

From **Technology-Inspired** towards **Utility-Centered** **Multimedia Information Retrieval**

Alan Hanjalic

Multimedia Signal Processing Group
Delft University of Technology
The Netherlands

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

MIR: MM content access in a broad scope



Google soccer

Opgeleverd: 106.000.000 resultaten (0,03 seconden)

Alles
 Abbeelden
 Video's
 Meer

Elk formaat
 Groot
 Gemiddeld
 Pictogram
 Groter dan...
 Exact...

Elk type
 Foto
 Clipart
 Lijsttekening
 Foto
 Clipart

Elke kleur
 Full colour
 Zwart-wit

Free Soccer Clipart Images, 757 x 593 - 56k - png clearfb.com

Fail Indoor Soccer: Intarnatural 600 x 600 - 110k - png pacificu.edu Soortgelijke afbeeldingen vinden

YouTube playing with fire

Related searches: rolling stones set this s

N-Dutz ft. Mr Hudson - Playing With Fire (Official Video with ...
 The latest new single from N-Dutz is Playing With Fire and features the brilliant Mr Hudson. Playing With Fire is taken from the new album from it ...
 by SteveMTW · 9 months ago · 12,894,375 views

EXCLUSIVE! Paula Seling & Ovi - Playing With Fire
 Powered by www.eurovision.tv: Watch the NTV video clip Paula Seling & Ovi will represent Romania with the song "Playing With Fire" at the 2010 ...
 by eurovision · 4 months ago · 1,586,847 views

Let's Play Pokemon Fire Red Episode 67: Operation Average Weedy
 One critical and Weedy's out.
 by StarHPC · 1 day ago · 4,392 views

Let's Play Pokemon Fire Red Episode 64: Hypno...toad
 That battle really couldn't have gone any better.
 by StarHPC · 2 days ago · 1,708 views

Romania - "Playing With Fire" - Eurovision Song Contest 2010 -
 More about the programme: www.ebc.eu.int Paula Seling & Ovi will be representing Romania in Eurovision 2010 with the song "Playing with Fire ...
 by BBC · 4 months ago · 1,895,766 views

OnMouseOver:
 Extra information about the event

Pictogrammic visualization of the event (goal)

Yellow card, foul, Makelele

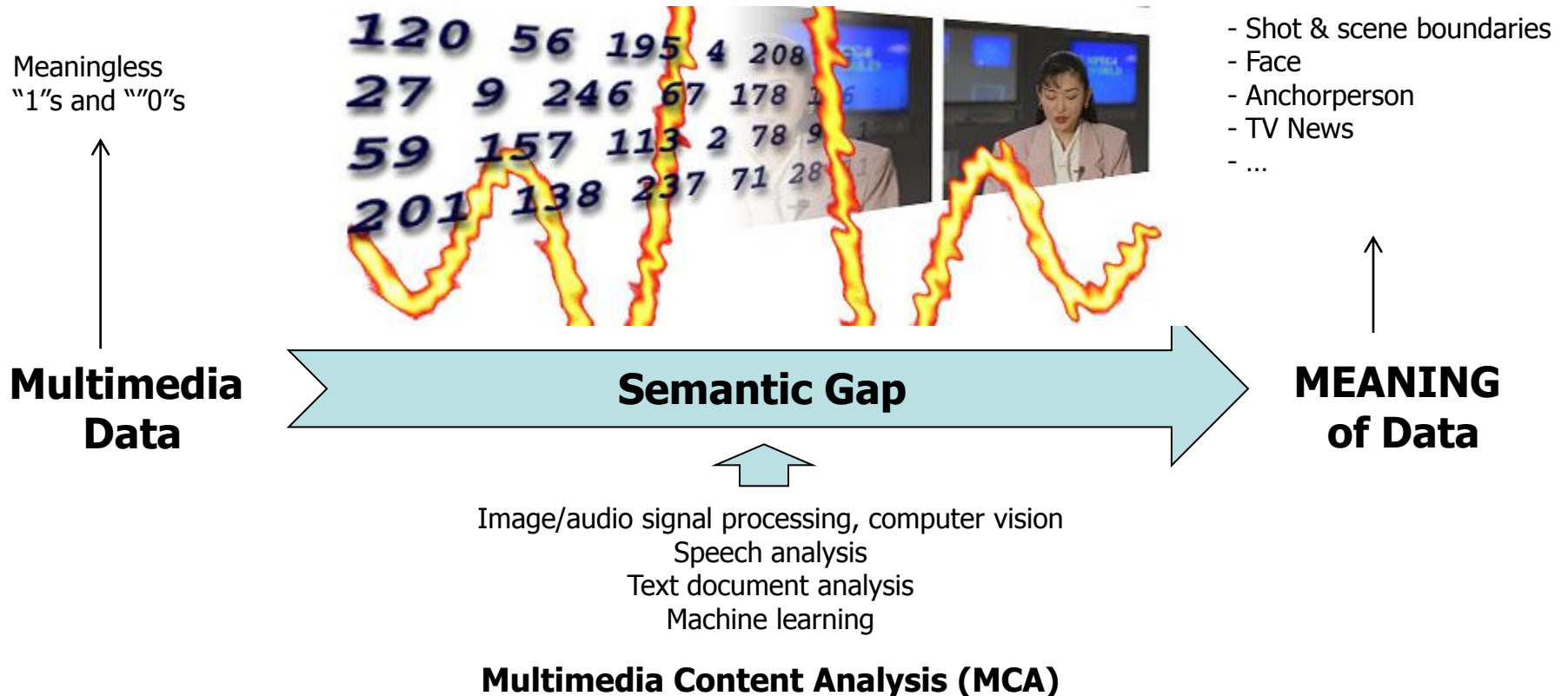
Currently playing:
 Fourth abstract segment (white dot)
 Pictogram indicates yellow card event

09:11 / 105:15

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Where it all started ...

