



TOSCA^{MP}

Distributed search for complex heterogeneous media

**Werner Bailer, José-Manuel López-Cobo, Guillermo Álvaro,
Georg Thallinger**

Search Computing Workshop

Brussels, Sept. 26, 2012

The TOSCA-MP project receives funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 287532.



Abstract



TOSCA-MP

Media search is definitely no longer a problem of finding content in centralised in-house repositories. Due to a strengthened collaboration paradigm and increased content syndication possibilities, it rather involves large-scale repositories distributed across a network. Furthermore, these repositories are often multilingual and heterogeneous in terms of structure and data model. Journalists on location are taking over more and more tasks that were previously done by other dedicated staff at a local studio or centrally. Today, most of user generated media are images, but the technical possibilities to broadcast user generated live video content are already there.

Media content is also increasingly provided to a diversity of media platforms, remixed, repurposed, etc., combining live and offline media. This happens both as part of professional production processes as well as by users, with the boundaries being more and more blurred. The provenance information from the reuse of content is crucial in understanding the context of media content snippets and assessing their authenticity and credibility.

Central indexing is no longer an option. We thus argue that novel tools and mechanisms for networked search over heterogeneous distributed repositories need to be provided, capable of handling not only offline but also live content. A major challenge is to provide end-users with a single entry-point with the ability of seamlessly searching over distributed repositories, enabling them to aggregate and perform filtering based on characteristics or topics among the unified set of results. The system must be able to consider context and provenance information on a fine grained level, not only to reduce redundancy and establish cross-links, but also to enable assessing the credibility of the media items returned. Finally, in such environments it is of utmost importance that networked media search systems learn from successful and unsuccessful retrievals based on relevant and reliable user feedback, both implicit and explicit.

Notes



- Presentation
 - Wed. Sept. 26, session starting at 11.00
 - 18 minutes timeslot

Outline



- Context
- Challenges
- Topics addressed by current research
- Open issues beyond current research activities

Context



- Professional media industry
- Media life cycle
 - Production
 - Distribution
 - Archiving and reuse
- Many business processes involve searching for media items
 - News production
 - Archiving
 - Producing documentaries
 - ...



TOSCA-MP

Media Production

- Searching not only in repositories of own organisation, but
 - Content providers such as other broadcasters, news agencies, ...
 - Networks such as Eurovision
 - User generated content
- Efficient annotation
 - For own reuse
 - Monetising own content by enabling it is found by others
- Strengthened collaboration paradigm, increased content syndication
 - large-scale repositories distributed across a network
 - multilingual
 - heterogeneous in terms of structure and data model
 - Journalists in the field do work previously done by dedicated staff in the studio

Media Production



- User generated media
 - Increasingly important, also for inclusion in professional content
 - Technical quality reaching (semi-)professional level
 - Today: images, number of videos growing fast, but still posted as “offline” items
 - Coming: live video and audio streams
 - Issues of context, authenticity, provenance

