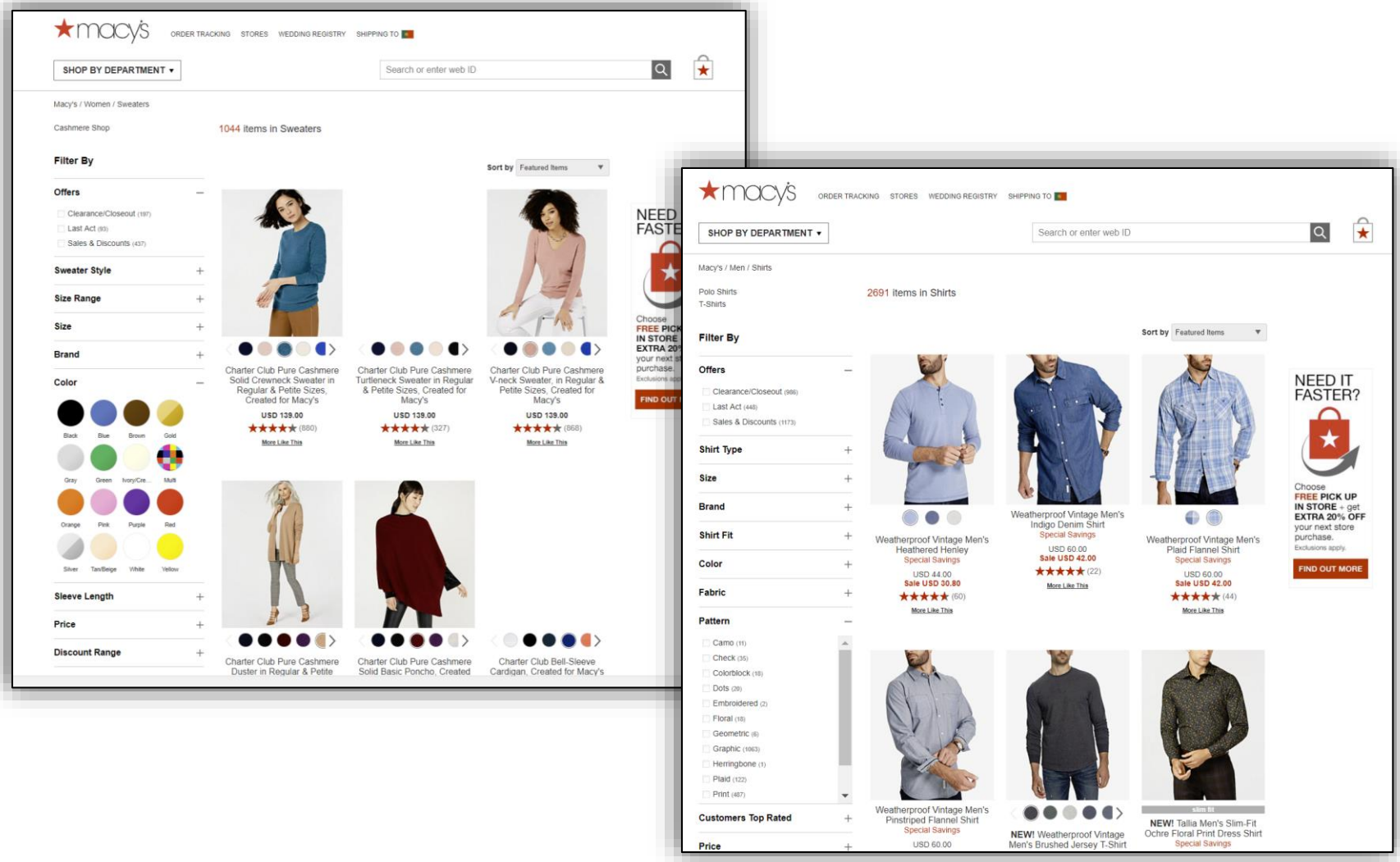
Add to your map
Taken with a Canon EOS Digital Rebel XT.
More properties
Taken on July 15, 2006 (edit)

Photo stats
Viewed 26 times (Not including you)

Edit title, description, and tags

Web data based search

Online shopping



Medical domain

Course grading

- 40% theoretical part (1 test or 1 exam)
- 50% for a 3 parts project
  - Submission 1
  - Submission 2
  - 100% Final submission
- 10% case study presentation
- Groups of 3 students, maximum 8 groups
- Additional rules:
  - Minimum grade on the labs or theory: 9
  - You may use one sided A4 sheet handwritten by you with your notes
    - It must be handed at the end of the test.

Project grading

- Scoring:
  - Implement. correctness 30%
  - Results analysis 30%
  - Critical discussion 40%
- Report:
  - Maximum of 8 pages.
  - No cover page.
  - Must include graphs, tables, etc.

- Report organization:
  - Introduction
  - Algorithms
  - Implementation
  - Evaluation
    - Dataset description
    - Baselines
    - Results analysis
  - Critical discussion
  - References

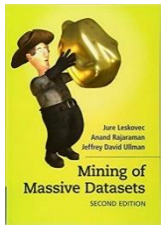
Case studies

1. Crowd Knowledge Enhanced **Multimodal Conversational** Assistant in Travel Domain
2. TweetFit: Fusing Multiple Social Media and Sensor Data for **Wellness Profile Learning**
3. Knowledge-aware **Multimodal Dialogue** Systems
4. Multi-modal Knowledge-aware Hierarchical Attention Network for Explainable **Medical Question Answering**
5. DeFacto - Temporal and Multilingual Deep **Fact Validation**
6. Modeling **Temporal Evidence** from External Web Collections
7. Ranking **News-Quality** Multimedia
8. **Crowdsourcing** facial expressions for affective-interaction

References

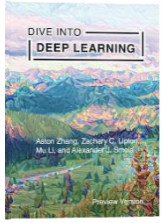
- Slides and articles provided during classes.

- Books:



Jure Leskovec, Anand Rajaraman, Jeff Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2011.

http://www.mmds.org/



Aston Zhang, Zachary Lipton, Mu Li, and Alex Smola, “Dive into Deep Learning”

http://d2l.ai/