Extracting meaning from data

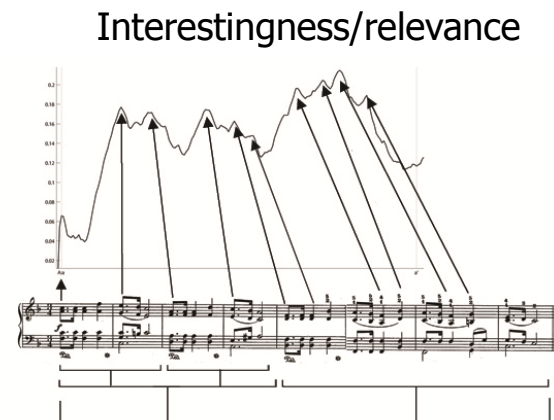
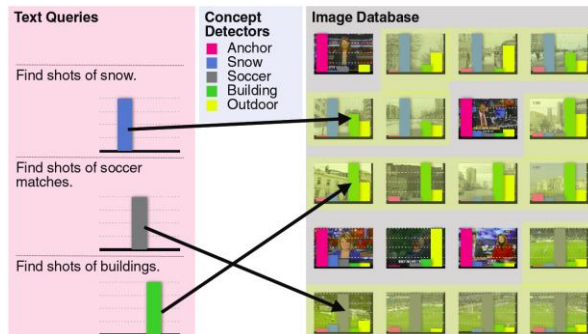
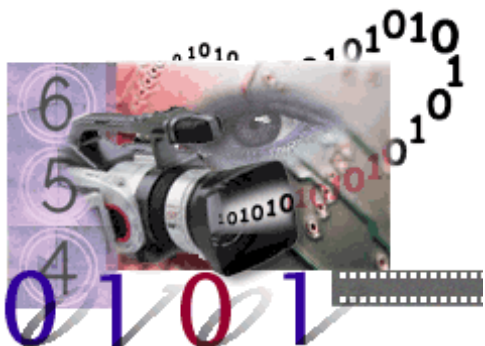


International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

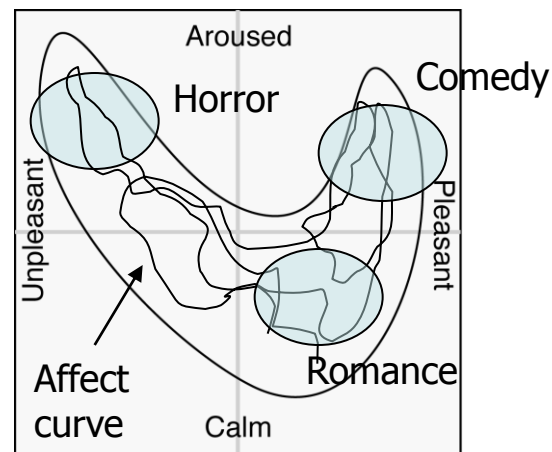
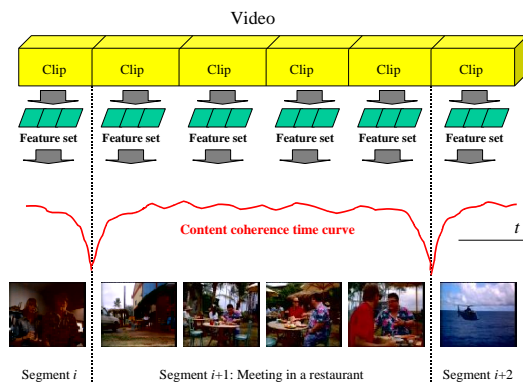
MIR around mid 2000s

Amazing growth of MCA

NIST TRECVID



Video Olympics



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

What also happened mid 2000s

Discovering the CONTEXT

- Basic idea
 - Relying on the information derived from the (social, web, use) context to generate annotations for multimedia items



From Context to Content: Leveraging Context to Infer Media Metadata

Marc Davis¹, Simon King¹, Nathan Good¹, and Risto Sarvas²

¹University of California at Berkeley
School of Information Management and Systems
314 South Hall, Berkeley, CA, USA 94720-4600
+1 510 643-2253
<http://garage.sims.berkeley.edu/>
(marc, simonpk, ngood)@sims.berkeley.edu

²Helsinki Institute for Information Technology (HIT)
P.O. Box 9800
02015 HUT, Finland
+358 9 694 9708
<http://www.hit.fi/risto.sarvas/>
risto.sarvas@hit.fi

ABSTRACT

The recent popularity of mobile camera phones allows for new opportunities to gather important metadata at the point of capture. This paper describes a method for generating metadata for photos using spatial, temporal, and social context. We describe a system we implemented for inferring location information for pictures taken with camera phones and its performance evaluation. We propose that leveraging contextual metadata at the point of capture can address the problems of the semantic and sensory gaps. In particular, combining and sharing spatial, temporal, and social contextual metadata from a given user and across users allows us to make inferences about media content.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]: Human Factors; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.5.1 [Information Interfaces and Presentation (e.g., HCI)]: Multimedia Information Systems; H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces; H.5.3 [Information Interfaces and Presentation (e.g., HCI)]: Group and Organization Interfaces; I.4m [Image Processing and Computer Vision]: Miscellaneous.

General Terms

Algorithms, Design, Human Factors.

Keywords

Mobile Camera Phones, Contextual Metadata, Content-Based Image Retrieval, Context-to-Content Inference, Wireless Multimedia Applications, Location-Based Services

1. INTRODUCTION

Multimedia researchers have been trying to solve the problems of content-based image retrieval and media asset management for well over the past decade [1, 14]. It is time to acknowledge that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'04, October 10–16, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-000-0/0004...\$5.00.

this research has not delivered on its promises. Neither fully automatic signal-based analysis nor manual annotation of media content has provided a workable solution to content-based multimedia access. A new direction and new solutions are needed. In order for media to be as accessible as text, descriptions of its content and structure (i.e., metadata) must be created that are computationally and humanly usable. Unfortunately, the low-level features that current algorithms can extract are not sufficient to meet the needs of how humans want to search for and use media content. This “semantic gap” is endemic to current multimedia information systems [3].

