



TOSCA^{MP}

Task-Oriented Search and Content Annotation for Media Production

Peter Schallauer (JOANNEUM RESEARCH, Austria)

Mike Matton (VRT, Belgium)

FIAT/IFTA Conference

London, 2012-09-30

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



Outline



TOSCA^{MP}

- TOSCA-MP Overview
- Application scenarios and task models
- EBU MIM/SCAIE group

TOSCA-MP



TOSCA-MP

- Task-oriented search and content annotation for media production
- EU ICT FP7 STREP, Oct. 2011-Mar. 2014
- Consortium



Motivation & Objectives



TOSCA^{MP}

■ Motivation

- media production and archiving workflows are changing rapidly
- workflows need to be increasingly networked and distributed
- 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

■ Objectives

- enable content holders to leverage scalable advanced distributed processing
- enable editorial staff in networked media production to seamlessly use content from distributed heterogeneous repositories
- make benchmarking of indexing and search methods an integrated part of the workflow

Research areas



TOSCA^{MP}

- Media Analysis
 - 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 search
- Task models and tools benchmarking
- Scalable distributed repository framework (indexing and search) for digital media production workflows
- Results are integrated in a SOA based framework, compatible with EBU/AMWA FIMS

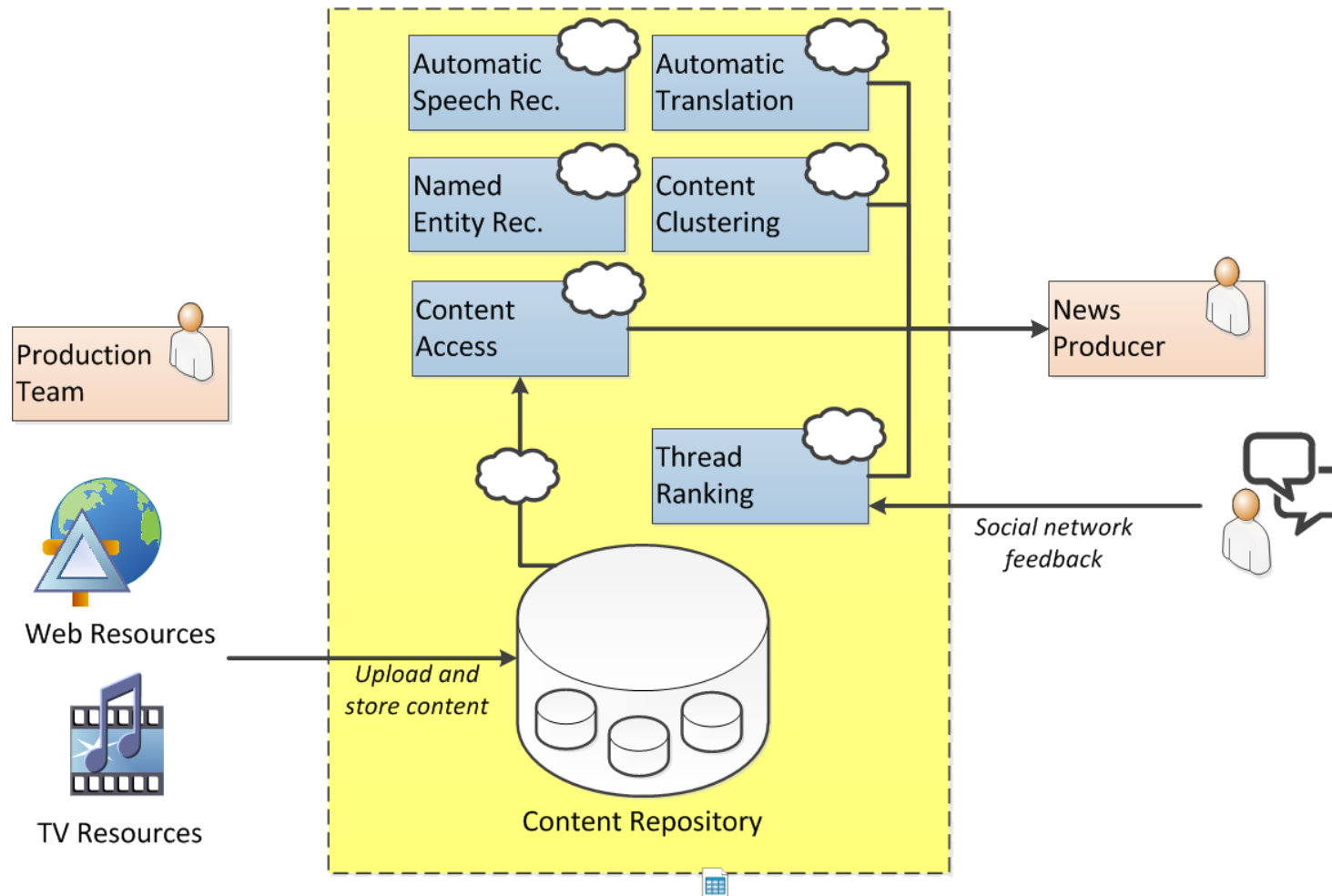
Usage Scenarios



TOSCA-MP

- 10 Business goals
 - High level goals to be achieved by broadcasters, service providers...
- 15 related scenarios
 - Concrete examples, how a business goal is achieved with tasks and actors involved
- Four categories
 - content access & retrieval
 - assisted production
 - news service distribution
 - infrastructure (repositories)
- Details on website (D6.1)

