Web Search

MSc level Course

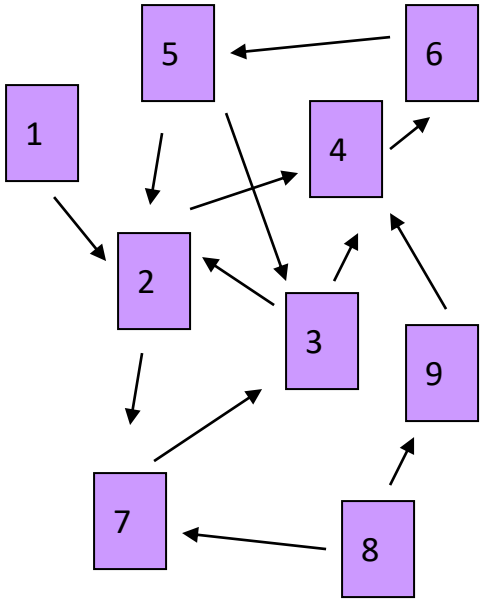
João Magalhães

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.

Web data based search

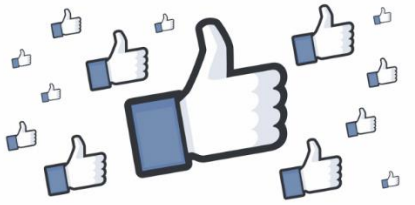
Web graphs



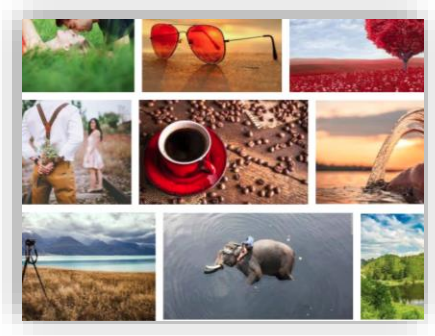
Links

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

Text

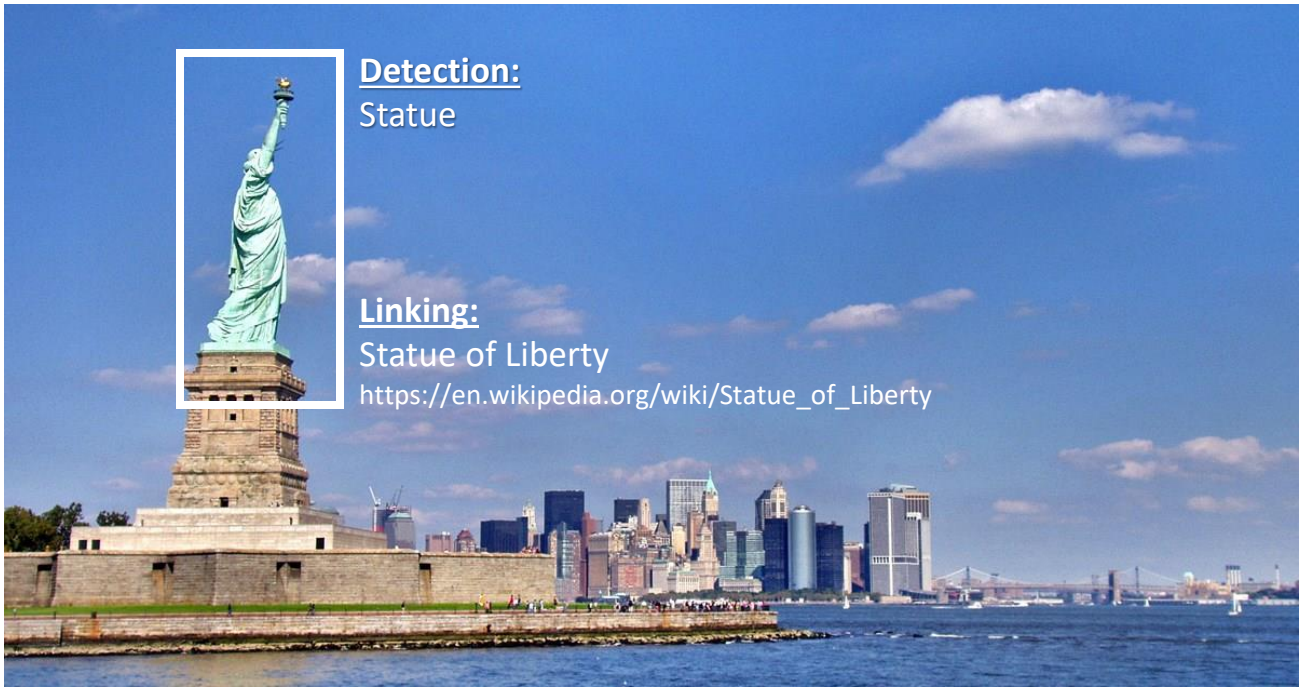


