

IM2 Phase II

Multimodal Content Abstraction (IP.MCA)

Structure and Achievements
2009



FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Stéphane Marchand-Maillet

Viper - UniGE

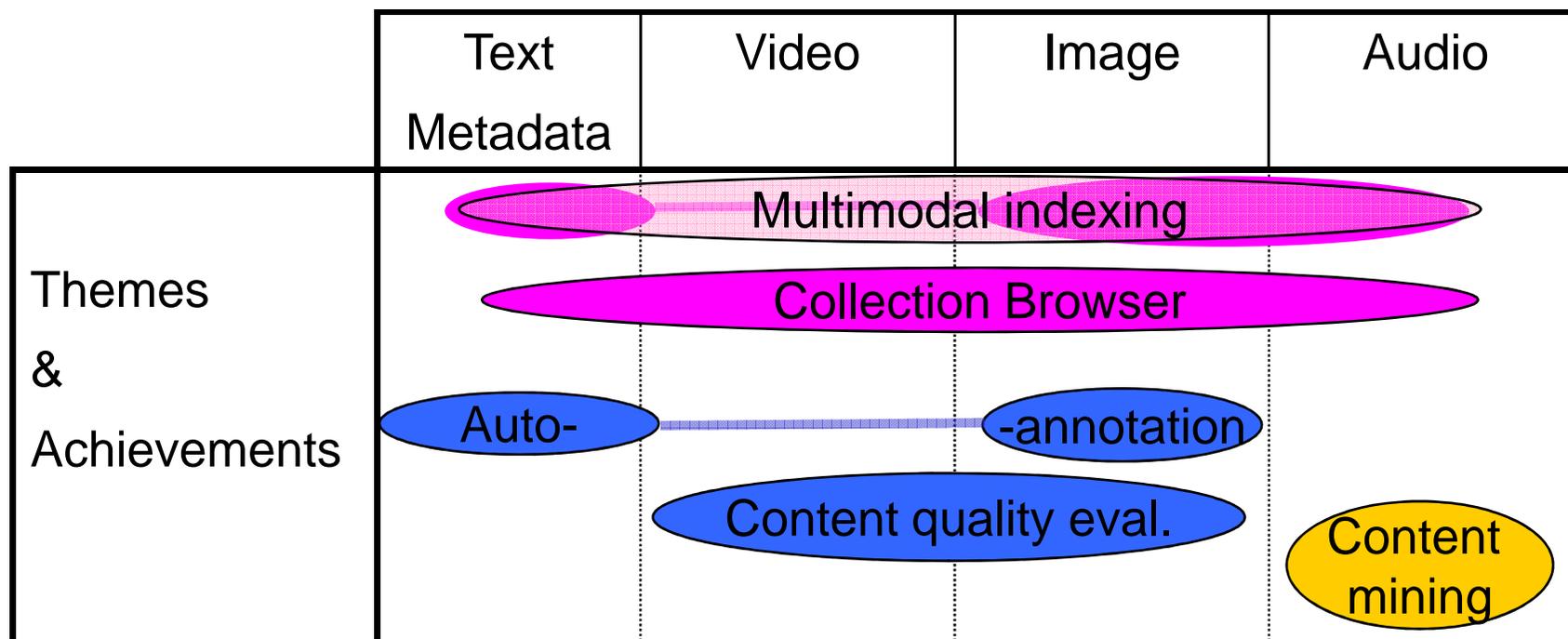


IM2 MCA Focus

Content-based indexing, retrieval, and semantic analysis and mining of multimedia data based on multiple modalities.