Most prior research in multimedia content analysis has suffered from assumptions that have impeded progress in the field and in making systems that can address users’ needs. These assumptions are: 1) media capture and media analysis are disconnected in time and space such that media must be analyzed long after they have been captured (and therefore effectively removed from their context of creation and the users who created them); 2) contextual metadata about the capture and use of media are not available to media analysis and hence all analysis of media content must be focused on the media signal; and 3) multimedia content analysis must be fully automatic and avoid user involvement (and therefore miss out on the possibility of “human-in-the-loop” approaches to algorithm design). These assumptions can be overcome by shifting the paradigm of media processing from an a-contextual, fully automatic model to a model that leverages the spatio-temporal-social context of media creation and use as well as interaction among devices and people (especially by taking advantage of new programmable networked mobile media capture devices). By making this shift in method and focus that takes advantage of new technology platforms for media creation, we have the promise of solving long-standing challenges in multimedia systems.

The devices and usage contexts of media capture are undergoing rapid transformation from the traditional camera-to-desktop-to-network pipeline to an integrated mobile media experience. We now see a new class of networked media capture devices (typified by camera phones) that combine: media capture (images, video, audio); programmable processing using standard operating systems, programming languages, and APIs; wireless networking; rich user interaction modalities; time, location, and user metadata; and personal information management functions. The confluence of this functionality at the point of media capture means that we have a unique opportunity to attempt to solve once intractable