Preferences



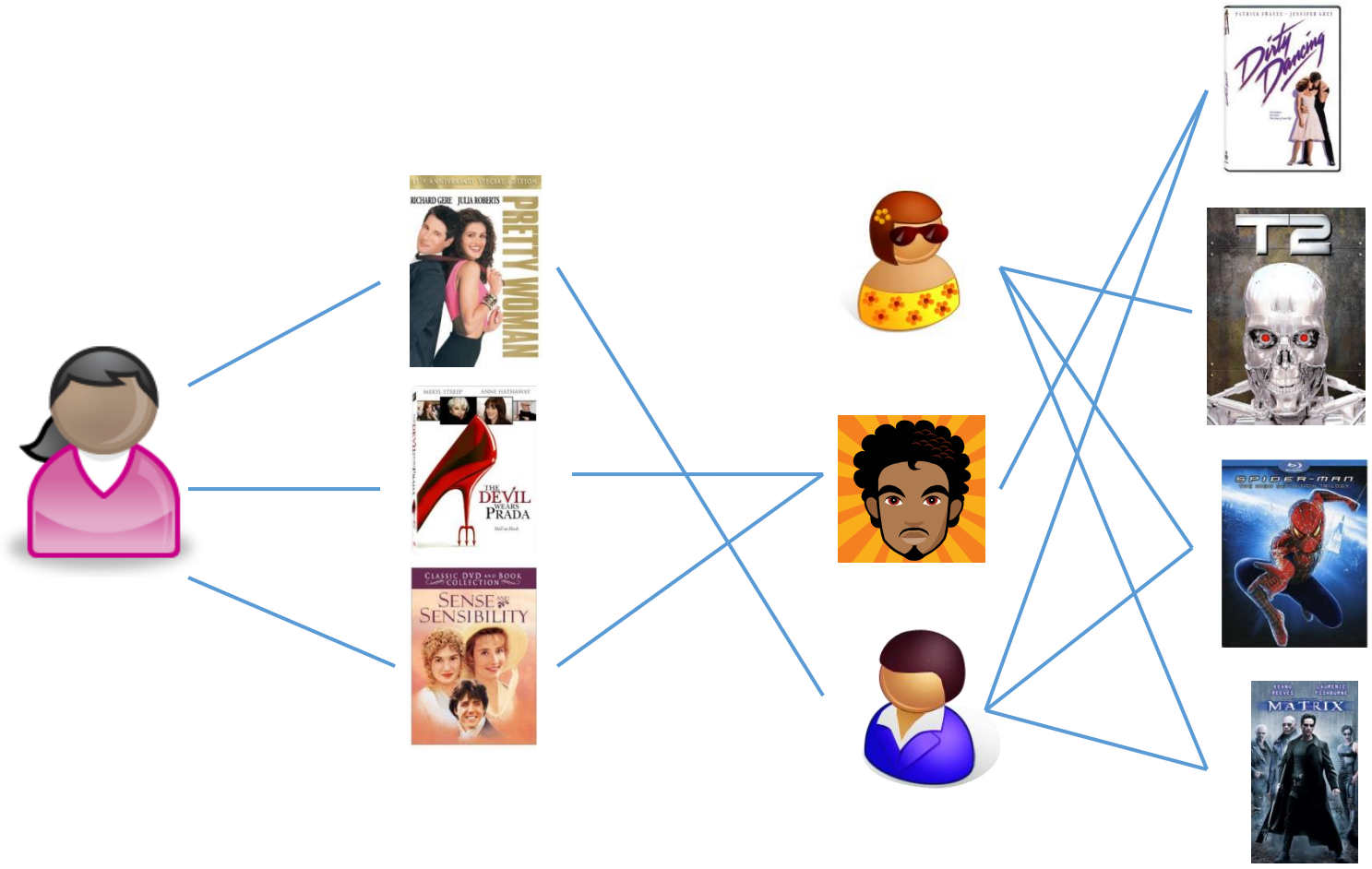
Images/videos

Classification, detection, linking



Classification:
Sea side
Statue
City
Sky

Collaborative filtering



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

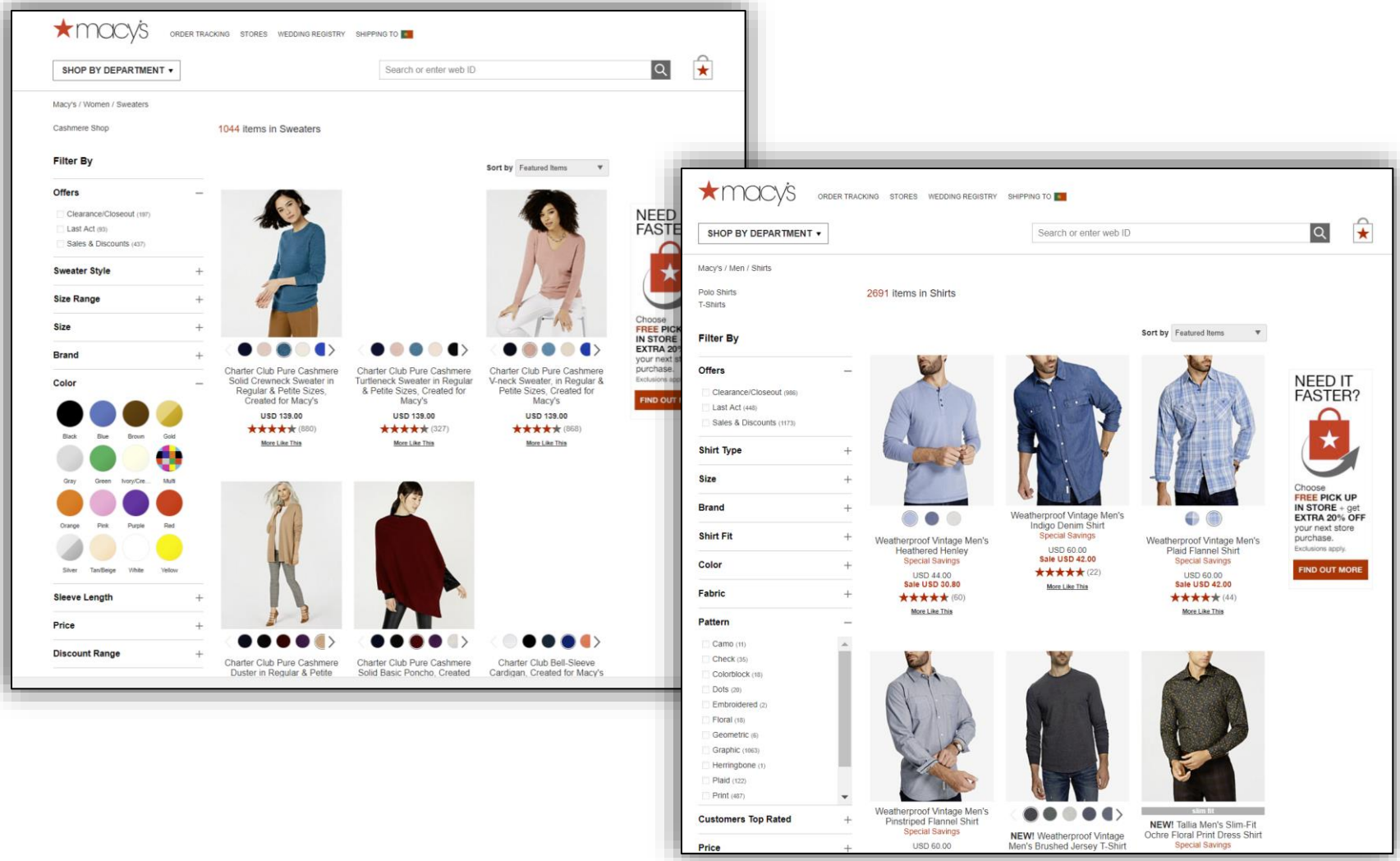
All rights reserved (edit)
Anyone can see this photo (edit)

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

Online shopping



Case-based search

Course plan

- Online graph analysis: social-networks, Web graphs, etc.
- Recommender systems: movies, books, friends, etc.
- Information tagging and annotation
- Image and video tasks: automatic captioning and visual Question-Answering
- Document summarization and answer generation

Course plan (in detail)