- Search and Retrieval
 - UniGE, EPFL, IDIAP
 - Document Content Abstraction
 - UniGE, EPFL, ETHZ
 - Multimedia Content Mining
 - IDIAP
- ⇒ Overall 7 tasks related to content analysis and indexing

IM2.MCA: Scope



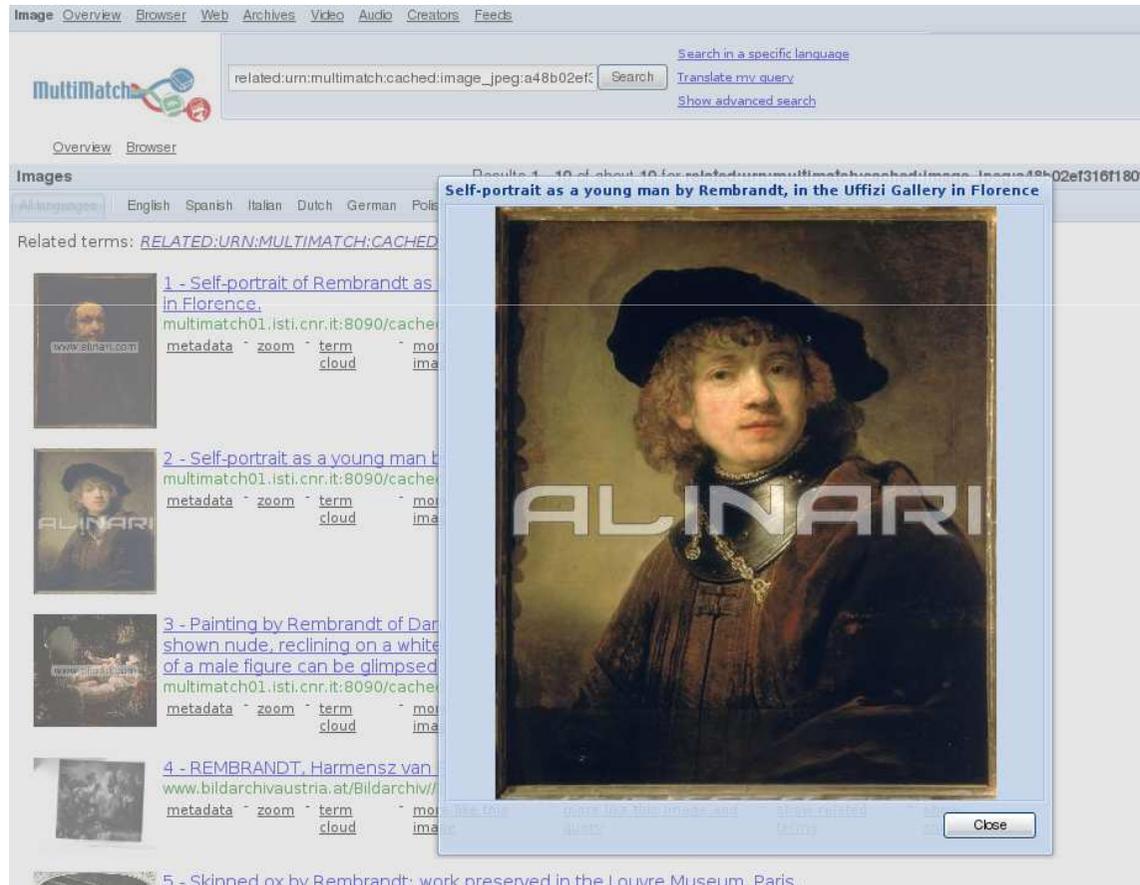
Tasks:

- A. → Content indexing + user interaction
- B. → Abstraction from content and/or usage
- C. → Content mining

Cross-modal Search Engine (A1)

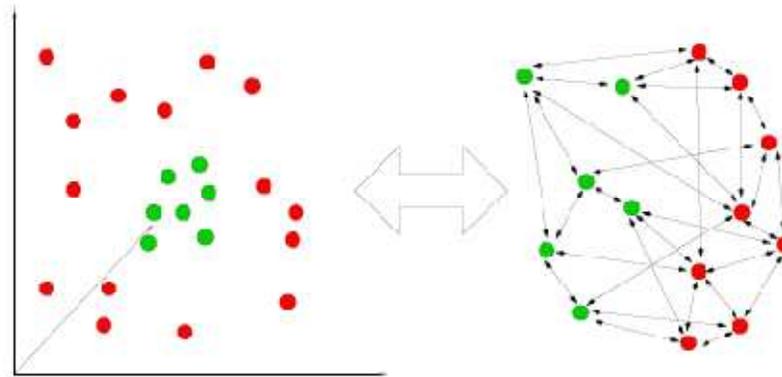
Search a collection by fusing text, images and relevance feedback

