

Topics in Multimedia Research

Dr. Michael S. Lew

Director of the Media Research Group
co-Head of the Imagery and Media Research Cluster
Leiden University
E-mail: mlew@liacs.nl

Multimedia Research Community

- *My group publishes regularly in*
 - *Major Scientific Conferences*
 - *ACM International Conference on Multimedia*
 - *ACM SIGGRAPH - computer graphics*
 - *ACM International Conference on Multimedia Retrieval*

LIACS Media Research Group

We are interested in the frontiers of multimedia research, technology which will often be commonplace 2-5 years in the future.

Faculty

(about 10 fulltime members in total)

- *Dr. Erwin M. Bakker, erwin@liacs.nl*

- <http://www.liacs.nl/~erwin/>
- Audio/Speech search and modeling
- BioInformatics



- *Dr. Michael Lew, mlew@liacs.nl*

- <http://www.liacs.nl/~mlew/>
- Machine learning
- Graphics & Multimedia Information Retrieval



Some Project Topics

- *Topics*
 - *Social Computing and Recommendation Systems*
 - *Social Media Analysis*
 - *Wikipedia Analysis*
 - *High Performance Signatures*
 - *Searching large image/video collections*

Social Computing & Recommendation Engines

- Recommendation engines are at the intersection of machine learning and societal usage.
- Just about every Internet shop wants to recommend new items to buy (Bol, Amazon)
- or People to connect to
- or Groups to join
- or What movie should you see next?

Social Computing & Recommendation Engines

- List of m users and a list of n Items .
- Each user has a list of items he/she expressed their opinion about (can be a null set).
- Explicit opinion - a rating score (numerical scale).

Collaborative Filtering (CF)

- The task of a CF algorithm is to find item likeliness of two forms :

Prediction – a numerical value, expressing the predicted likeliness of an item the user hasn't expressed his/her opinion about.

Recommendation – a list of N items the active user will like the most (Top-N recommendations).

Collab Filter - No Problems

	Star Wars	Sense and Sensibility	Inception	Pride and Prejudice	The Empire Strikes Back
Alex	9	6	9	3	10
Jamie	7	8	7	10	7
Terry	9	5	9	?	?

Problem: User Conflict

What do you do when user ratings conflict?

	Star Wars	Sense and Sensibility	Inception	Pride and Prejudice	The Empire Strikes Back
Alex	10	5	9	3	10
Jamie	9	5	8	10	7
Terry	9	5	9	?	?

Project - Video Recommendation

MovieLens Project has complete ratings databases of varying size: 100K, 1M, etc.

<http://www.grouplens.org/node/73>

Plan:

- (1) Download a MovieLens DB
- (2) Write a simple prediction algorithm
- (3) Design and write a more sophisticated algorithm
 - indirect links
 - other machine learning algorithms
 - etc.

Reading

A large-scale evaluation and analysis of personalized search strategies

Zhicheng Dou, Ruihua Song, Ji-Rong Wen

In Proceedings WWW '07 Proceedings of the 16th International Conference on World Wide Web

Evaluating collaborative filtering recommender systems

Jonathan L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl

ACM Transactions on Information Systems (TOIS), Volume 22 Issue 1, January 2004

Social Sentiment Analysis

- For a social post or a new product, there are often reviews and comments.
- How can one automatically summarize the positive/negative view toward the post or item?
- Starting point: analyze occurrence of certain categories of words which have positive or negative meanings.

Social Trend Analysis

- What is the next big event? Ebola, Climate Change, Separation from the EU?
- How can one measure trends on the Internet?
- Starting point: download WWW news information and model by clustering. Design new measures for automatically analyzing trends.

Wikipedia Analysis

- *Everyone uses Wikipedia*
 - What is the level of spam in Wikipedia?
 - How accurate is Wikipedia?

Wikipedia Analysis

- *Wikipedia Search Engine is Poor*

- Design a new search engine for Wikipedia

- How can one measure how good/complete a Wikipedia page is?

