Enterprise Multimedia Integration and Search

José-Manuel López-Cobo[1] and Katharina Siorpaes[1,2]

[1] playence, Austria, [2] STI Innsbruck, University of Innsbruck, Austria
{ozelin.lopez, katharina.siorpaes}@playence.com

1   Motivation

"The next generation Web should not be based on the false assumption that text is pre-dominant (…). The Web is a multimedia environment, which makes for complex semantics." [1]

With increasing bandwidth, cheaper storage of data, and improved hardware, multimedia content is gaining importance. This does not only show in Web portals but also in any other content-intensive environment in many verticals. As rich medium, video can transport and conserve more information than text ever could. This new type of content also creates new issues with respect to search, integration, management and preservation: up until now, heterogeneous data formats in text-based resources were a major challenge, now integration of non-textual contents has to be tackled. Naturally, a rich media asset is related to information and data residing in various information systems. While text analysis has been researched since years and mature solutions for tackling text annotation exist [2], the semantic annotation of multimedia is still a hard problem. This cannot only be traced back to the problems of image and audio analysis, but also to the fact that the combination of both can lead to entirely new high-level semantics. The automatic analysis of multimedia is only feasible to a very limited extent: still, human contribution for interpreting meaning to create metadata for rich media is required. In this paper, we describe playence Media, a system that aims at the semi-automatic annotation, search, and integration of multimedia and textual resources across information systems, taking semantics beyond text. We first describe the system and then discuss future challenges of this work.

2   Multimedia holistic view in the enterprise

When dealing with multimedia assets, current information systems can only rely on a weak annotation process due to the fact that only a limited set of low-level metadata can be extracted: file title, resolution, size, length, and other technical properties. In the best case, in specialized domains like media production and consumption, annotation or tagging is done manually, using a shared vocabulary and/or thesaurus.

However, the level of integration of these multimedia assets with the rest of the organization's information systems is limited, due to the cost and the lack of a holistic semantic model that could facilitate integration.

In an organization, maintaining different processes for each data format is not affordable. Therefore, we envision a holistic view on multimedia involving five information-related processes (Figure 1). The most important feature of this conceptual architecture is that is it based on semantic technologies, enabling scalable data integration [3]. playence Media relies on ontologies, providing shared vocabularies and
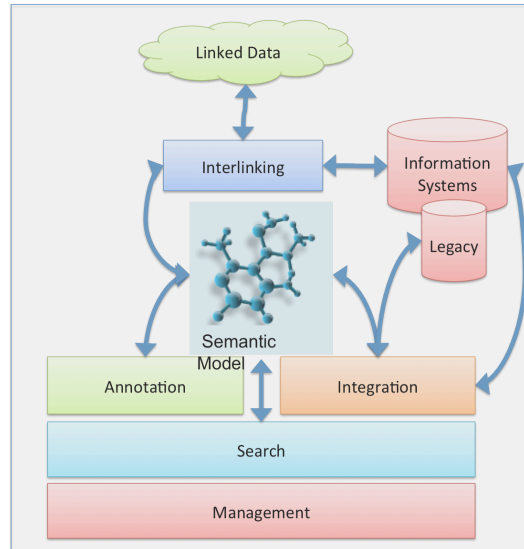

Figure 1 - Multimedia holistic view

terminologies. These models support all information-intensive processes: annotation, interlinking, integration and search.

2.1 Annotation process

The annotation process in playence Media can be done automatically by the system or manually. The accuracy and precision of the automatic process depends on the type of content.

In the case of text, the system is able to annotate content using domain ontologies, pinpointing the occurrence of a concept or an instance in text. The annotation process can be done using several ontologies, thus empowering users to have more than one point of view (e.g. sales department and engineering department have differing foci and priorities on the
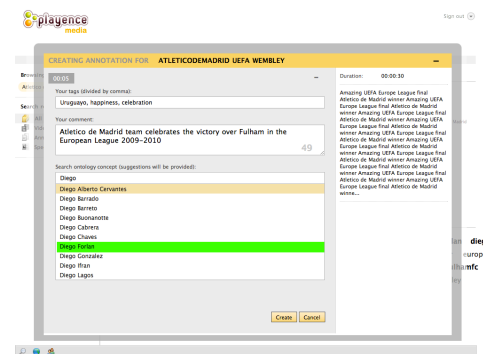

Figure 2 - playence Media annotation process

minutes of the same meeting).