Cultural Heritage Search Engine (MultiMATCH) powered by IM2 technology

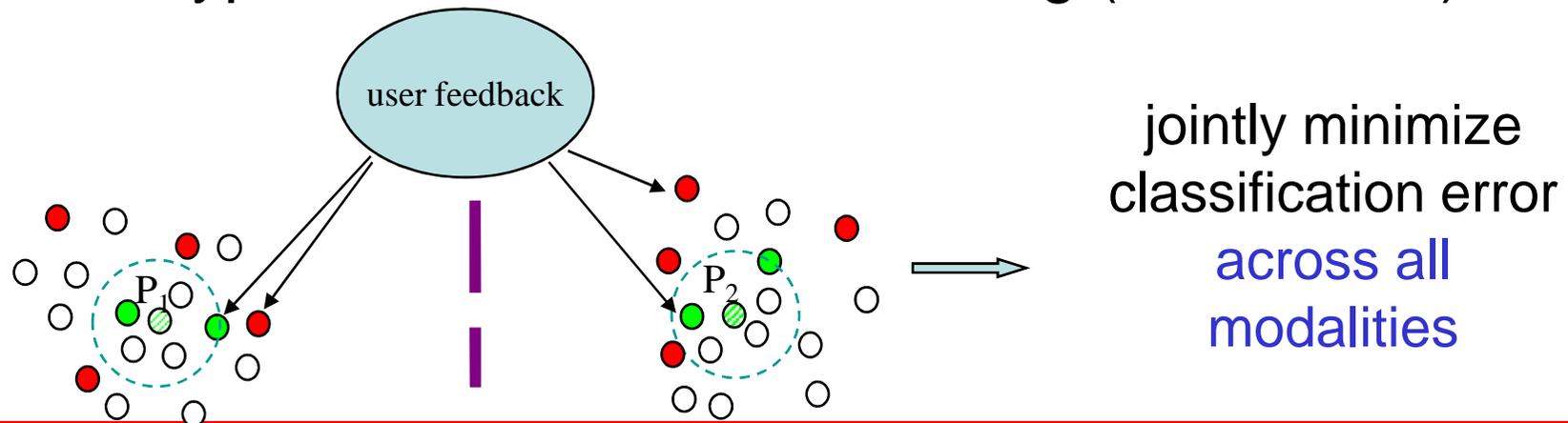


Cross-modal Search Engine (A1)

- Transform heterogeneous features into homogeneous similarities



- Prototype-based fusion and learning (**RankBoost**)