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Games with the purpose

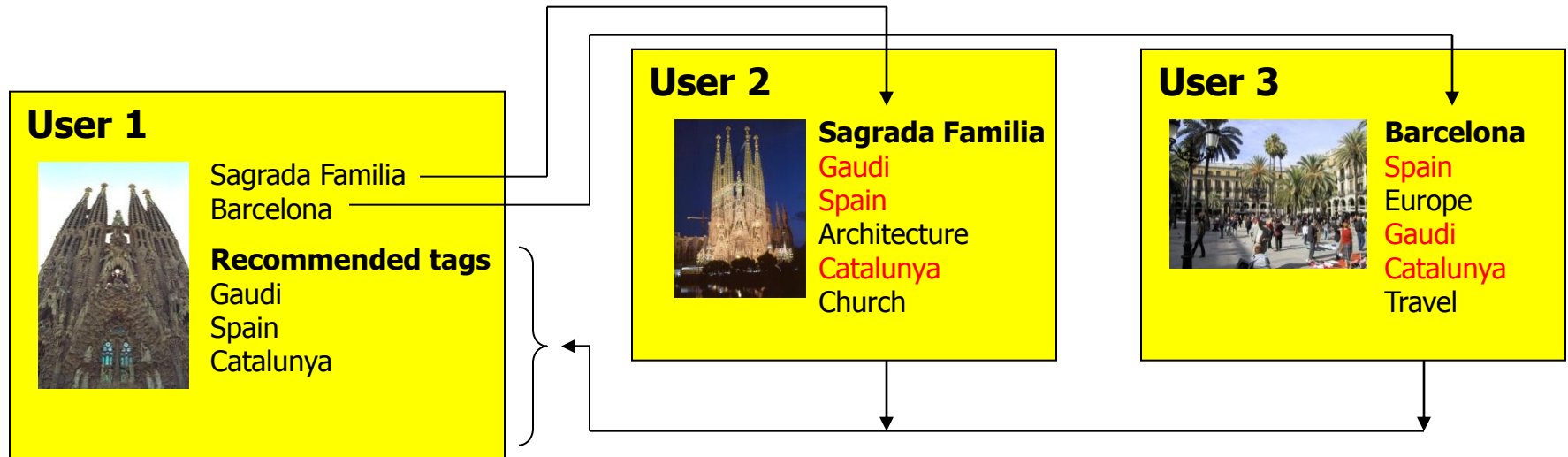
Von Ahn, IEEE Computer 2006



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Tag Recommendation

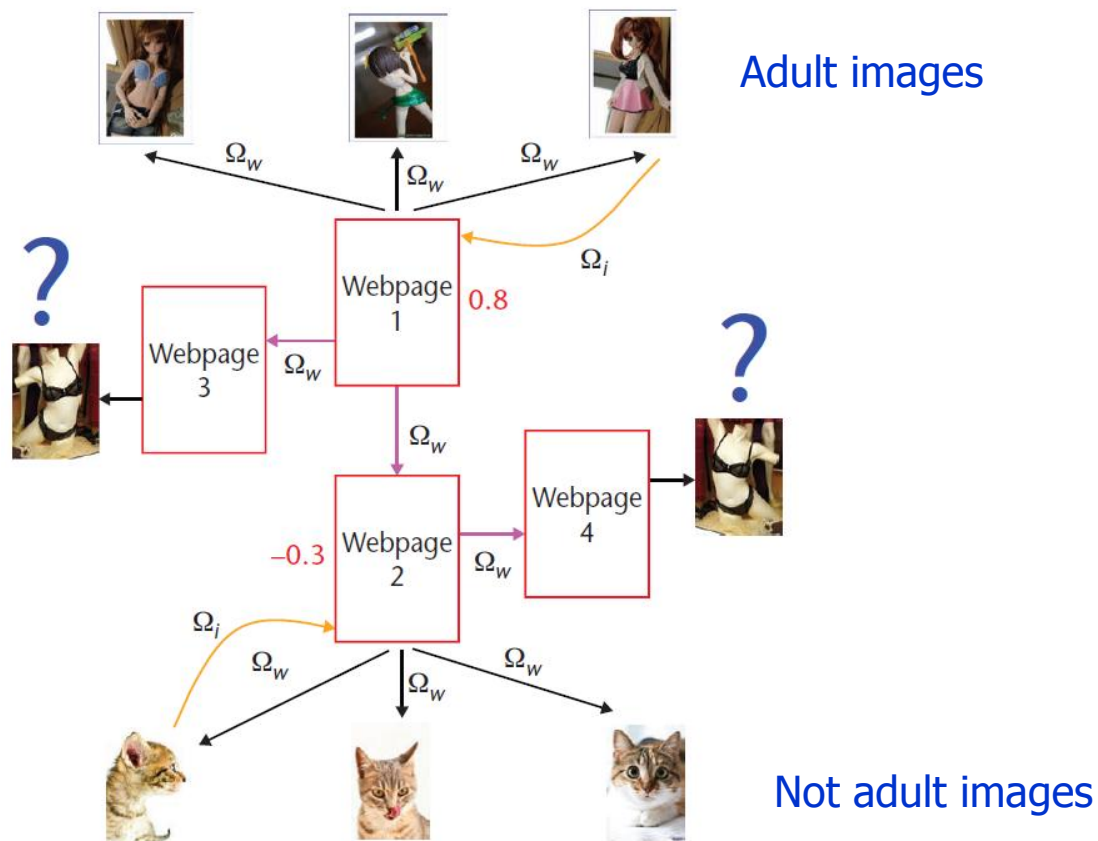
Sigurbjornsson & van Zwol, WWW 2008



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Image Classification using Web Graph

Mahajan & Slaney, ACM Multimedia 2010

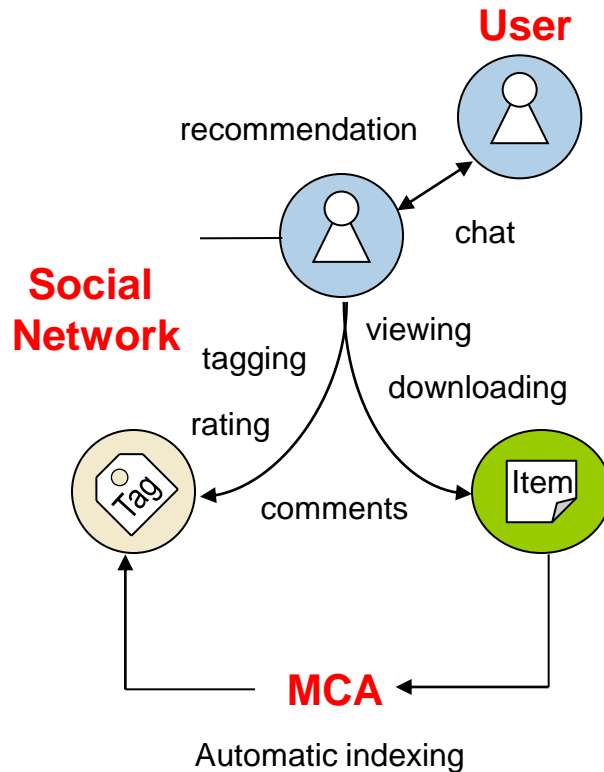


International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Trend: CONTENT + CONTEXT

Triple-Synergy Paradigm

www.petamedia.eu



user preferences

user's search intent

relations among users (friendship, trust)

authority distribution

similarity between users

item relevance

item quality (e.g. trivial?)

implicit relations among multimedia items

implicit links between items and tags

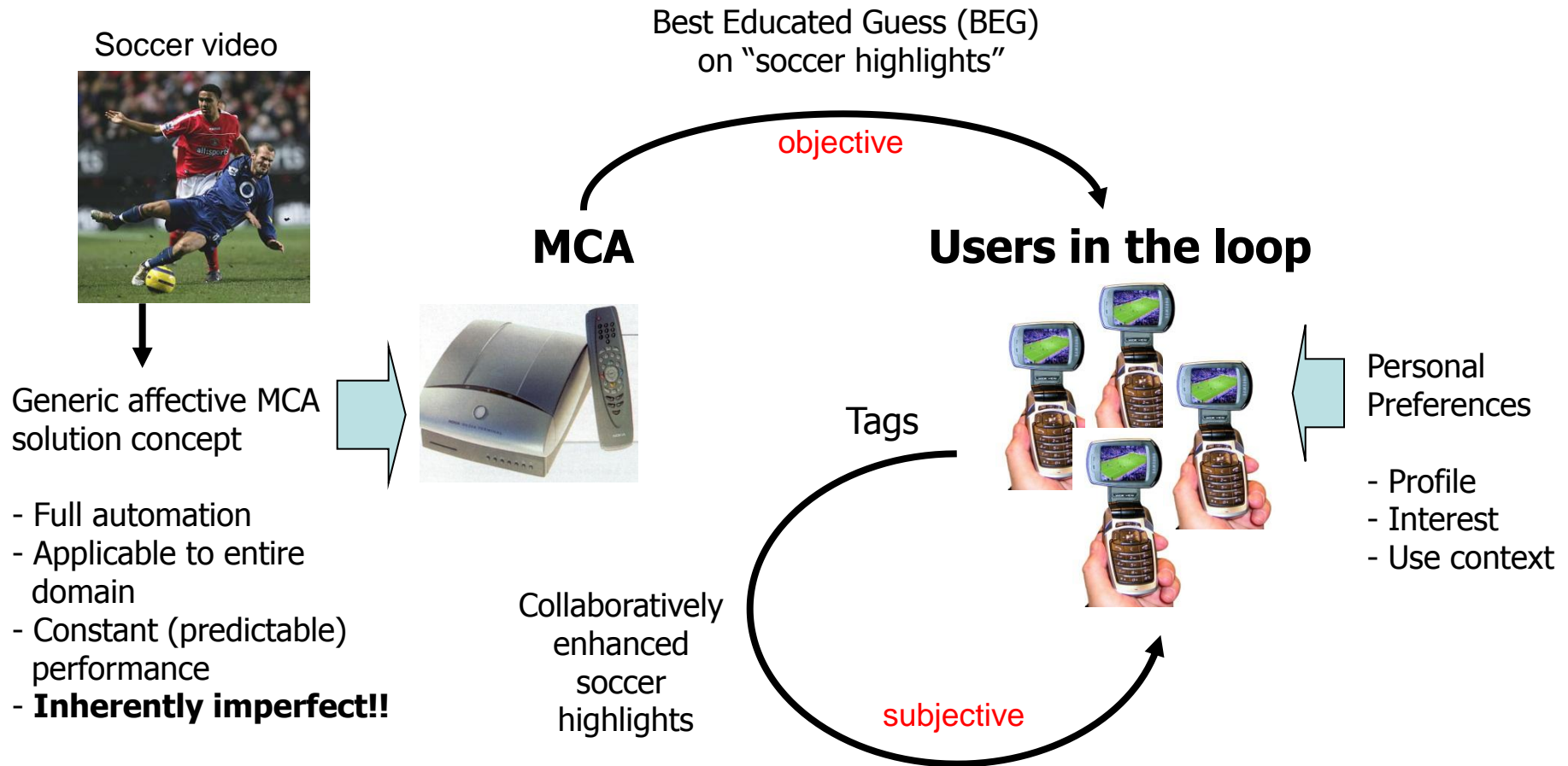
verification/enhancement of text-search or MCA results

