MediaEval Multimedia Benchmark Initiative
Looking back and looking forward to year five

Martha Larson, Delft University of Technology
Search Computing and Social Media Workshop
NEM Summit, 28 September 2011, Torino Italy

TUDelft

Overview

- What is MediaEval?
- What is a benchmark?
- What happened in 2011?
- What will happen in 2012?

TUDelft

What is MediaEval?

- … a multimedia benchmarking initiative.
- … evaluates new algorithms for multimedia access and retrieval.
- … emphasizes the "multi" in multimedia: speech, audio, visual content, tags, users, context.
- … innovates new tasks and techniques focusing on the human and social aspects of multimedia content.
- … is open for participation from the research community

http://www.multimediaeval.org

What is a benchmark?

The Merriam-Webster Definition

¹bench·mark *noun* \'bench-ˌmärk\

Definition of BENCHMARK

1 usually **bench mark** : a mark on a permanent object indicating elevation and serving as a reference in topographic surveys and tidal observations

2 **a** : a point of reference from which measurements may be made

   b : something that serves as a standard by which others may be measured or judged

   c : a standardized problem or test that serves as a basis for evaluation or comparison (as of computer system performance)

What is a benchmarking initiative?

- A benchmarking initiative is a forum that organizes tasks for the research community.
- Researchers are invited to develop algorithms that address the tasks.
- Because everyone is carrying out the **same task** on the **same data** and uses the **same evaluation metric**, it is possible to directly compare the performance of algorithms.

Components of a task

- A task definition that describes the problem to be solved
- A data set provided to the benchmark participants
- Ground truth against which participants' algorithms are evaluated
- An evaluation metric.

Flickr: S P Photography

TU Delft

Benefits of benchmarking

- Efficient use of resources
- Reduction of duplicated research effort
- Easy entry for researchers into a new field
- Tracking improvement in the state of the art
- Stimulate industry innovation with techniques fresh from the lab
- Inspire researchers in academia to solve specific problems

MediaEval History

- Founded as VideoCLEF by Gareth Jones and myself in 2008
- Ran in 2008 and 2009 at the Cross Language Evaluation Forum (CLEF)
- Ran in 2010 and 2011 as an independent benchmark
- For 2012: we have just start process of selecting tasks
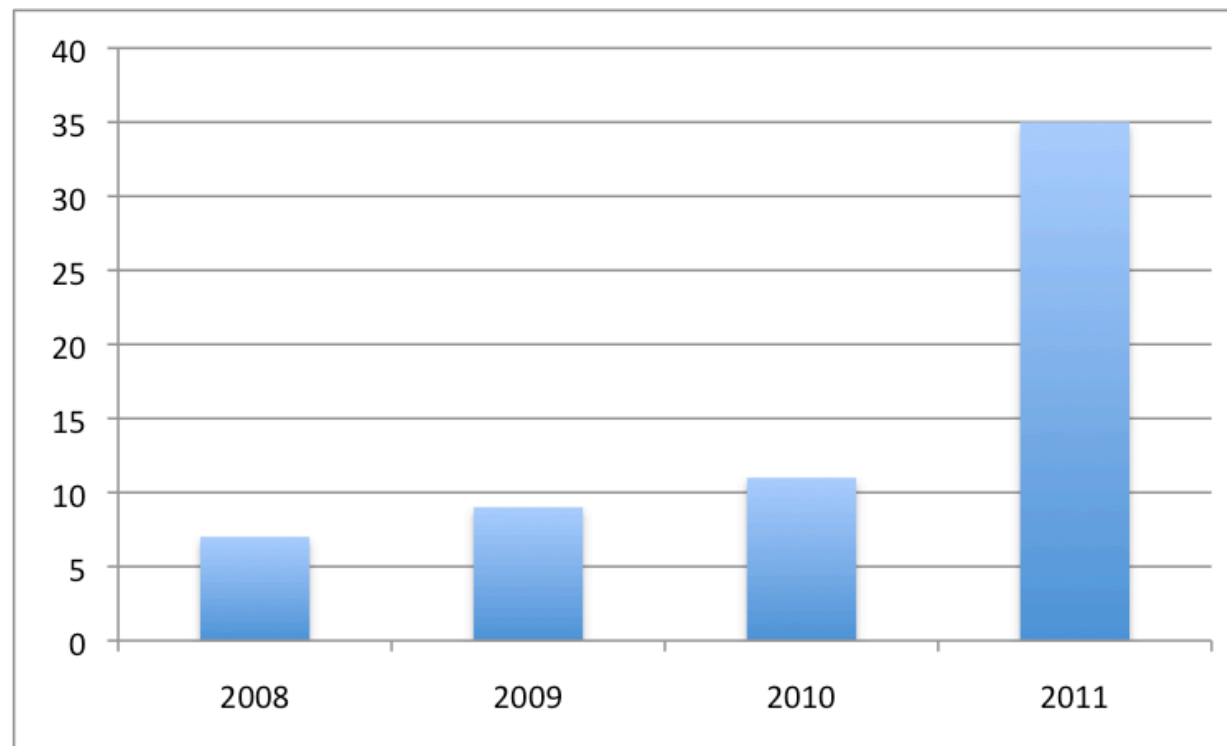
MediaEval and PetaMedia