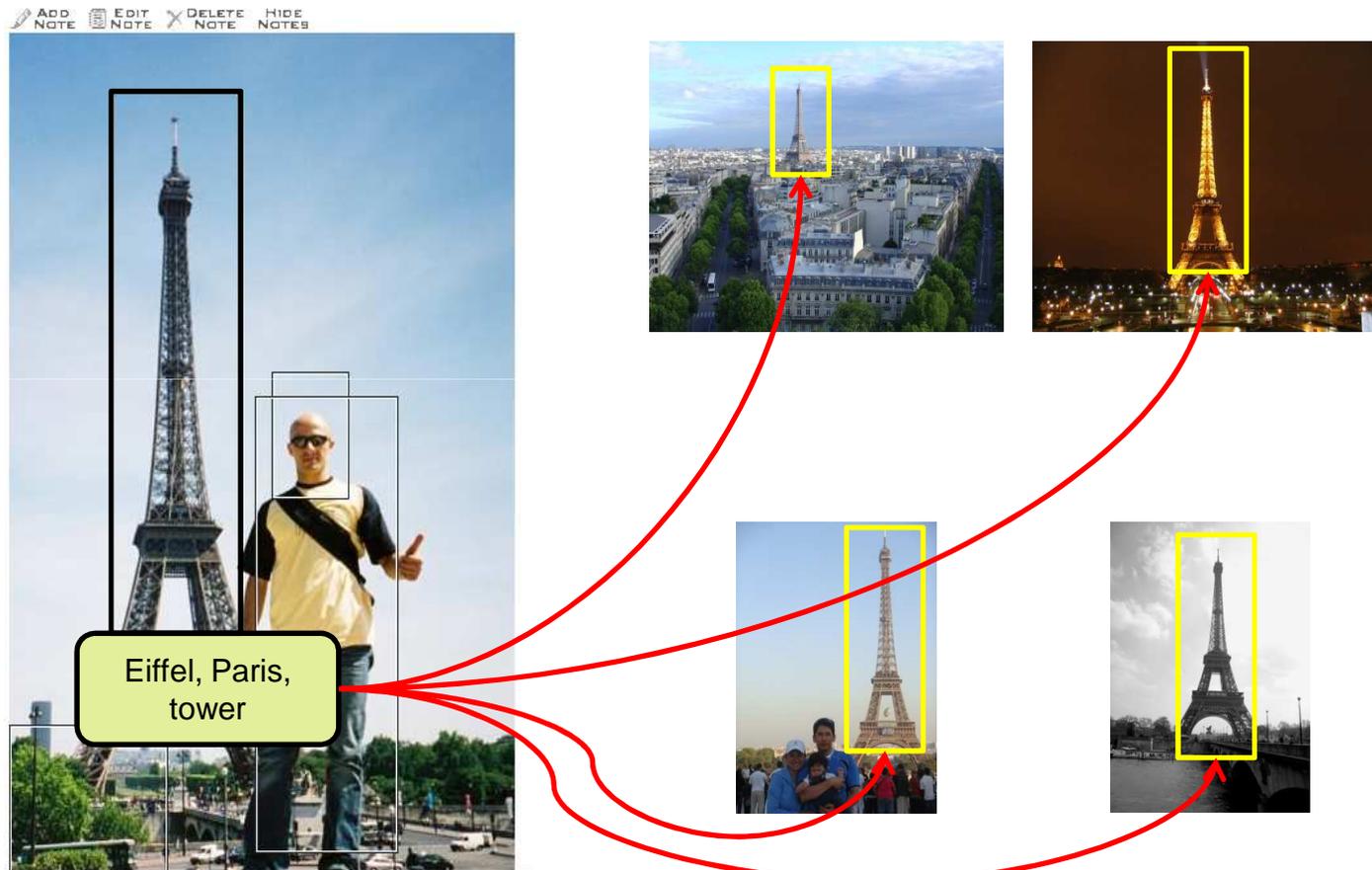
Tagged Media-Aware Multimodal Content Annotation (A2)

Goal: image annotation and tag propagation

- **Rich user interaction:** Add tags, mark and label objects, search annotations, browse
- **Object duplicate detection** algorithm (SURF features) used to propagate tags to all detected images with the same object in the dataset
- **Social Media Tag Demonstrator** – web-based tool available online at: <https://ltslinux18.epfl.ch/mmspg6/image>



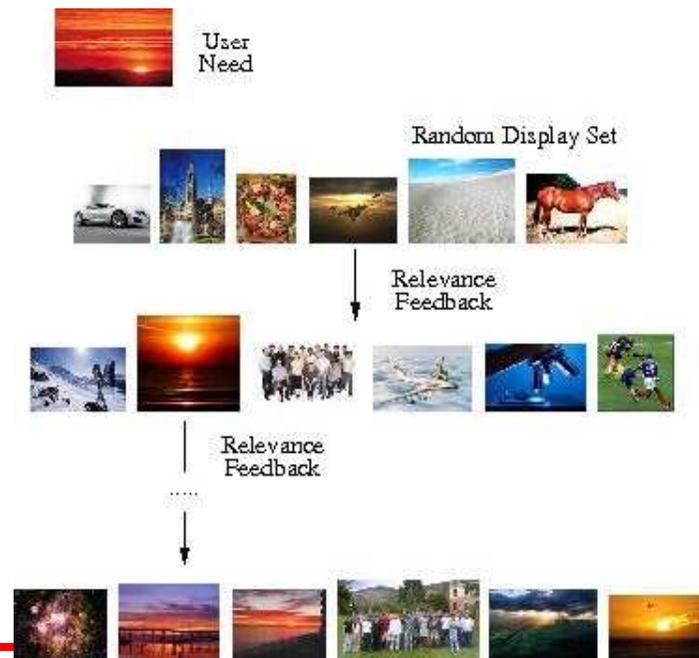
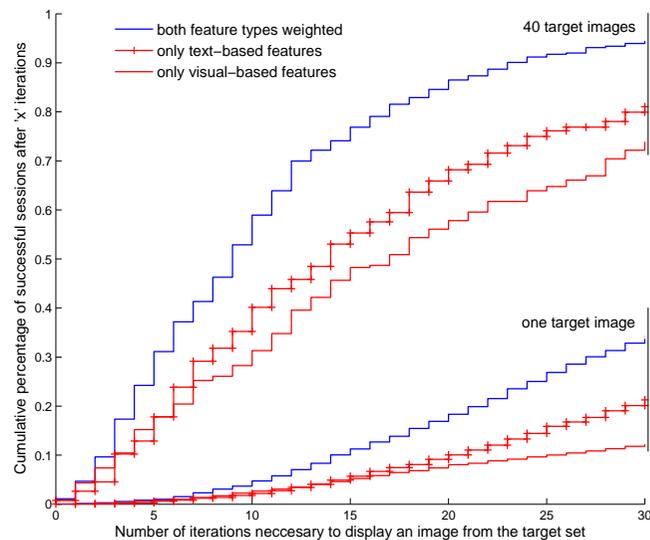
Social Media Tag Demonstrator (A2)



