Use Case Description Template

Use Case 0 - Definition		
Name:	*Benchmarking of Automated Metadata Extraction, Linking and Search*	**Current Version:** *1.1*
Code:	*BAMELS*	
Summary:	*Services for automated metadata extraction (e.g., automatic speech recognition), linking of metadata (e.g., associating named entities with authority files and between media items) and advanced search services (e.g., using semantic technologies) are becoming increasingly important in media production processes. However, assessing their performance for supporting a particular task in the process is important for leveraging the potential of these technologies. This use case complements the AIEMPro^2 and search use cases by describing benchmarking services that can be used to assess and optimise these services and their configuration.*	
Description:	*The performance of automatic metadata extraction and advanced search services may strongly vary depending on content properties, such as genre, resolution, spoken language etc. Thus benchmarking (i.e., checking the quality of metadata/results generated against a validated reference) such a service for a new type of content in order to (i) assess whether it is applicable, (ii) select an appropriate configuration and (iii) build a tailored process for a certain type of content is relevant for using such services in a practical workflow.* *A practical workflow might include a range of metadata extraction and search services from different vendors, and even more than one service instance for one type of operation, as they might be complementary in terms of different content types. Thus a standardised, service oriented approach to benchmarking enables efficient and integrated evaluation of metadata extraction and search services.* *A benchmarking process for a metadata extraction or search service (called "service to be evaluated" in the following) produces evaluation results and possibly updated parameters based on metadata generated by this service and reference metadata ("ground truth"). In more detail, the process can be described as follows:* - *A set of metadata documents or search results for a defined set of media items, generated by the service to be evaluated, needs to be checked. In addition, the configuration used to obtain the result (parameters, queries, etc.) is available.* - *For (a subset of) these media items, ground truth, i.e., manually annotated/validated metadata, relevant search results, etc., are available.* - *Additionally, data for weighting the importance of specific aspects of the result or types of errors might be given (e.g., penalty for missing a named entity vs. missing a verb in the ASR transcript, cost of reporting an irrelevant vs. missing a relevant result in copy detection or search).*	

	- *Based on these inputs, the benchmarking service produces evaluation results in terms of one or more metrics implemented by the benchmarking service, represented as metadata documents. Possibly, the service also outputs visualisations of these results.* - *Optionally, the benchmarking service might output a modified/optimised set of parameters for the service being evaluated. This typically requires either a dependency between the benchmarking service and the service being evaluated or a standardised set of parameters for a certain type of metadata extraction/search task.* - *If the service being evaluated performs analysis on collection level, the same considerations about collections as described in the AIMEPro[2] use case apply.*
Initiating Actor:	*Archive Administrators, Automation*
Supporting Actors:	*Application Monitoring Staff, Software Specialists, R&D Staff*
Inputs:	*Content flows* - *one or more XML documents describing results of automatic metadata extraction, linking or search services for a set of media items, conforming to relevant metadata standards (e.g, ISO/IEC MPEG-7 Audiovisual Description Profile, EBU Core)* - *one or more XML documents carrying ground truth data, conforming to relevant metadata standards (e.g, ISO/IEC MPEG-7 Audiovisual Description Profile, EBU Core)* - *one or more XML documents carrying configuration information and parameters of these services* - *zero or more XML documents containing weighting information for assessing the results* - *optionally a reference to the media files on which the results are based* <hr> *Information flows* - *a set of or a reference to a collection of automatically generated metadata documents or result sets* - *a set of or a reference to a collection of ground truth documents* - *a set of configuration parameters of the services being evaluated* - *a set of parameters for the benchmarking service* - *optionally, a reference to a set of multimedia content items stored in a media file already ingested in the system or to be ingested contextually* <hr> *Control flows*

	- *Synchronous status monitoring at task instance level, including exceptions, management of retries, fallouts, termination of tasks and task model instances* - *Dynamic prioritisation of task instances* - *Basic access and browsing of produced metadata (e.g., HTTP access to XML metadata documents being produced, access to plots of result statistics)*
Outputs:	*Content Outputs* - *A set of (XML) documents containing the results of the benchmarking task in terms of appropriate metrics* - *A set of textual files containing detailed logs of each individual task execution* - *Optionally, a set of modified/optimised configuration parameters for the service being evaluated*
	Information Outputs - *A (reference to) a set of benchmarking results* - *A (reference to) a (portion of) logging information where relevant* - *Optionally, a (reference to a) set of configuration parameters*
Pre-conditions:	*An automated metadata extraction, linking and search service has been performed on a set of media items in the system and the results are available in the system as XML documents.* *Manually generated or automatically generated and manually validated ground truth is available on part of the content metadata extraction has been performed for.*
Post-conditions:	*Absence of severe or unmanaged exceptions, actual execution of tasks, presence of output metadata instances, syntactic and semantic conformance of the metadata instances to the intended schemas.*
Non-functional requirements:	*Execution and memory/storage resource allocation should be pre-estimated and communicated in some form to the caller in advance. Estimated completion time should be provided. Progress reporting and error notification should be supported by the service.*

Default flow:

1. *The process initiator selects the set of automated processing results, the set of ground truth documents and configuration parameters for processing by a benchmarking service*

2. *The process initiator asks for the execution of a task instance t on the collection C or on the individual item I*

3. *The system estimates needed resources, and completion time for the execution of t and*

communicates these data to the process initiator.

4. *If the initiator agrees on the provided estimation, the task instance is internally scheduled for activation*

5. *t gets activated, and the process initiator is notified of the activation*

6. *The initiator can optionally decide to postpone the actual start of execution or to manage prioritisation of the pool of activated task instances*

7. *t gets executed, and the process initiator is notified of the start of execution*

8. *The initiator can monitor the execution of t and browse the available benchmarking results as soon as they are completed*

9. *when the execution is complete, the system notifies the initiator and communicates coordinates of the results. These may include information for downloading files from network repositories or data from databases.*

10. *the initiator accesses the results, optionally verifies and updates proposed configuration parameters and feeds back the updates to the related automated metadata extraction/search component(s)*

Exception Handling:

- *if resource estimation fails system-level predefined boundaries (e.g., maximum execution time, maximum allocated resources per task model instance), the initiator must be notified immediately*

- *if the initiator rejects the resource and completion time estimation, there should be a negotiation phase*

- *if part of the internal transfers for ensuring execution-near-data fail, the system should re-estimate execution times and notify these back to the initiator.*

- *If any task-level unmanaged exception occurs during the execution of the task instance, the initiator must be immediately notified. The initiator can decide to retry the task, or to terminate its execution*

- *If the results are not accessible, the initiator must be able to notify the event to the system*

Optional Actions:

None relevant at this phase

Questions:

None relevant at this phase

Related Documentation:

[1] SMPTE 377M-2004 "Material Exchange Format"

[2] ISO / IEC Audiovisual Description Profile - http://mpeg.chiariglione.org/technologies/mpeg-7/mp07-avdp/index.htm

[3] EBU Tech 3293 - EBU Core Metadata Set			
Use Case History:			
Version	Date	Author(s)	Changes
1.0	19th December 2011	Werner Bailer (JRS) on behalf of TOSCA-MP project	
1.1	20th December 2011	Werner Bailer (JRS) on behalf of TOSCA-MP project	Revision based on feedback from TOSCA-MP consortium

Figure 1: OASIS REFERENCE MODEL APPLIED TO FIMS