tag relevance

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Affective MCA in a Social Context

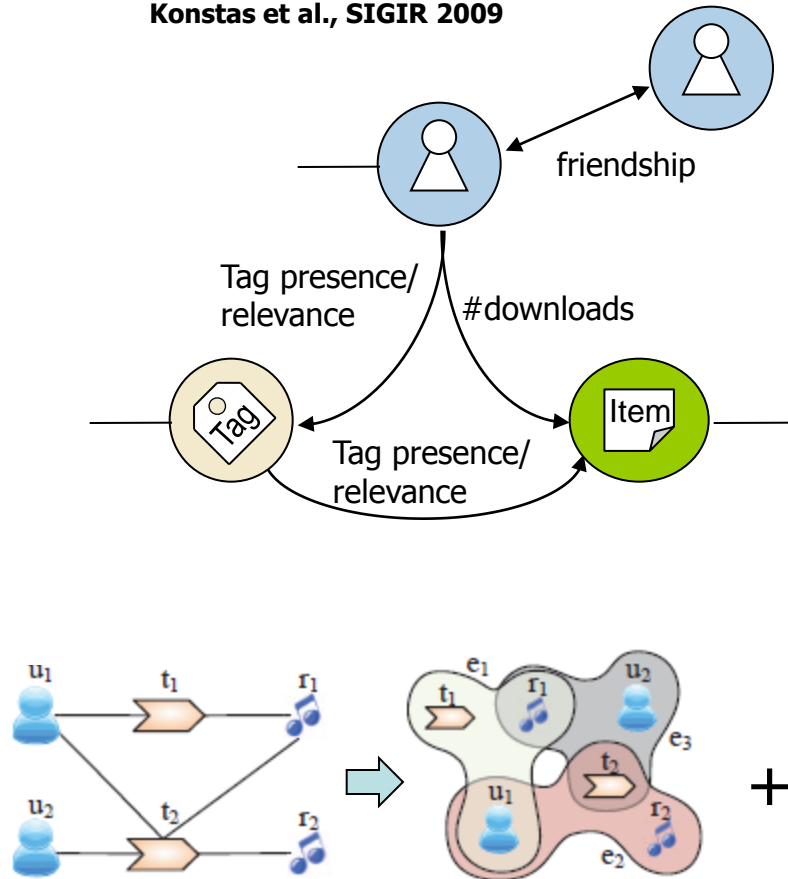
Smits & Hanjalic, IEEE Multimedia 2010



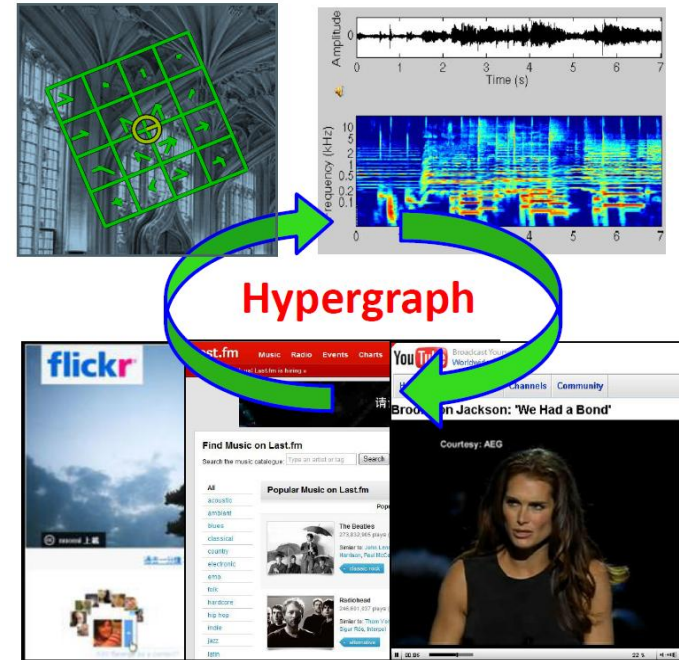
International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Recommendation using Social Graph