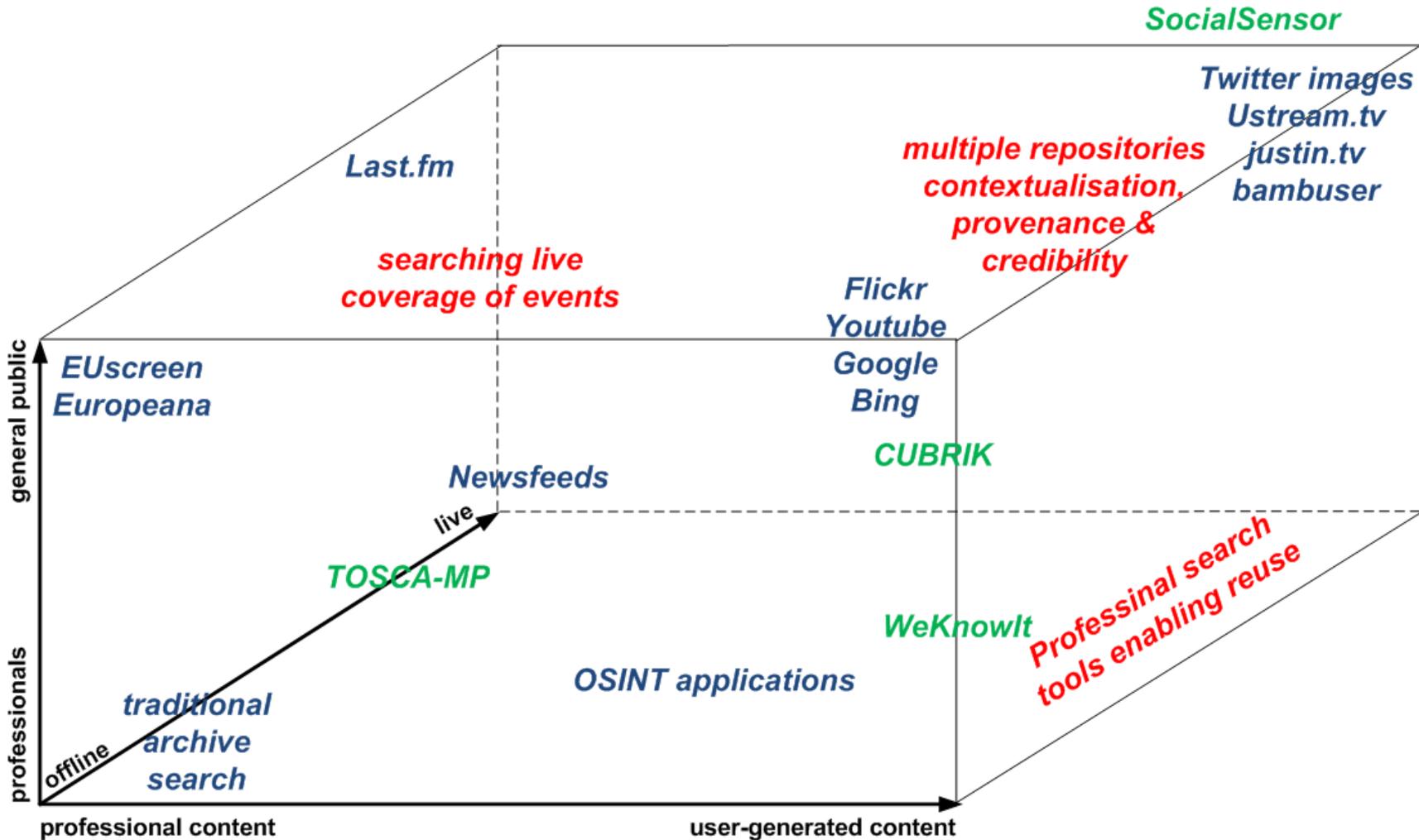
Distribution and Reuse

- media content is increasingly provided to a diversity of media platforms
- remixed, repurposed
- combining live and offline media.
- both by professionals and users
- modeling provenance, assessing their authenticity and credibility
 - archive (cf. MPDI)
 - cf. W3C ProvXG, focus on linked data, further specific work related to multimedia needed

Challenges



TOSCA-MP



Challenges

- central indexing is no longer an option
- networked search over heterogeneous distributed repositories are needed
 - combining heterogeneous result sets
 - provenance and credibility
- annotating and searching live content
 - contextualisation
 - cross-linking
- professional search tools over distributed user generated content repositories

Topics addressed by current research



TOSCA-MP

- Search across offline media in distributed heterogeneous repositories
- Integrate social context, places and events in searching offline user generated media
- Live search of user generated content

TOSCA-MP



- Task-Oriented Search and Content Annotation for Media Production
- STREP started in Oct. 2011
- enable media search in distributed large-scale repositories in the network (multilingual and heterogeneous)
- off the shelf tools, able to adapt to users' tasks are needed
- enable content holders to leverage scalable advanced distributed processing
- integrate benchmarking of indexing and search methods into the workflow

TOSCA-MP research areas



- Multilingual speech metadata extraction
- Content-adaptive visual metadata extraction and enrichment
- Aligning and linking metadata
- Task-adaptive search & retrieval and user feedback
- User interfaces for annotation and result presentation
- Task models and benchmarking
- Scalable distributed repository framework for digital media production workflows



Open issues ahead

- Professional search tools for user generated content
- Handling provenance and authenticity across repositories
- Establish links between results from different (professional/UGC) repositories

Future Search Tools

- capable of handling live content
- provide end-users with a single entry-point for seamlessly searching over distributed repositories
- enabling users to aggregate and perform filtering based on characteristics or topics among the unified set of results
- consider context and provenance information on a fine grained level
- reduce redundancy and establish cross-links
- enable assessing the credibility of the media items returned
- Learn from implicit and explicit user feedback

Questions?



