Towards a Semantic Turn in Rich-Media Analysis

Tobias Bürger; Georg Güntner

Salzburg Research Forschungsgesellschaft m.b.H.
A-5020 Salzburg, Jakob-Haringer-Strasse 5/III, Austria
e-mail: {tobias.buerger, georg.guentner}@salzburgresearch.at

Abstract

Typical application scenarios in the area of rich-media management, such as the continuous digitisation of the media production processes, the search and retrieval tasks in a growing amount of information stored in professional and semi-professional audio-visual archives, as well as the availability of easy-to-use hard- and software tools for the production of rich-media material in the consumer area, lead to an increasing demand for a meaning-based management of digital audio-visual assets. This "semantic turn" in rich-media analysis requires a semantic enrichment of content along the digital content life cycle and value chain: The semantic enrichment of content can be achieved manually (which is expensive) or automatically (which is error-prone). In particular, automatic semantic enrichment must be aware of the gap between meaning that is directly retrievable from the content and meaning that can be inferred within a given interpretative context. Each solution has its benefits and drawbacks. Our paper discusses the relevance of semantic analysis of rich-media in certain application scenarios, compares two possible approaches, a semi-automatic and an automatic approach, and presents a case study for an automatic solution. Following the observations of the case study, we come up with recommendations for the improvement of the semantic enrichment by an manual annotation step.

Keywords: semantic web; multimedia content management; semantic indexing

1 Introduction

As a motivation for the application of semantic technologies in the area of rich-media analysis we want to highlight the following application scenarios: Firstly, the continued digitisation of the media production process at professional content providers and content distributors (e.g. broadcasters, telecommunication companies) not only leads to an exponential growth of highly unstructured digital material, but also to an increased demand for a reliable classification of audio-visual material along all stages of the digital content value chain. Secondly, in the consumer area and the semi-professional area (e.g. corporate media archives, or small and medium sized audio-visual archives) the easy-to-use production tools lead to an unmanageable amount of audio-visual material, that rather later than sooner has to be managed and indexed in some way, whereas the meaning of the digital essences often is locked in the raw content. Thirdly, even if basic metadata is available to describe the content and its meaning, user-centred applications are increasingly demanding the utilisation of the benefits of the true semantic search approach, i.e. inference and reasoning, narrowing down and widening the search by using some kind of formal knowledge representation.

To exemplify the above scenarios, imagine a user who wants to find out recordings of performances of sacred music by Wolfgang Amadeus Mozart in and around the city of Salzburg during the Salzburg Festival 2005. This query is full of hidden semantics (see also figure 1 for a schematic presentation of this query and the associated knowledge model): Location based semantics (what does "in and around the city of Salzburg" mean?), time based semantics (when was "Salzburg Festival 2005"?), factual semantics: e.g. which works are considered to be "sacred works"; which musical forms are known to be "sacred works" in general (e.g. a choral, a mass); which particular works by Wolfgang Amadeus Mozart are sacred works?

The importance of the knowledge related with the query is, that nothing of it has to be encoded in the media essences or their description: all this knowledge can be modelled, described and used without any particular relation with the digital essences.

Our research group is currently investigating different approaches and methods for the combination of media content with semantic annotations and for the usage of pre-existing knowledge (i.e. the "context") for inferring further knowledge about the content automatically. We have faced the question of annotation from low-level content analysis recently, in the national research project Smart Content Factory (SCF) and we are going to address the issue of merging new content with existing knowledge in the IST project LIVE (see section 2 and 3). In another project, Smart Content Factory, we tried to automatically derive the semantics of TV news clips in order to make them browse- and searchable. To do this, we combined information extracted from raw multimedia content with domain knowledge about multimedia data.

In LIVE which deals with broadcasting of media events by integrating different videos streams with background information about these media events, we investigate how domain knowledge and background information can be efficiently combined to deduce further knowledge from broadcast live video-streams.