Konstas et al., SIGIR 2009



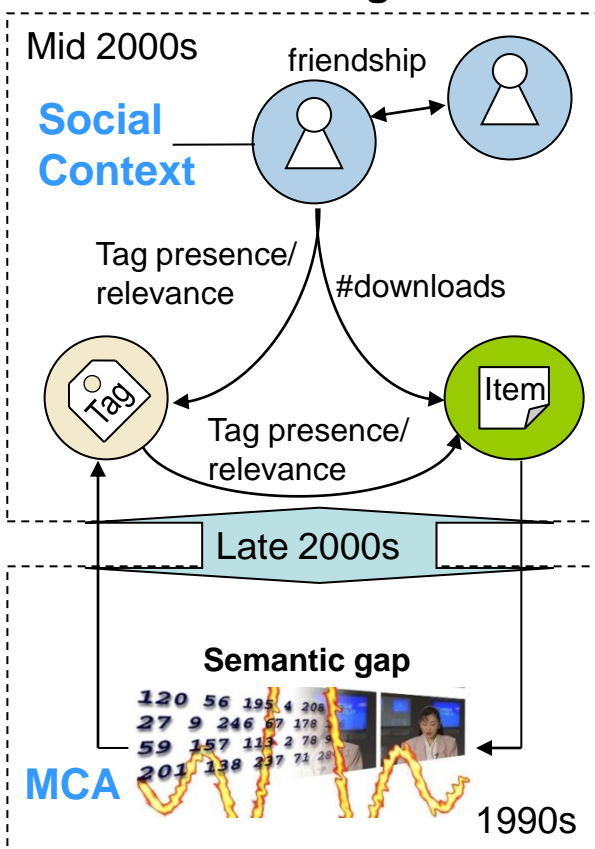
Bu et al., ACM Multimedia 2010



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Where are we now?

Resource integration



Content I like ...



... anytime ...



... and anyplace!

Utility gap

Hanjalic, IJMIR, October 2012

- The gap between the expected and de facto usefulness of MIR systems
 - Where does the gap come from?
 - Development and evaluation of MIR solutions insufficiently biased towards users
 - Departing from what we CAN do and not so much from what we SHOULD do
- **Technology inspired approach**

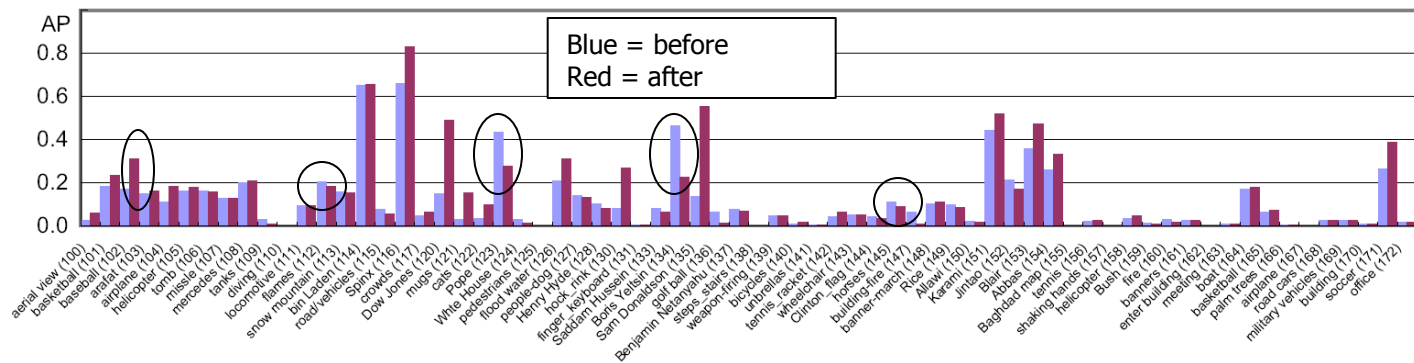
International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Technology-inspired approach

Expectations versus reality: An example

- Visual search reranking

- Approaches focusing towards optimizing the MAP, while the user is interested in the performance of the individual query
- Why not focusing on the problem of detecting whether to rerank or not?



“If MAP is your criterion, are you solving a relevant problem?”
(after Slaney, IEEE Multimedia, 2011)

Utility-by-Design


Hanjalic, IJMIR, October 2012


- Understanding the user's information need
 - Not only the "what" but also the "why" behind the search request
- Addressing the information need in all its complexity
 - From "simple" object/entity search towards really exciting search problems
- Maximizing the value offered to the user
 - From technically relevant towards maximally informative results
- Explaining the retrieval results
 - Can a search engine admit it does not know the answer and still help the user to find the right answer?


International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Search Intent: An Example

high note singing 🔍

- 

Singing Lesson: how to sing high notes!
 More **Singing** Lessons: www.gatmusic.com Private voice lessons over Skype: www.gatmusic.com if you want to learn faster as well as learn more ...
 by abgar8 | 3 years ago | 702,337 views
- 

IU sings 3 high notes ..
 by OkSincha | 11 months ago | 491,631 views
- 

Adam Lopez - Highest Vocal Note- Guinness World Record
 Adam Lopez - Guinness World Record holder in the category of "Highest vocal note by a male". Recorded live at the Channel 7 Studios, Australia, 2005
 by AdamLopezMusic | 3 years ago | 1,304,968 views

Tutorial video
 (~ Learn something)

Entertainment show
 (~ Being entertained)

Informational show
 (~ Getting informed)

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Search intent: Two sides of a coin

- **User side**

- Discovering what the user had in mind when inserting the query
- **Search intent inference**

- **Collection side**

- Discovering what multimedia item would respond well to a given inferred intent
- **Search intent classification**



Search intent classification

Example approach

Hanjalic, Kofler, Larson, ACM Multimedia 2012, Brave New Ideas Session, to appear

- **Discovering intent categories through social-Web mining**
 - Inspired by Broader's Taxonomy of Web Search, done by analyzing expressions of information needs from Yahoo! Answers and via crowdsourcing
- **Five discovered categories**
 - **Information** (improving the knowledge about a topic)
 - **Experience - Learning** (acquiring skills, learning "how" by experience)
 - **Experience - Exposure** (watching how people experience situations)
 - **Affect** – watching a video to change one's mood
 - **Object** – using a video as a "tool" or to serve a particular purpose
- **Discovering features for automatically inferring intent category**
 - Style, vocabulary and phrases in the spoken content, patterns in tempo, duration and frequency of video shots

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Really exciting search problems?