MediaEval draws on the "Triple Synergy":
- Multimedia content analysis
- Social network structures
- User-contributed tags and annotations
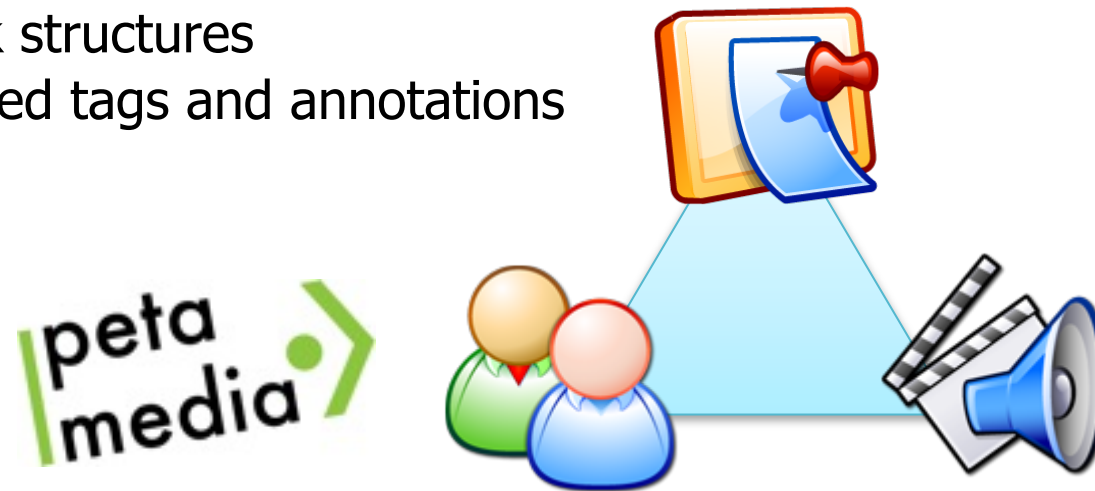
PetaMedia Network of Excellence: Peer-to-peer Tagged Media

Participating Teams

MediaEval Tasks 2011

(Number of teams who crossed the finish line for that task)

- Placing task (6)
- Spoken Web Search task (5)
- Affect task (6)
- Genre tagging task (10)
- Rich Speech Retrieval task (5)
- Social event detection task (7)

Placing Task

- **Task:** automatically assigning geo-coordinates to Flickr videos using one or more of: Flickr metadata, visual content, audio content, social information
- **Data:** Creative Commons Flickr data, predominantly English
- **Organizers:**
  Adam Rae, Yahoo! Research
  Pascal Kelm, TU Berlin
  Vanessa Murdock, Yahoo! Research
  Pavel Serdyukov, Yandex
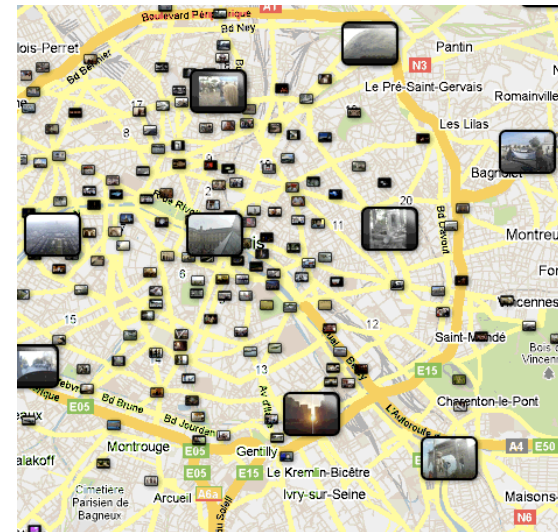
Spoken Web Search

- **Task**: search FOR audio content WITHIN audio content USING an audio content query. This task is particularly interesting for speech researchers in the area of spoken term detection.
- **Data**: Audio from four different Indian languages -- English, Hindi, Gujarati and Telugu. Each of the ca. 400 data item is an 8 KHz audio file 4-30 secs in length.
- **Organizers**:
Nitendra Rajput, IBM Research India
Florian Metze, CMU

World Wide Telecom Web
(i.e., the "Spoken Web")