TOSCA-MP

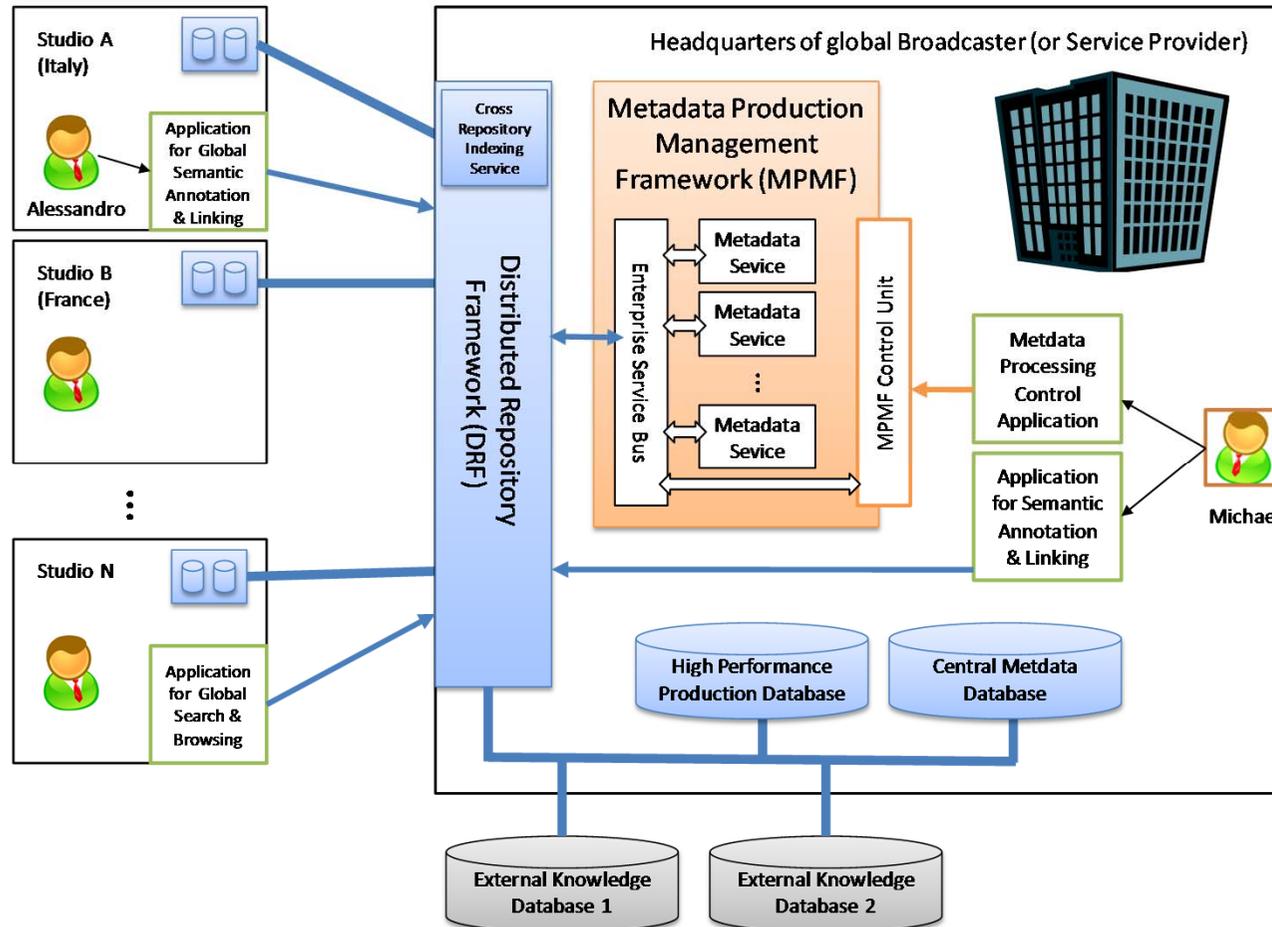
The research leading to these results has received funding from the European Union's Seventh Framework Programme under the grant agreement no. 287532, "TOSCA-MP" (<http://tosca-mp.eu>).



Scenario 1 - Annotation



TOSCA-MP



Scenario 2 – Search & Retrieval