As we have experienced in the Smart Content Factory [1], semantic descriptions of content can enhance fast and easy navigation through audio-visual repositories. Semantics - i.e. the interpretation of the content - is important to make content machine-processable and to enable the definition of tasks in workflow-environments for knowledge workers in the content industries. Some of the recent research projects in the area of semantic (or symbolic) video annotation try to derive the semantics from the low level features of the audiovisual material or from other available basic metadata, e.g. by audio-classification of classification of camera movement. Some of the projects aim at highly automated indexing using the results of automatic speech recognition however error-prone they may be. Most of these approaches are - as also pointed out in [2] - not capable to derive the semantics of multimedia content because in many cases the results of the analysis cannot be related to the media context [3]. For humans the construction of meaning is an act of interpretation that has much more to do with pre-existing knowledge (the "context") than with the recognition of low-level-features of the content. This situation is commonly referred to as the "semantic gap" [4].

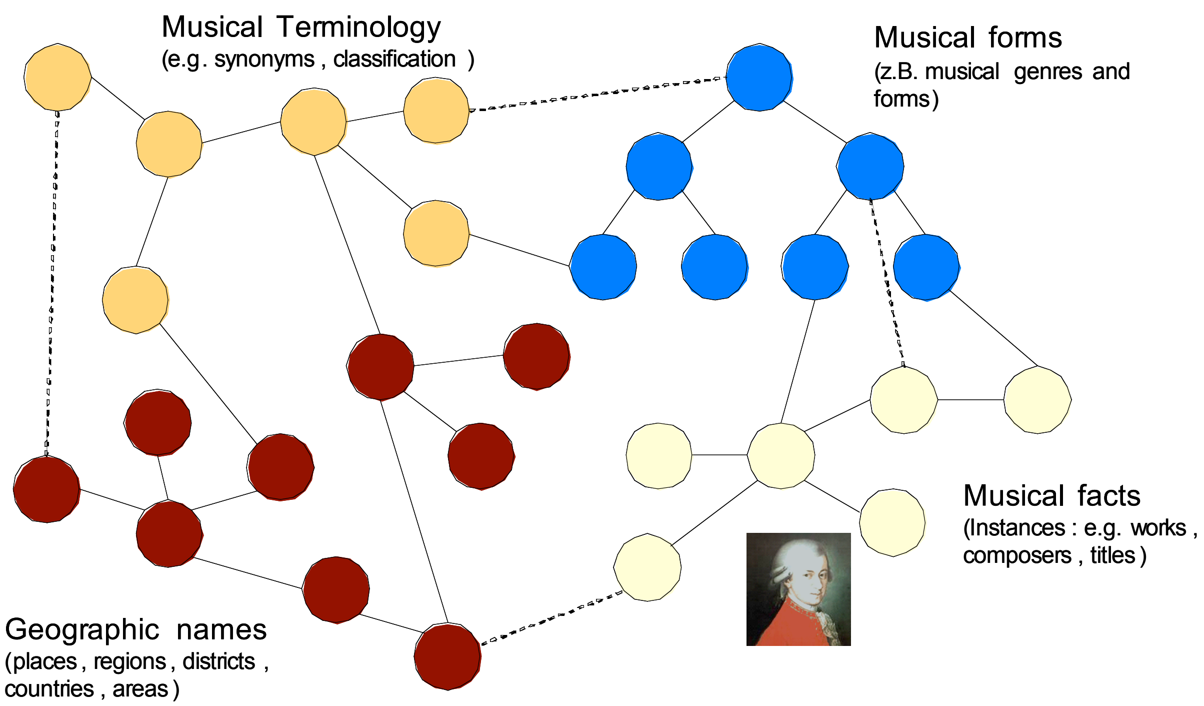

Figure 1: The knowledge-base for the semantic query about Mozart's sacred works

Two solution paths have emerged for this problem: The first one is to provide rich annotations created by humans as training data for the system to learn features of videos for future automatic content-based analysis. The second approach does not rely on training, but purely on analysis of the raw multimedia content. The training approach is not well suited for scenarios in which a great amount of content has to be annotated before any training and automation can be done or in which the application domain is very broad. The second approach usually only works well in settings where the relevant concepts can easily be recognized. However, most content based services demand richer semantics. As pointed out in section 4, popular examples on the Web show that there are currently many service-based platforms that make use of their users' knowledge to understand the meaning of multimedia content.