High Performance Search

- In distributed web downloading (i.e. Google), you do not want the overhead of cross checking signatures (*cryptographic hash function*) between download nodes.
 - webpage -> signature (hopefully unique)
 - minimize number of bits in signature
- Which signatures algorithm (i.e. MD5, CRC) performs the best for web downloads at a particular number of bits or optimize aspects relevant to a particular problem?
- How do they perform against a modern pseudo-random number generator?
 - uniqueness (or collisions) vs. cpu generation time

Image Search

- Research Problem & Questions:
 - Most image search are 2D:

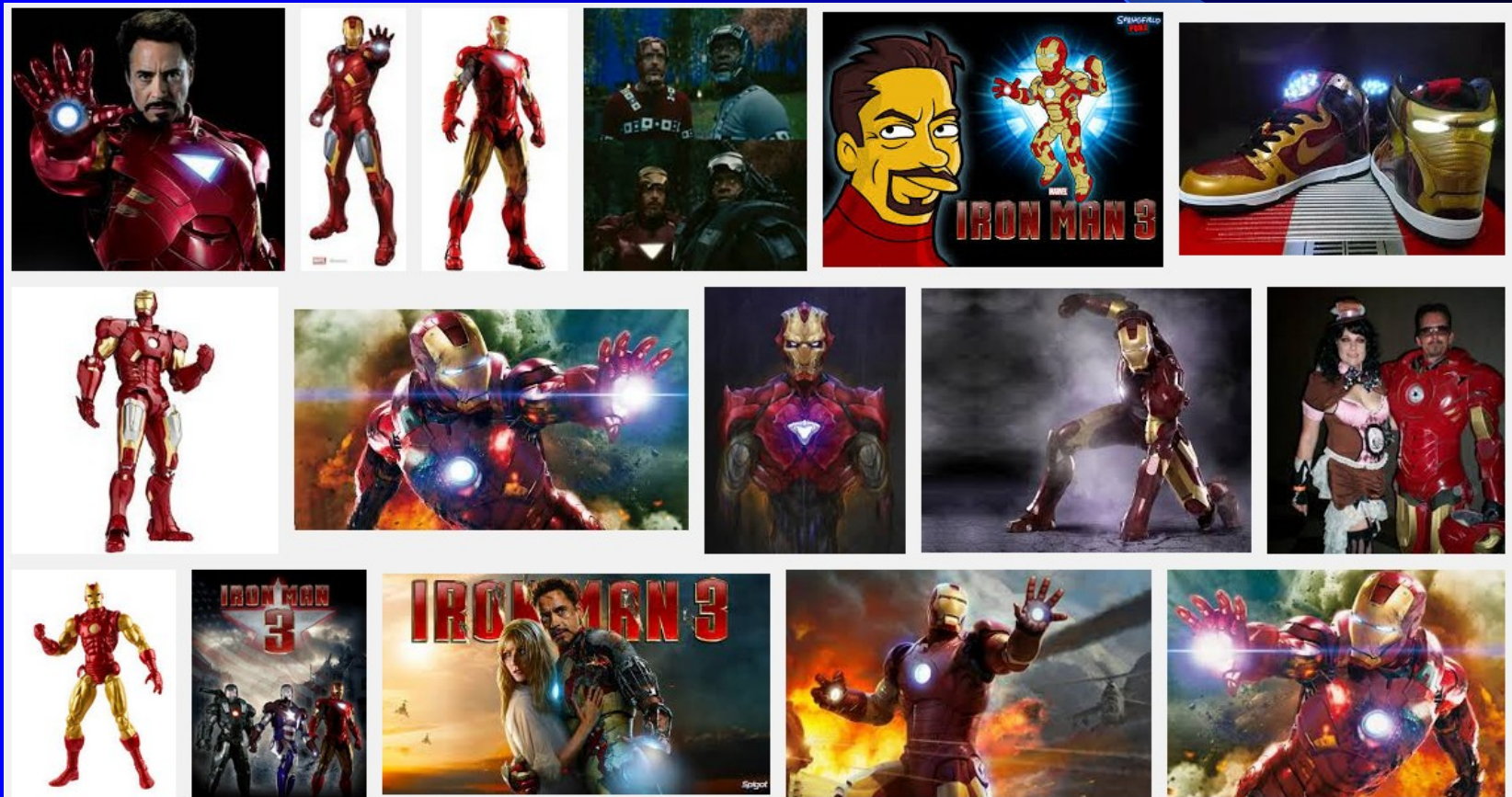
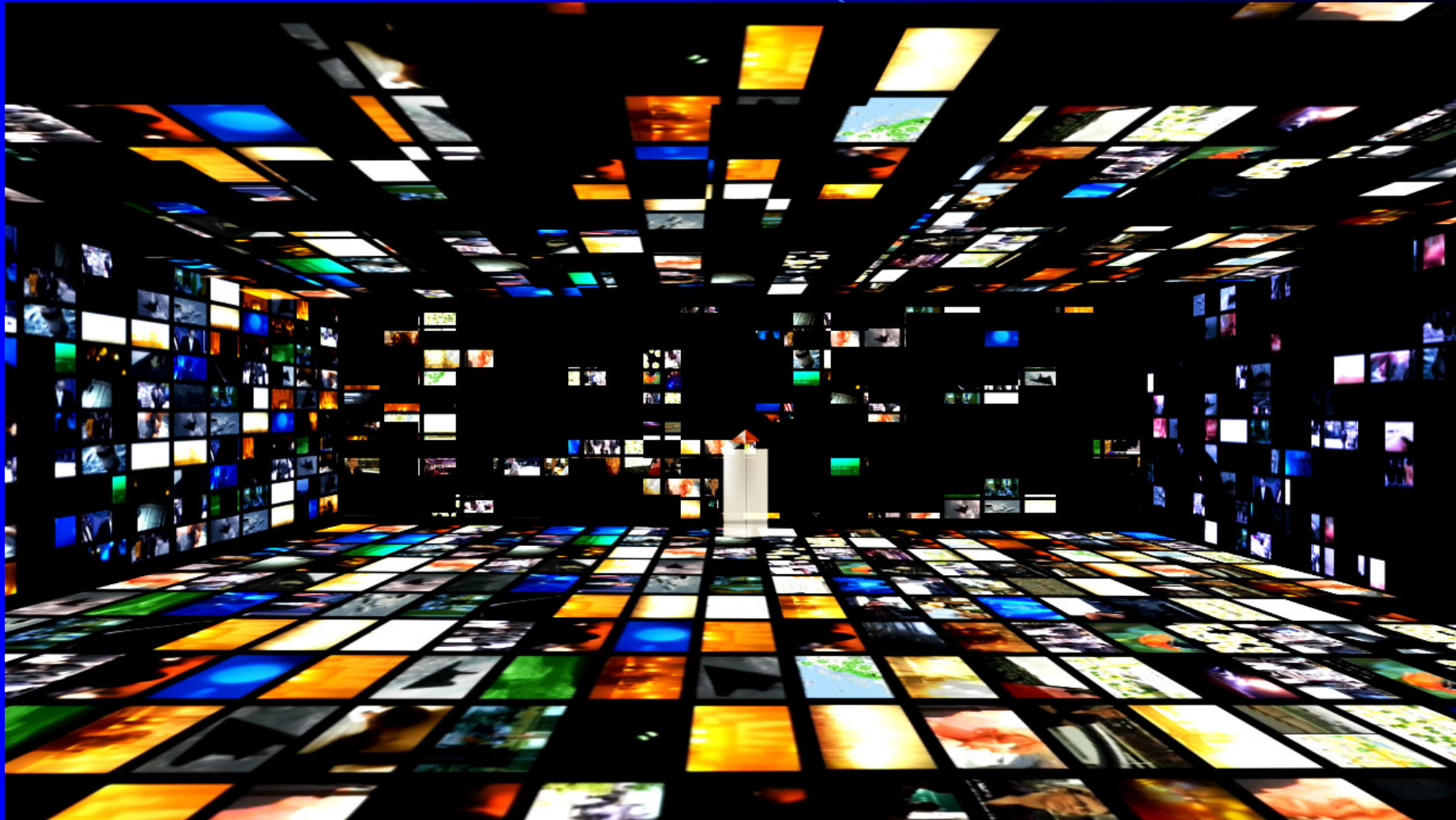


Image Search

- Research Problem & Questions:
 - **Starting Point:** Use Google Image Search to find the initial results
 - Download and re-rank the results based on
 - image similarity
 - user preferences
 - prior feedback from user

Image Search in 3D

What about displaying the results? What would be an effective interface?



Concluding Remarks

The project descriptions are only *starting points*.

The research question/goal and approaches will evolve with discussion.

If these subjects are interesting to you, please email me for a chat about what you would like to do.