Web Data Mining and Search		
Week	**#**	**Lecture**
11/mar/21	1	Introduction
18/mar/21	2	Web data driven tasks
25/mar/21	3	Recommendation
01/abr/21		**Easter**
08/abr/21	4	Web Graphs
15/abr/21	5	Web document categorization
22/abr/21	6	Visual data search
29/abr/21	7	Word and Entity embeddings
06/mai/21	8	Modeling sequence data
13/mai/21	9	Transformer
20/mai/21	10	Mixed vision-and-language models
27/mai/21	11	Document summarization
08/jun/21		**Invited lecture: Data Science for Social Good**
17/jun/21		**Project presentation / Test**

Course grading

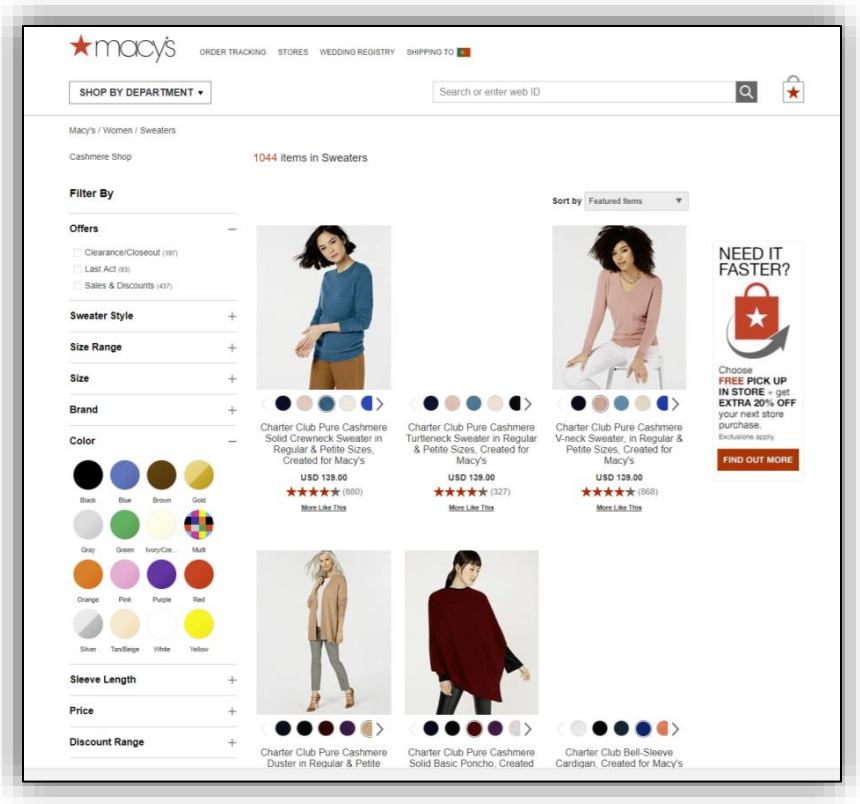
- 40% test / exam
- 50% project
- 10% project presentation
- Groups of 3 students
- Additional rules:
  - Minimum grade on the labs or theory: 9
  - You may use one sided A4 sheet handwritten by you with your notes
    - It must be handed at the end of the test.

PW is a project-based learning course!!!

- Fashion product search
- Recommendation
- Web images captioning
- Visual QA
- Timeline summarization
- Open-domain search assistant