In our paper we concentrate on different approaches to close this semantic gap and provide insight into two solution paths, one automatic and one semi-automatic and demonstrate a case study on a prototypical solution for the "semantic augmentation" in the area of audio-visual archives.

2 Methodology: Automatic Vs. Semi-Automatic Semantic Rich-Media Analysis

In general, metadata generation systems can be classified in manual-, semi-automatic- and automatic annotation tools: The aim of automatic and semi-automatic tools for the analysis of rich-media content is to extract as much useful information from the raw media file as possible. Manual annotation tools aim to provide support for users to add metadata by hand.

Currently many systems try to expose the semantics of multimedia data by adding metadata to it. However, most of them do not derive these annotations just from the low-level features detected in the raw media data, but instead for example either analyze the different modalities of a video, analyse the usage context of the media or rely on human annotation/interpretation to derive higher-level semantic features from multimedia data.
In this section we want to introduce different approaches that we applied in two research projects making use of semantic technologies for rich-media analysis.

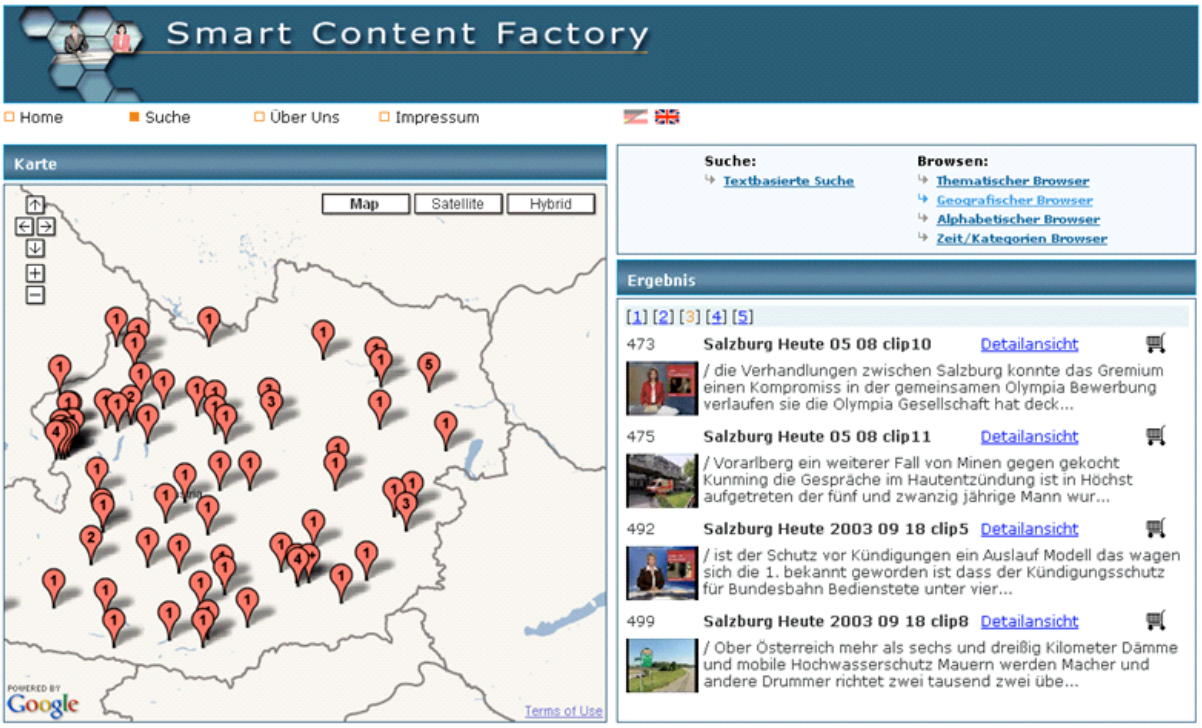

Figure 2: Location based navigation in the Smart Content Factory

Firstly, a project called "Smart Content Factory" (2003-2006) was driven by the idea to develop a system infrastructure for the knowledge-based search, retrieval and navigation in audio-visual archives of news clips. The approach was highly relying on an automatic feature extraction, mainly the speech-to-text transcription during the first phase. In a late phase this automatically extracted features were supported by additional reliable sources available in the digital production process at the Austrian Broadcasting Corporation (ORF). Section 3.1 describes the results and major findings of the selected approach with respect to the semantic indexing.
Secondly, an ongoing integrated project called "LIVE staging of media events" (started n 2006) is driven by an approach to combine the methods of automatic and semi-automatic detection, extraction and annotation of content with a knowledge-base under the control of a semantic based media framework. Moreover the framework propagates knowledge and contextual information to s recommender system which thus to some degree becomes aware of the meaning of the media. Section 3.2 describes the current state of this approach.