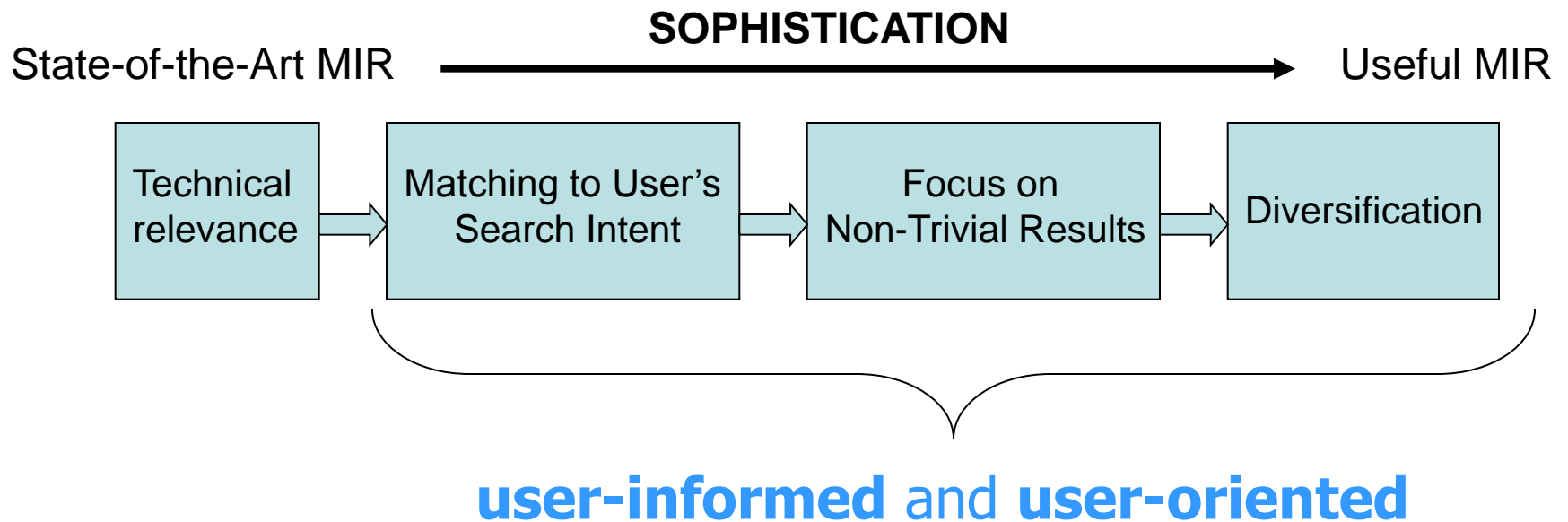
- Give me all recently uploaded videos about XYZ!
 - The emphasis is on ABOUT!
- Where was this photo/video taken?
 - No constraints!
- Give me more YouTube results like this one!
 - ...by taking my local, affective and “deep” personal context into account
- I want to know more about this location!
 - How to get the application like Google Goggles to always work well?