In the case of video or audio, automatic analysis has less accuracy and can only contribute to the annotation partially. For content containing speech, techniques of ASR (Automatic Speech Recognition) can be applied, obtaining 60-80% of accuracy depending on the technical conditions of the audio channel, enabling speaker diarization [4] (e.g "who said what and when"). For video or image analysis, high-level features can be obtained, like face detection, detection of objects (or persons, or whatever other recognizable object), daylight classification and other basic features.

In those cases, collaborative manual contribution is still needed when addressing media annotation. playence Media supports manual annotation in three ways: informal tags, free text, and formal concepts and instances. Concepts and instances can be chosen from a drop-down list following predictive search or can be selected from a tree structure.

2.2   Integration

Annotation will produce a corpus of documents – regardless the type of content - annotated with the same set of domain ontologies. Thus, integration is facilitated because content residing in all information systems (documents, multimedia assets and previous legacy systems and databases) can be accessed using the same queries. Data can easily be located, mashed-up and displayed regardless its original source.

In playence Media, we approach integration from a holistic point of view, allowing users to find and relate assets with the same set of tools (domain ontologies). A user of playence Media trying to find information about a specific project (e.g. in a company where meetings are video recorded and annotated through playence Media) will be able to locate not only project meeting minutes (in text and video), but also other videos related with the project (similar topics, similar people), memos and related documents as well as previous meetings where the project was discussed.

2.3   Search

playence Media performs semantic search, using annotations and applying faceted search and semantic navigation to narrow the set of results (Figure 3). These techniques are complementary, allowing hybrid search and filtering through concepts and tags. This empowers the user to find the specific asset she is looking for regardless the technique.

When searching, playence Media makes use of Natural Language Processing techniques like lemmatization or spell check. Semantic features are used in query expansion, like synonym expansion through SKOS, or generalization-specialization expansion, using the "is-a" relationship and instances from concepts involved, or using more complex relations in query expansion. Thus, precision is enhanced as relevant assets (e.g. those in which concepts and instances from the query and those expanded appear) will be presented in the first results.

Figure 3 - playence Media structure

3   Discussion and conclusion

The challenges associated with multimedia integration in the enterprise are manifold: Videos are dynamic and often only one portion of the video might be relevant. Search must point to exactly these pieces of information. Additionally, extracting text from speech as one source for annotation, the quality of the transcription is a big concern. The creation and evolution of the domain vocabulary or ontology underlying the semantic annotations and search must be created so that it reflects the vocabulary relevant in the organization. Additionally, evolution and consistency of annotations are issues that have to be addressed. As multimedia analysis is not advanced, manual input is required to a certain extent. Thereby, unobtrusive, workflow-integrated and collaborative annotation and validation must be supported. Creating high-quality links to the relevant documents or even portions in documents is a must: media assets cannot be treated as isolated information islands. Making use linked data for multimedia annotation involves questions like selecting useful data sets and nodes as well as ensuring the quality of these annotations. Interlinking of contents - regardless of their type, whether they are rich media or textual contents - is a crucial requirement for efficient and comprehensive knowledge management.

References

1. Tim Berners-Lee et al. "A Framework for Web Science" in: Foundations and Trends in Web Science, Vol. 1, No 1, pp. 1-130, 2006.
2. Lawrence Reeve and Hyoil Han "Survey of semantic annotation platforms" in: SAC '05: Proceedings of the 2005 ACM symposium on Applied computing, ACM, 2005.
3. Tom Heath and Enrico Motta. "Ease of interaction plus ease of integration: Combining Web2.0 and the Semantic Web in a reviewing site". Web Semantics: Science, Services and Agents on the World Wide Web, Volume 6, Issue 1, February 2008, Pages 76-83
4. G. Friedland, O. Vinyals, Y. Huang, C. Müller: "Prosodic and other Long-Term Features for Speaker Diarization", IEEE Transactions on Audio, Speech, and Language Processing, Vol 17, No 5, pp 985--993, July 2009.