3 Results: Towards A Semi-Automatic Reliable Semantic Analysis Framework

This section describes the results of two research projects with respect to their selected approach for the semantic analysis.

3.1 Smart Content Factory – An Automatic Approach to Semantic Rich-media Analysis

In a research project called Smart Content Factory [5], we developed a prototype of a system infrastructure for the automatic knowledge-based refinement of audio-visual content repositories based on state-of-the-art digital asset management systems. In the project an automatic approach was used which was based on a two step indexing pipeline:

> In the first step a primary index is created by methods provided by state-of-the-art media analysing tools (i.e. Virage Video Logger™ and the Smart Encoding™ process). Video clips are passed to the Factory in MPEG-1 format. A polling mechanism informs the indexing and contextualisation components about newly available video clips and triggers the indexing process. The primary indexing results in the creation of key frames, the automatic detection of scenes, the transformation of speech to text, the recognition of speakers, etc.
>
> Subsequently a semantic indexing process is started which is based upon the results of the previous content based indexing and relies primarily on the speech to text transformation (i.e. "audiologging", for which different solutions have been tested in the course of the project). The dependency on the results of the audiologging is a weak point in the concept of the Smart Content Factory in so far as the subsequent semantic indexing builds on the results of a per se error-prone automatic extraction process. The semantic indexing is using the Lucene indexing framework [6] and the ontologies and thesauri forming the knowledge base of the Smart Content Factory which are accessible via "pluggable" RDF knowledge components described in the previous section. By applying and using knowledge models during the semantic indexing we create a set of "smart indices", allowing search, retrieval and reasoning along various dimensions of the information space.