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

From relevant to useful search results

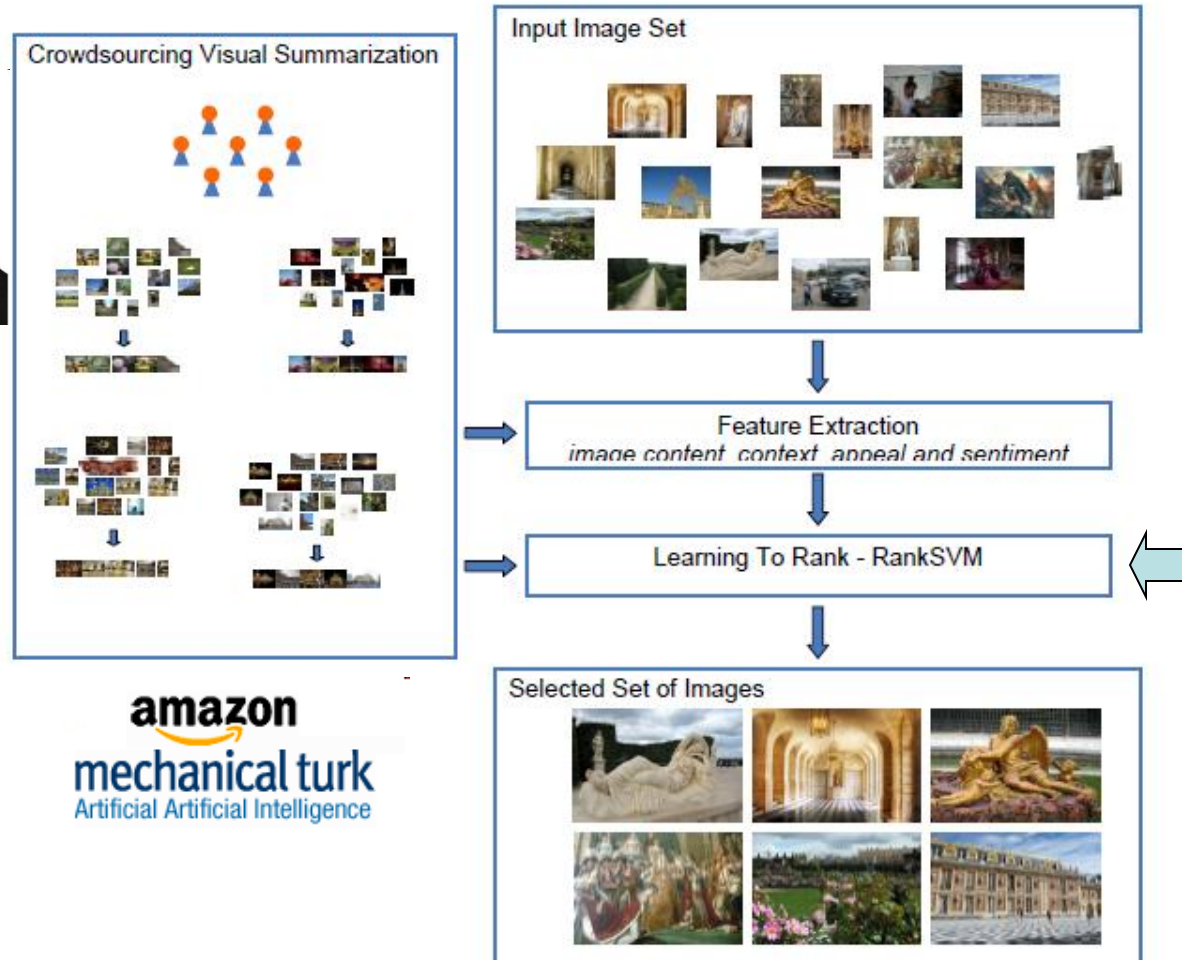
Hanjalic, IJMIR, October 2012



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

User-Informed Summarization of Image Collections via Crowdsourcing

Rudinac, Larson, Hanjalic, submitted to IEEE Trans. on Multimedia



amazon
mechanical turk
Artificial Artificial Intelligence

Explainability of the search output

Hanjalic, IJMIR, October 2012

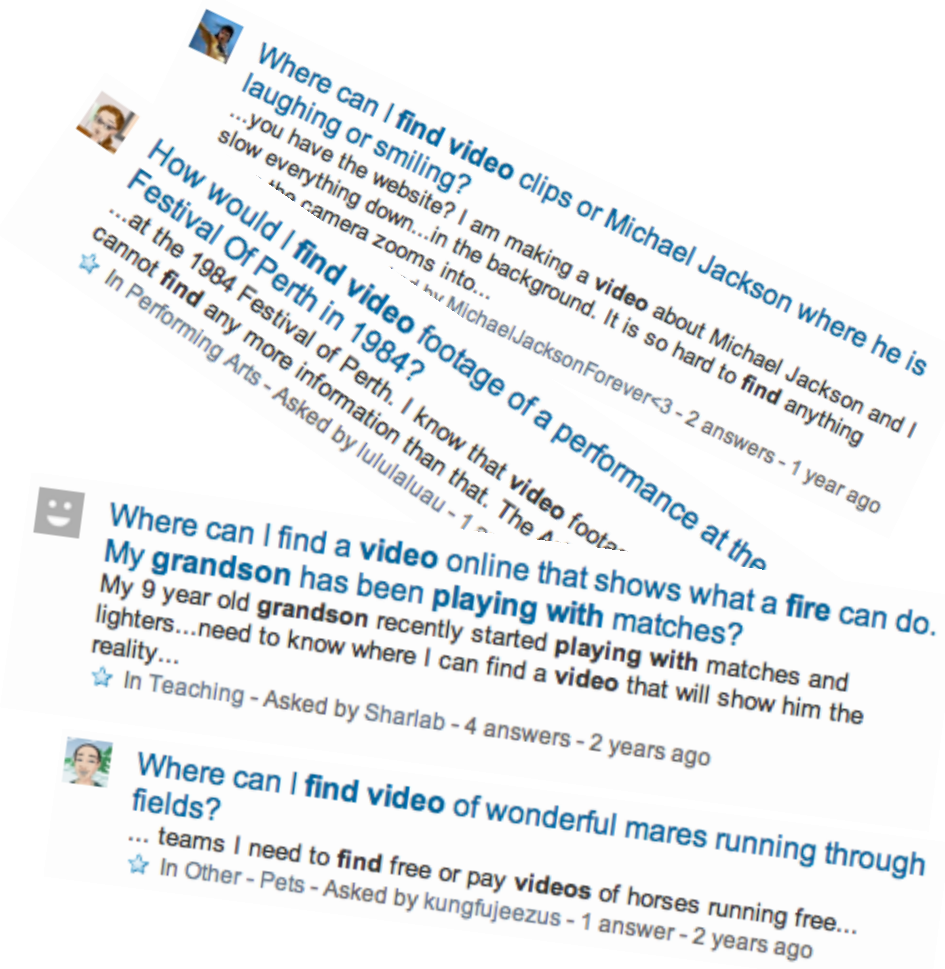
- Usefulness of a search results list is measured by the extent to which it helps the user in her further actions
- We can't guarantee optimal results list, but we can try to provide information
 - how reliable the results are and
 - where the imperfections/ambiguity come from
- Information can be
 - offered to the user to guide the query adjustment or selection of an alternative search mechanism
 - used internally by the search engine to make decisions in ambiguous info-spaces

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

Predicting Web MM Search Failure

Kofler, Larson, Hanjalic, ACM Multimedia 2012, to appear

- What are typical failure-prone queries?
- Can we predict such queries?
- Approach by
 - analyzing the search log
 - deriving information from the log to capture patterns of query (re)formulation and the click-through data in a search session



International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium

To take away

- Continue to benefit from resource integration
 - Imperfection (algorithm) → Adequateness (user)
- Make your approach more user-informed
 - Great potential of crowdsourcing still waits to get exploited!
 - Great challenges of deploying crowdsourcing input in developing automatic MIR solutions
- Keep the user central and avoid technology-push pitfalls
 - User (interaction) data still under-represented compared to multimedia data
 - Evaluate user experience and not algorithmic performance alone

Thank you

Multimedia Signal Processing Group

<http://msp.ewi.tudelft.nl>

Delft Multimedia Information Retrieval Lab

<http://dmirlab.tudelft.nl>

Alan Hanjalic

- URL: <http://dmirlab.tudelft.nl/users/alan-hanjalic>
- Email: A.Hanjalic@tudelft.nl

International Workshop on Search Computing, 25-26 September 2012, Brussels, Belgium