Example: “Assisted production of news stories using distributed multilingual sources”

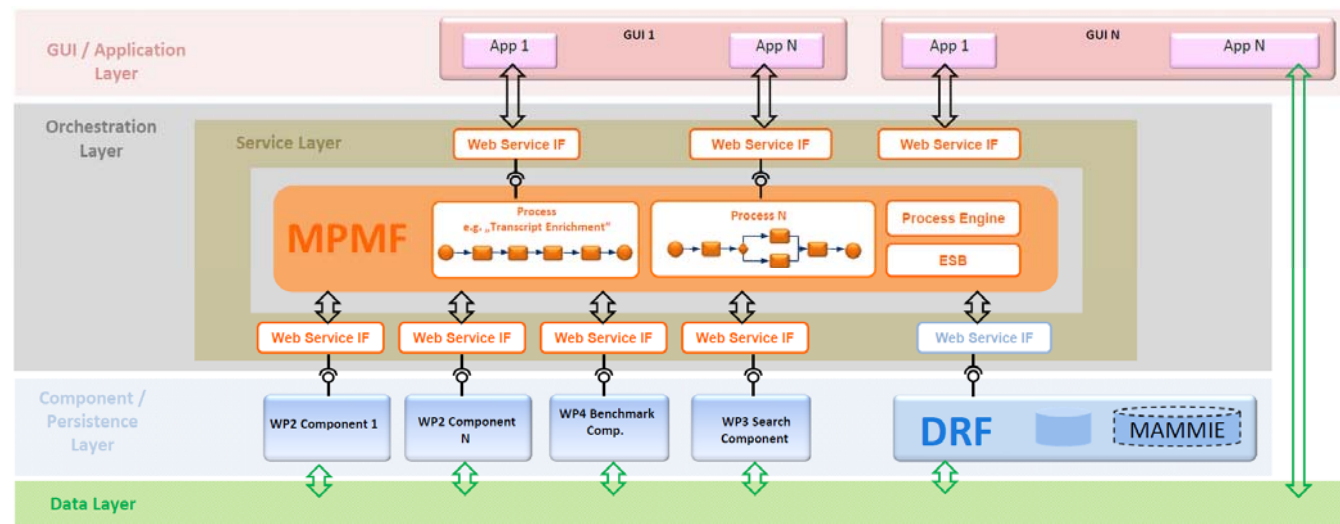


System architecture



TOSCA-MP

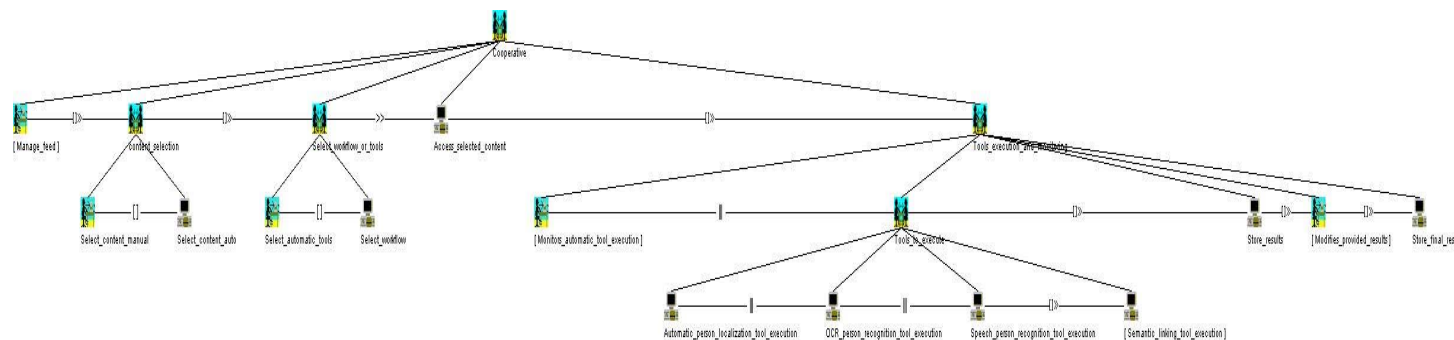
- Four subsystems
 - Metadata Production Management Framework (MPMF)
 - Distributed Repository Framework (DRF)
 - sets of services
 - graphical user interfaces
- SOA based framework
- Compatible with EBU/AMWA FIMS



Task models



- Collecting a set of high-level task descriptions from users
 - e.g. specific types of annotation, specific types of search, content summarisation
 - <http://tosca-mp.eu/tasksurvey>
 - **Survey still open! Contribute information about your tasks.**
- Create formalised task models (ConcurTaskTrees, CTT)



- To automate the validation of tools in different media production workflows (scenarios)