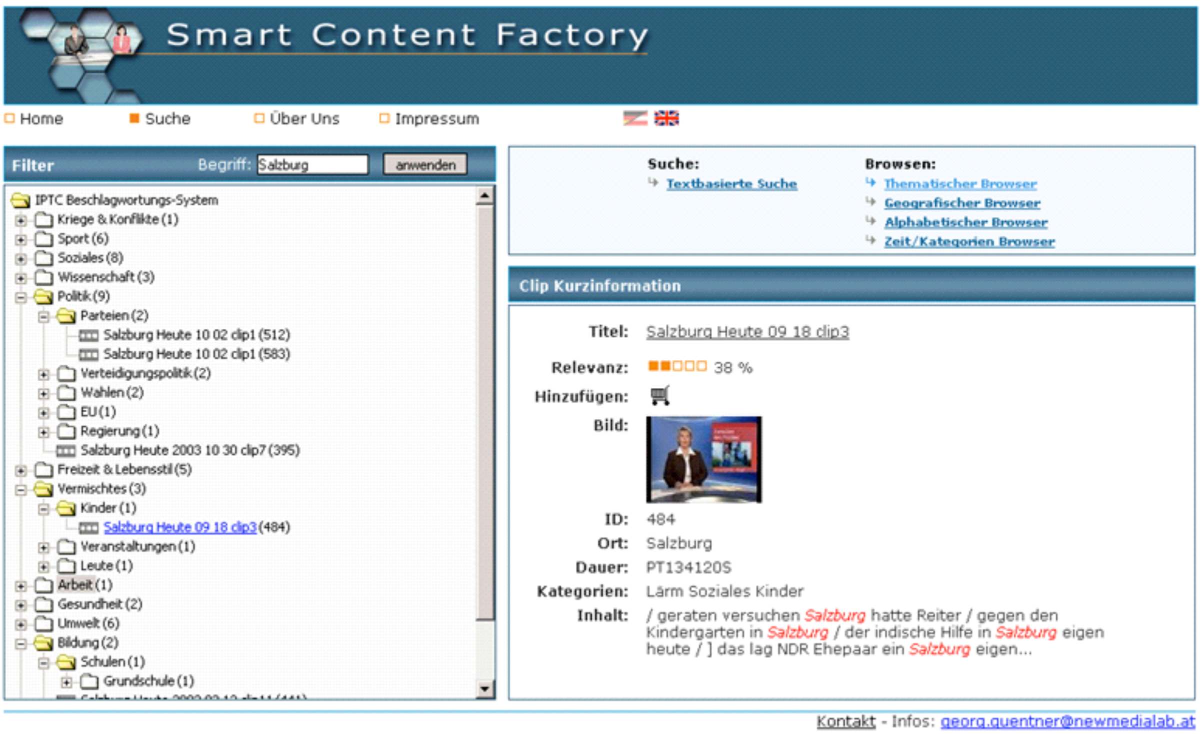

Figure 3: Category based navigation in the Smart Content Factory

One of the key issues of our semantic indexing framework of the Smart Content Factory was the use of an extensible set of formal knowledge models (accessible via the Jena RDF framework [7]):

(1) The first knowledge model ('locations') contains a thesaurus of geographic names. The thesaurus extends the properties of gazetteers by modelling the hierarchical relations between the geographic locations (e.g. 'village' is-part-of 'political district'). The geoname thesaurus is based on data structured according to the ADL Feature Type Thesaurus [8] and is represented in RDF, the gazetteer is structured according to the ADL Gazetteer Content Standard [9]. This model allows specialisation and generalisation of search queries by location concepts. The recognition of location names is further supported by an engine for the recognition of named entities. In the Smart Content Factory we used LingPipe [10] to resolve ambiguities:

Scanning texts for occurrences of common known Austrian place names like 'Wien' (Vienna) or 'Salzburg' is rather easy. But it is a lot harder to semantically distinguish an appearance of the name of the Austrian village 'Haus' from the German word for house (the building). Due to the lack of a bigger training set, the location name recognition's precision/recall is not very high, but it serves as a starting point for distinguishing false positives from true positives. A simple tf/idf-based ranking is used to determine the 'most significant' location which a video is related to. Tf/Idf means 'term frequency-inverted document frequency' [11].

(2) To identify thematic categories we used the IPTC thesaurus (International Press and Telecommunications Council, [12]) which defines a hierarchical structure of thematic news categories (e.g. sports, policy, economy). For our purpose the IPTC thesaurus is represented in RDF according to SKOS Core 1.0, an RDF schema defined by W3C for the description of thesauri and similar types of knowledge models [13]. Similar to the location name identification process, this process identifies terms from a controlled vocabulary provided by the IPTC thesaurus.

(3) A web-based synonym service for (German) words was integrated into the semantic indexing process. The service was created and is maintained by the University of Leipzig ('Deutscher Wortschatz' [14]). The Web service interface is based on the SOAP protocol. By means of this service the index is enhanced with synonyms of non-stop words. All models are either integrated via a Web service interface or stored in a database and are accessed via the Jena RDF framework [7], which also provides a powerful inferencing mechanism (e.g. traversing of hierarchical relations).

In the Smart Content Factory, one of the objectives for the introduction of the knowledge-based index was the enhancement of search and retrieval and navigation support [1].

The main benefits of this approach were the little need for human intervention in the process as the annotations were totally generated automatically. Another benefit was that the approach was extensible as other thesauri/knowledge models could be easily plugged in to recognise for example events or dates. One drawback however was the amount of false positives locations that were recognised, which was mainly due to the bad results of the text to speech engine. Another drawback was the amount of time needed to index a video, which doesn't allow real-time indexing of the video.

Figure 2 shows the application of the location based semantics for the map based search nd navigation in the audio-visual archive. Figure 3 exemplifies the category based navigation, using the IPTC thematic thesaurus which is represented in RDF as described above. Both navigation paradigms were highly appreciated by the test user group during a user evaluation in 2006, whereas other forms, e.g. the hyperbolic tree navigation paradigm were ranked low in the users' interest profile.