- Consists of a network of VoiceSites hosted by telecom infrastructure
- Voice Sites are interlinked voice-driven applications created by users
- Spoken web is parallel and complementary to the existing WWW
- Users need only a phone (any phone!!)
- Users don't need to read

http://interactions.acm.org/content/?p=1094

TUDelft

Affect Task:
Violent Scene Detection

- **Task:** deploy multimodal features to automatically detect portions of movies containing violent material.
- **Data:** A set of ca. 15 Hollywood movies (that must be purchased by the participants.)
- **Organizers:**
Mohammad Soleymani,
      Univ. Geneva
Claire-Helene Demarty,
      Technicolor
Guillaume Gravier, IRISA

Flickr tylluan

Genre Tagging

- **Task:** Given a set of genre tags (how-to, interview, review etc.) and a video collection, automatically assign genre tags to each video based on on the combination of modalities
- **Data:** Creative Commons internet video, multiple languages mostly English
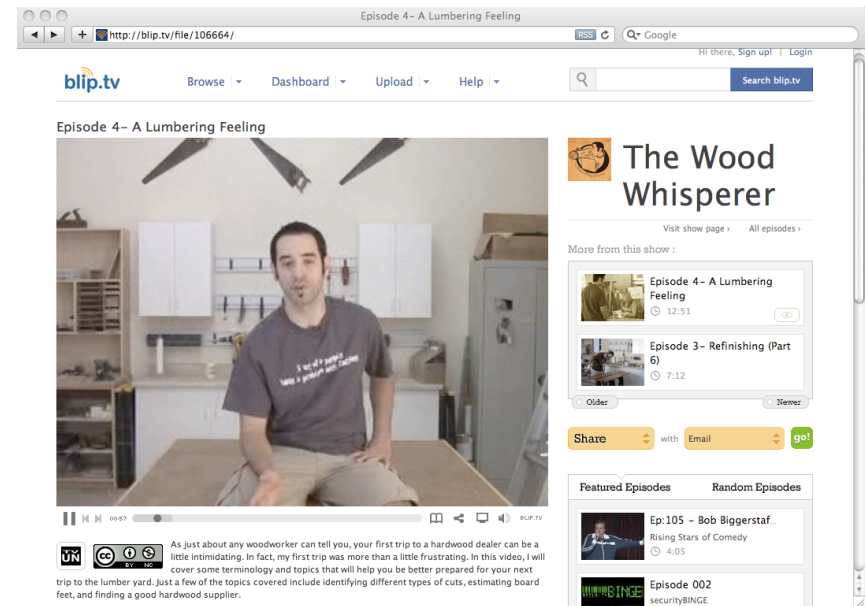- **Organizers:**

Martha Larson, TU-Delft

Sebastian Schmiedeke,

TU-Berlin

Christoph Kofler, TU-Delft

Isabelle Ferrané,

Université Paul Sabatier

Rich Speech Retrieval

- **Task:** Given a set of queries and a video collection, participants are required to automatically identify relevant jump-in points into the video based on the combination of modalities
- **Data:** Creative Commons internet video, multiple languages mostly English
- **Organizers:**
  Roeland Ordelman, Univ. Twente and B&G
  Maria Eskevich and Gareth Jones, Dublin City University

MediaEval task selection

- **Task selection process is community based**
  - First, collect task proposal from researchers and projects
  - Then, run a survey to assess task popularity and preferences
- **Tasks must have:**
  - Real-world use scenario
  - Data set that can be distributed (ideally Creative Commons)
  - A method to generate ground truth given available resources
  - "Task champions" who are willing to be task coordinators
  - Five core partners per task who are committed to task completion and to supporting the coordinators.

Social Event Detection Task

- **Task**: Discover events and detect media items that are related to either a specific social event or an event-class of interest.
- **Data**: A large set of URLs of videos and images together with their associated metadata
- **Organizers**:
  Raphael Troncy, Eurecom
  Vasileios Mezaris, ITI CERTH

MediaEval 2011 Workshop

- Held at *Santa Croce in Fossabanda – a medieval convent in Pisa, Italy* – 1st -2nd September 2011
- Official satellite event of Interspeech 2011
- 39 two-page working notes papers (13 in 2010)

MediaEval 2011 Workshop

MediaEval Project Support

- **Genre Tagging Task:** PetaMedia
- **Rich Speech Retrieval Task:** AXES and IISSCOS with support from PetaMedia
- **Affect Task:** Violent Scenes Detection: PetaMedia and Quaero
- **Social Event Detection:** PetaMedia, Glocal, weknowit, Chorus+
- **Placing Task:** Glocal with support from PetaMedia

MediaEval 2012 Schedule

- **End 2011** Survey to gather community input on proposed tasks
- **Early 2012** Decision on which tasks to offer; publication of Call for Participation
- **Spring 2012** register and return usage agreements
- **Late Spring** release of development data
- **Mid-Summer** release of test data
- **Early September** run submission
- **Mid-September** working notes paper submission
- **Early October** MediaEval 2012 Workshop

Flickr Marius B

Thank You
Questions?