Multimodal relevance feedback (A3)

Multimodal retrieval approach combining **textual and visual features** extracted from annotated media (medical images accompanied by reports, songs with lyrics, etc.):

- **No query image required**
- **Automatic selection of features** most likely to account for user needs



Automated content description (B1)

Make use of media usage (retrieval engines' logs) to infer document description

- Determine image similarities using using techniques such as SVD, NMF, PLSA-based User Relevance Model
- Propagate annotations and meta-data over similarity graph

Evaluation over community-based annotation efforts (Corel, Flickr,...)

Automated content description (B1)



(deer, grass, water, white-tailed)



forest, snow, trees, wolf
(grass, shade, trees, wolf)



bear, river, snow
(bear, grizzly, stream, water)



head, lion, mane, rocks
(cats, field, grass, lions)



dust, elephant, sky, water
(bull, elephant, sky, water)



grass, hippo, pair, river
(grass, hippos, wallow, water)

Predicted tags above, groundtruth tags in parentheses

Multimodal quality metrics (B2)

Challenge: *Measurement of perceived quality* plays a fundamental role in the context of multimedia services and applications.

Focus:

- Subjective quality assessment of high resolution still pictures in the context of performance evaluation and comparison of compression algorithms.
 - Subjective and objective quality assessment of visual and audio-visual sequences, compressed and transmitted over error-prone networks, in multimedia applications.
- Qualcomm best student paper award at the International Conference on Quality of Multimedia Experience (Qomex2009) for the paper "Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel"
-

Multi-modal auto-annotation (B3)

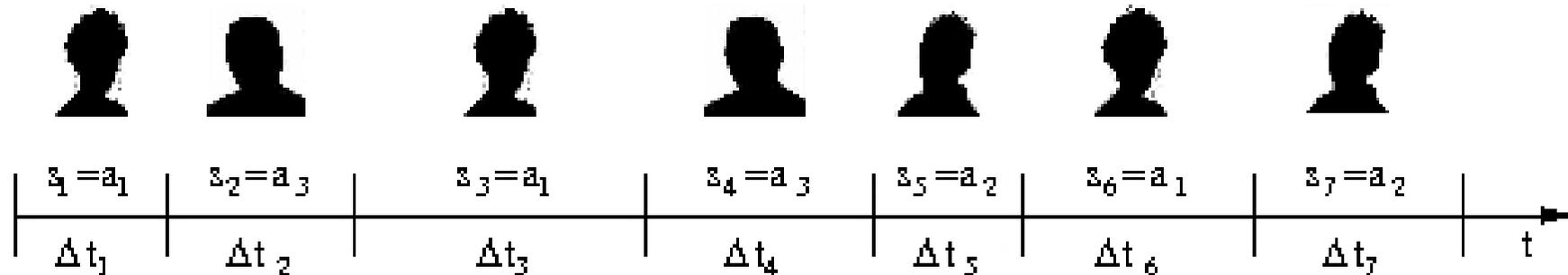
- Multi-modal crawling technology to create reference database for visual search from online photo sharing communities
- Auto-annotation system for holiday snaps building on the multimodal reference database