3.2 LIVE – Real-Time Semi-Automatic Annotation of Videos with the Intelligent Media Framework

The integrated project "LIVE Staging of Media Events" (LIVE; FP6-27312, [15]) aims at the creation of novel intelligent content production methods and tools for interactive digital broadcasters to stage live media events in the area of sports, such as the 2008 Olympic Games. In the terminology of the project, "staging live media events" is a notion for the creation of a non-linear multi-stream video show in real-time, which changes due to the interests of the consumer (end user). From a technical viewpoint, this requires a transformation of raw audiovisual content into "Intelligent Media Assets". LIVE will develop a knowledge kit and a toolkit for an intelligent live content production process including dynamic human annotation and automated real-time annotation. As a consequence novel iTV video formats for live events will evolve.

In the LIVE project we applied the lessons learnt from the automatic approach of the Smart Content Factory to overcome the weaknesses of automatic metadata extraction and extended the system architecture to meet the requirements of real-time semantic indexing. In the first phase of the project we started to design an Intelligent Media Framework that is taking into account the requirements of real-time video indexing to combine several automatic and manual annotation steps. The Intelligent Media Framework thereby integrates the following components of the LIVE production support system:

(1) The Intelligent Media Asset Information System (IMAIS) providing access to services for the storage of media, knowledge models and metadata relevant for the live staging process and providing services for the creation and management and delivery of intelligent media assets. This will be the central component of the Intelligent Media Framework and will semantically enrich incoming metadata streams;
(2) The Recommender System, giving content recommendations to the user based on the user's personal profile and on previous user feedback;
(3) The Metadata Generation System, dealing with the detection, extraction and annotation of knowledge from audiovisual material;
(4) The Video Conducting System, dealing with the real-time staging of a live event.

The components of the indexing pipeline in LIVE are shown in figure 4: the Automatic Analysis Application, the Human Annotation Tool and the Intelligent Media Framework. The role of the Intelligent media Framework is to accept and handle partial information about particular media items, to add semantic information to the items and to infer and attach contextual knowledge to the items that is probably related to the event that is staged. It furthermore provides knowledge services that offer controlled vocabularies related to the current context of a stream to guarantee the unambiguousness of the terms used.