EBU MIM/SCAIE group



TOSCA^{MP}

- Group on automatic metadata extraction
 - Study usefulness of AIE tools in media production and archiving (benchmarking)
 - <http://tech.ebu.ch/groups/pscaie>
 - Standardization of time-based descriptive metadata
 - Published request for technologies (RfT)
- Link with TOSCA-MP project
 - Addresses automatic information extraction technologies in RfT
 - Develops novel benchmarking methods for assessing the value of automatic metadata extraction tools in media production
 - Mainly technological development
 - Know how will be shared with SCAIE group

EBU MIM/SCAIE group



TOSCA^{MP}

- MPEG-7 Audiovisual Description Profile (AVDP)
 - Tailored to media production and archiving applications
 - Standardised by ISO in 2012
 - Joint work of EBU, NHK, RAI, VRT, JRS
 - Allows metadata interoperability between AIE tools and systems
- Technology evaluation
 - Request for technologies
 - Collect content + ground truth
 - Evaluate (benchmark) technologies

EBU MIM/SCAIE group



TOSCA-MP

- Effort to collect content + ground truth data
 - Goal: evaluate / compare automatic information extraction tools of different providers (industry and research institutes)
 - Example tasks
 - Speech recognition
 - Copy detection
 - Content summarisation
 - Video OCR
 - Speaker recognition
 - Etc...
- Content made available for research & evaluation by EBU and content providers via mammie platform

EBU MIM/SCAIE group



TOSCA-MP

- The mammie platform

EBU MIM-Scaie Media Exchange Platform

Create New
Log Out
Help

Explore all

Free search

search

Filter by Contributor

RAI	100
VRT	64
NHK	44
IRT	39
DR	27


Filter by Task Set

Filter by Available Ground Truth

Showing items 1 - 10 of 274 found matching *.*.

Click an item's thumbnail to play the content or click the title to show a detailed view.

 [324009_2_journal_d_f4003_28122003_11.mpg](#)
Added by IRT on 06 November 2009

 [327916_3_journal_eng_f4612_30082005_18.mpg](#)
Added by IRT on 06 November 2009

 [326147_3_journal_f4365_26122004_22.mpg](#)
Added by IRT on 06 November 2009

 [327904_3_journal_spa_f4610_28082005_01.mpg](#)

EBU MIM/SCAIE group




TOSCA-MP

- The mammie platform: item detail & download

JSB20030626.mpg
Added by NHK on 06 November 2009
[delete item](#) | [back to overview](#)

ID: [ifip:cbc98560-ad10-012c-e5fd-00145e2a38b6](#)



00:00 00:00

description

contributor:
(click to enter contributor)

coverage:
(click to enter coverage)

creator:
NHK

date:
June 26, 2003

format:
MPEG-1

language:
English, Japanese

relation:
(click to enter relation)

rights:
NHK

source:
(click to enter source)

subject:
Japan's Scenic Beauty (Niigata/Sado Island)

type:
(click to enter type)

Download the item: [high-res](#) | [low-res](#)

EBU MIM/SCAIE group







TOSCA-MP

- The mammie platform: ground truth

Ground Truth

Below you see an overview of the ECM Scaie tasks and the sets this item belongs to (green background). If ground truth data for training is available, you can download it by clicking the link on the right.

 ITEM SEGMENTATION use for train: yes use for test: no	source Ground Truth Metadata
 FACE DETECTION use for train: no use for test: no	No Ground Truth available for this task.
 COPY DETECTION use for train: no use for test: no	No Ground Truth available for this task.
 SPEECH RECOGNITION use for train: no use for test: no	The Ground Truth source is hidden because this is not train data!

EBU MIM/SCAIE group



TOSCA^{MP}

- Call for participation
 - Content providers
 - Provide content
 - Provide ground truth data (very important)
 - Provide requirements on automatic information extraction tools
 - Automatic information extraction tool providers
 - Both academic & industry
 - Test tools on provided content and provide valuable feedback to EBU
 - Provide tools such that content owners can test the tools
- You are welcome to participate
 - Registration on <http://ebu-scaie.lab.vrt.be/mammie>
 - EBU-SCAIE members welcome your research results and feedback
 - EBU group website: <http://tech.ebu.ch/groups/pscaie>

Some cases for AIE



TOSCA-MP

- Legacy archive
 - Preservation: content is digitized because physical carriers degrade
 - Currently: speed-up of digitization process
 - At the cost of generating less manually annotated metadata
 - Big opportunity for AIE to unlock the archive faster (help archivists to annotate more and faster in a (semi-)supervised way
 - E.g. Concept detection
 - E.g. Automatic detection of location
- Broadcast news
 - Metadata is only added after broadcast date
 - AIE: opportunity: metadata upon ingest
 - E.g. Speech recognition on interviews
 - Face recognition spotting people appearing
 - Allowing for faster reuse of the content

Info



TOSCA^{MP}

TOSCA-MP

Survey on production and archiving tasks

<http://tosca-mp.eu/tasksurvey>

Peter Schallauer

peter.schallauer@joanneum.at

EBU SCAIE

Content collection & RfT

<http://tech.ebu.ch/groups/pscaie>

Mike Matton

mike.matton@vrt.be

<http://www.tosca-mp.eu>

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>).