The screenshot shows the photocontext interface with a central photo of the Arc de Triomphe. The interface includes a navigation bar at the top with links for 'photocontext', 'Auto-tagging', 'View & Share', 'Explore', 'Settings', and 'Logout'. The main content area is divided into several sections:

- Object-level annotation with bounding box:** A yellow bounding box is drawn around the top part of the Arc de Triomphe.
- Object title from Wikipedia article:** A yellow box highlights the text 'Arc de Triomphe' at the bottom of the photo.
- Mined Wikipedia Pages:** A list of related websites including 'Arc de Triomphe', 'Triumphal arch', 'Paris', 'Civil religion', and 'The Amazing Race 1'.
- Related tags from itemset mining:** A list of tags including 'paris', 'arc triomphe', 'arc detriomphe', 'arch', 'monument', 'francia', and 'toile'.
- Location estimated from mean location of photos in cluster:** A map showing the location of the Arc de Triomphe in Paris.
- Other users who have taken photos of the same object:** A list of users, including 'the_cha' who took a photo on '28.10.2006'.

At the bottom of the interface, there is a small disclaimer: 'This product uses the Flickr API but is not endorsed or certified by Flickr. Flickr | Our Privacy Policy | Terms & Conditions'.

Social Network Analysis (C1)

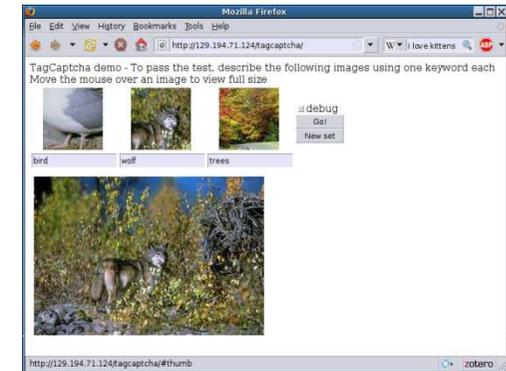
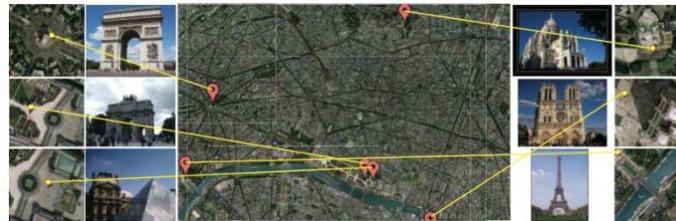
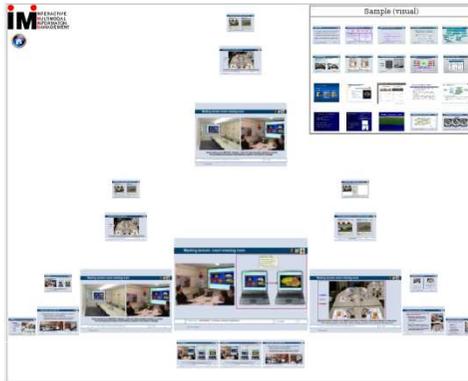


Extraction of **socially relevant information from turn-taking patterns**:

- **Role recognition** with Bayesian approaches and Hidden Markov Models (more than 80% accuracy over 90 hours of material)
- **Detection of agreement and disagreement** in competitive discussions (70% accuracy over 30 hours of material)
- **Detection of story related social groups** in conversations (topic purity of 0.75 over 27 hours of material)

Further contributions

- Research prototypes



- International visibility
 - 22 Publications (inc. 6 journal/chapters)
 - ImageCLEF (MM Wiki Track)
 - Projects, Networks
 - ACM CIVR2009
-

Plans for future

- Multimodality is key for IM2.IP1 in Phase 3
- Some of the work will continue
- Emphasis will be placed on technology transfer

Thank you...

Questions?