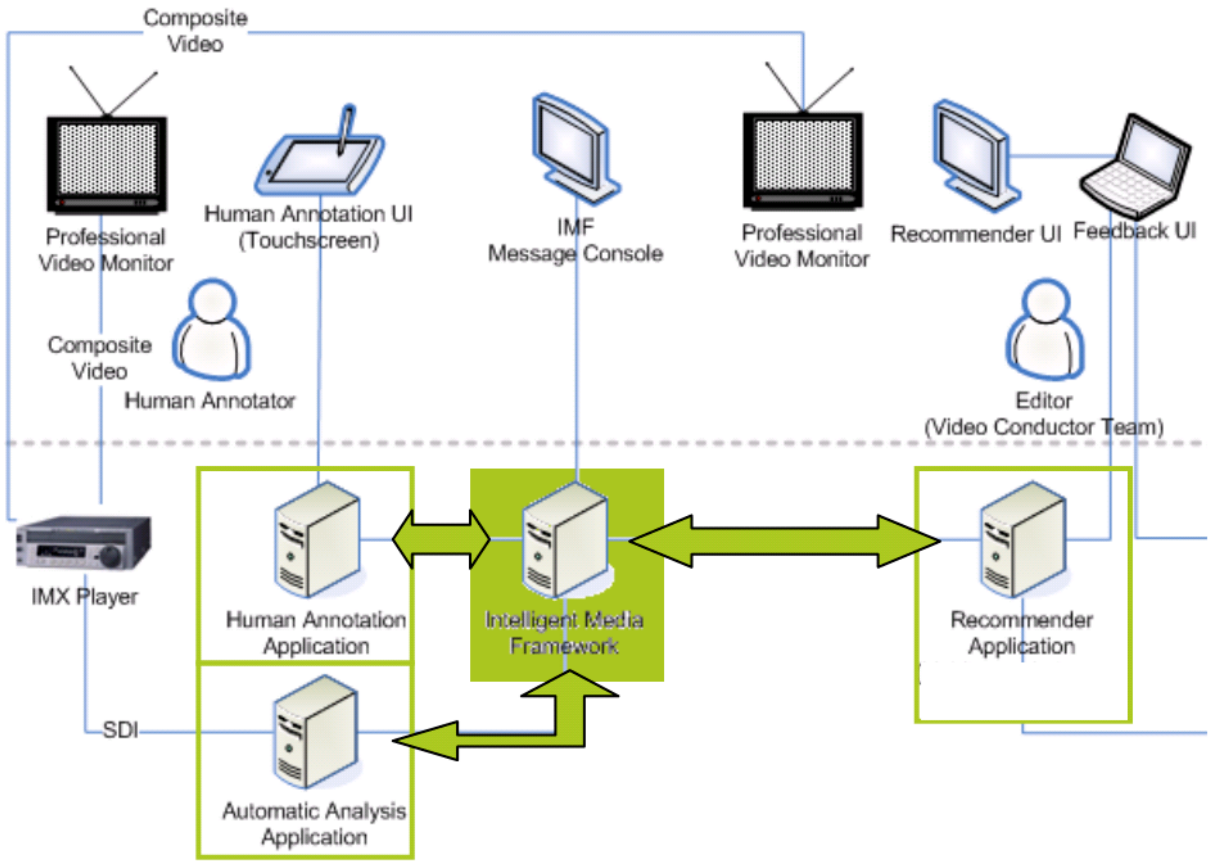

Figure 4: Category based navigation in the Smart Content Factory

The semantic enrichment process in LIVE is twofold: The Automatic Analysis Application detects close-ups, shots, faces, camera-motion, colour schemes, scenes and artists. This information is enriched in (1) a manual step done by the human annotator through the human annotation tool (2) in the Intelligent Media Framework that has knowledge about the context of the analysed media item. In (1) terms from the controlled vocabulary are assigned to the low-level information that was extracted in the basic analysis step. In (2) these terms are used to attach more semantic information of the current action or event to the media items that is possibly inferred by the current event schedule or other particular information that was detected in the course of this event.

This semantic indexing process is more reliable than the approach from the Smart Content Factory, because it is neither based on error-prone text transcripts nor totally relies on automatic analysis tools. One key enabler of this semantic indexing step is the use of existing information systems at the broadcaster's side that have knowledge about the staged event, the participants and so on. The most important step in this process, however, is the human annotation that is later inferred by the Intelligent Media Framework in a reliable way. This allows us to act in real-time (with a maximum delay of approx. 20ms) and provides high-level metadata helping to bridge the gap between the raw audio-visual essences and their intended meaning.

Figure 4 shows parts of a prototypical demonstration setup shown during a LIVE review meeting in Vienna in March 2007. Both, a human annotation tool and the automatic annotation system, are using a semantic aware middleware, the Intelligent Media Framework, which provides context information and the controlled vocabulary for the annotation process and propagates the detected "meaning" to a recommender system for the professional user (editor, video conductor). The video conductor team decides which live streams and switching possibilities are offered to the consumers.

4 Discussion

In this section we list existing approaches that try to extract knowledge from rich-media items and try to relate these approaches to the LIVE approach. Indexing and metadata generation is a common task to media analysis systems and there are sound algorithms and methods in the area of computer vision, pattern recognition, natural language processing and signal processing that can be applied, most of them applicable for extracting low-level features from the essences. Currently many systems try to expose the semantics of multimedia data by adding annotations to it. However, the common trait of all the following examples is that they do not derive these annotations just from the low-level features detected in the raw media data, but instead for example either analyse the different modalities of a video, analyse the usage context of the media or rely on human annotation/interpretation to derive higher-level semantic features from multimedia data.