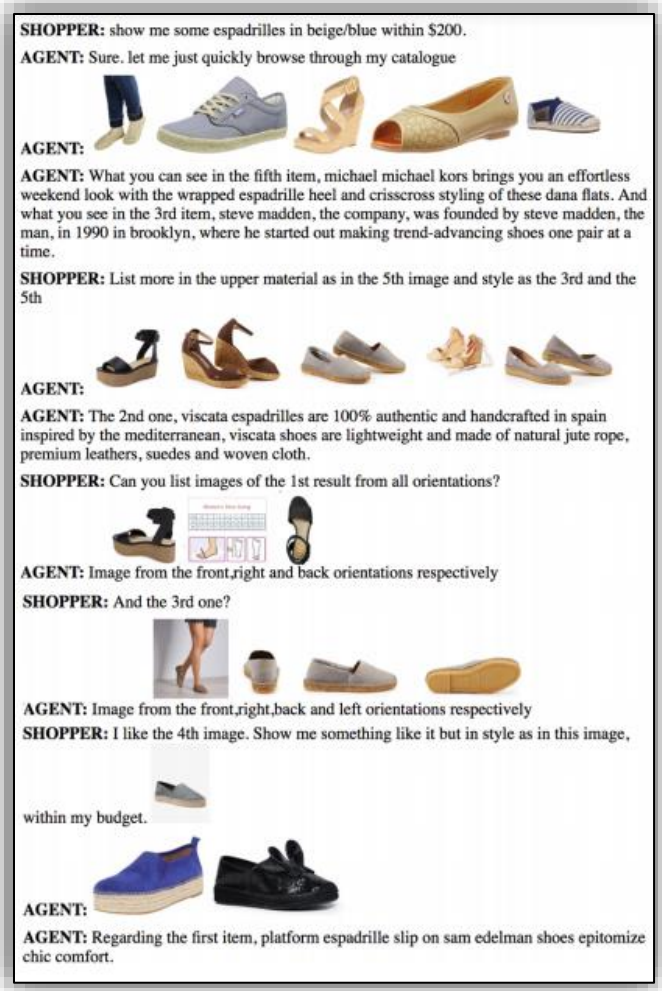
Fashion product search

- Fashion dataset
- ElasticSearch
- Visual Embeddings
- Faceted search and browsing



Recommendation

- Click-through dataset
- Model past likes/don’t likes
- Predict product to be purchased



Web images captioning

- MSCOCO dataset
- PyTorch to generate captions
- External knowledge is optional

A young boy is playing basketball. 	Two dogs play in the grass. 	A dog swims in the water. 	A little girl in a pink shirt is swinging. 
A group of people walking down a street. 	A group of women dressed in formal attire. 	Two children play in the water. 	A dog jumps over a hurdle. 

Visual QA

- MS COCO dataset
- PyTorch to generate tags
- External knowledge is optional
- Classification type of answers

Attributes:
umbrella
beach
sunny
day
people
sand
laying
blue
green
mountain

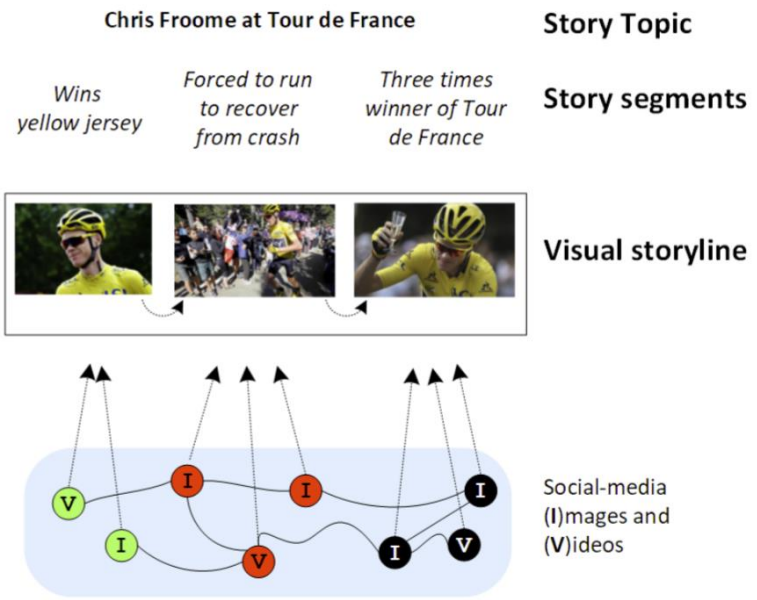
Internal Textual Representation:
A group of people enjoying a sunny day at the beach with umbrellas in the sand.

External Knowledge:
An umbrella is a canopy designed to protect against rain or sunlight. Larger umbrellas are often used as points of shade on a sunny beach. A beach is a landform along the coast of an ocean. It usually consists of loose particles, such as sand....

Question Answering:
Q: Why do they have umbrellas? **A :** Shade.

Visual Stories - Summarization

- Create visual summaries of events
- Dataset will be provided
- Summarization of timelines



https://github.com/davidfsemedo/TRECVID-VisualStories

Open-domain search assistant

- Wikipedia powered chatbot
- Open-ended conversations
- One single answer instead of document list

https://knowledge-answer-generation.herokuapp.com/#
https://www.arxiv-vanity.com/papers/2101.08197/

Project grading

- Report:
  - MAXIMUM 8 PAGES!!!
  - No cover page.
  - Must include graphs, tables, etc.
- Scoring:
  - Algorithms imp./int. 30%
  - Critical discussion 40%
  - Results analysis 30%

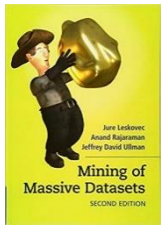
    OR
  - UI implementation 30%

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

References

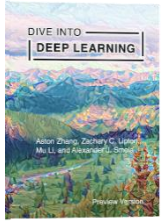
- Slides and articles provided during classes.

- 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/