Recent research efforts try to combine automatically derived features like speech-to-text transcripts with background knowledge and related information on the Web: an application called Rich News [16] deals with automatic annotation and extraction of semantics from news videos: In a first step, the system extracts text from speech and then it tries to extract the most important topics from that. With these extracted topics, the system starts a Google search to find news stories on the web that cover the same topic(s). Google did exactly the same in a research project [17] to enhance American TV news with background information from the Internet: They extract important topics from the subtitle channel that is broadcasted with every news show and give consumers the possibility to get background information about these news items by displaying links to related web pages. MediaMill [18], a system that was developed at the University of Amsterdam, uses all modalities of a video to derive the semantics of it, which is especially important when the visual content is not reflected in the associated text like close captions or speech-to-text transcripts. Besides these examples from the research community there are also some major trends in industry: Some of the established industry-strength systems come from the classical document management area with its sophisticated full text indexing and information retrieval methods. The industrial players are now extending these methods to audiovisual content (e.g. Convera's RetrievalWare [19] or Autonomy's IDOL Server [20]). Other vendors come from the digital asset management sector and extend their systems to full text indexing and knowledge management methods (e.g. Virage's VS Archive [21]).
These systems rely heavily on metadata and on full-text indexation which in turn, is based on speech-to-text extraction, but also make use of statistical methods to classify content according to taxonomies or thesauri. The leading search engines (e.g. Google or Yahoo) are currently extending their search features to audio-visual media, but they are still to a high degree relying on metadata and on text transcripts provided along with the media assets. Besides that, Google and Yahoo both released versions of their search engines to discover contents in videos: Google for example has been indexing news stories from an U.S. broad-caster. There the search index is based on the closed caption provided along with news clips. They also scan the Web for videos and images: In that case - as also described in [22] - additional metadata are generated from adjacent information on the website (e.g. text blocks or the video file name).

However, there are also other services dealing with content which are already very popular among a large user community: Two of the most popular content-based services are Flickr [23] and Last.FM [24], both of which get their users to classify (tag) images or music files within these systems. These systems do not classify the content according to predefined categories, but instead, taxonomies evolve from the users' tags (folksonomies). These tags can be regarded as user-based knowledge about certain items that on these platforms, is used to filter or recommend content to other users. The main requirements in LIVE are (1) reliability of annotations as fast decisions have to be made based on them and (2) the real-time assignment of these annotations. The mentioned approaches are shortly discussed according to their usefulness in LIVE.

As mentioned above Web sites such as Flickr [23], or also Riya [25] recently began to apply algorithms for automatic extraction of content metadata (e.g. shapes, colour or texture features) and some of them already use high level pattern recognition technology such as face detection (e.g. Riya) However the metadata provided by them is not reliable enough. Other approaches that tend to provide reliable metadata information like MediaMill [18] or the commercial tools like Virage's VS Archive [21] perform the necessary analysis not fast and accurate enough. Manual annotation tools like Vannotea [26], Advene [27] or M-Ontomat-Annotizer [28], to name just a few existing annotation tools, are to complex to be utilizable in the LIVE real-time situation. LIVE or the IMF also goes one step beyond the research as it tries to include contextual information in the analysis as much as possible, however at the moment solely in the LIVE domain.

5 Outlook and Conclusion

In the Smart Content Factory we experienced that it on the hand is possible to automatically extract low-level features from video and augment them through the use of semantic technology but on the other hand this information is sometimes error-prone as it has to be based on incomplete or wrongly extracted features. In LIVE wrong or incomplete information could lead to wrong decisions in the live staging process which to some extent forces us to augment extracted information manually to reach a high degree of reliability. The use of controlled vocabularies and semantically enhanced metadata throughout the whole LIVE system introduces a common language that will lead to fast and reliable decisions during the staging process. The Intelligent Media Framework is responsible for the augmentation of the automatically extracted information that is additionally manually refined. Our next steps in the development of the IMF are the further addition of contextual information to the metadata sets of single media items, such as information related to the event or the athletes.

Acknowledgements

Smart Content Factory is one of the lead projects at Salzburg NewMediaLab (SNML [29]), the Austrian centre of excellence in the area of digital content engineering and new media technologies. Salzburg NewMediaLab is funded by the Austrian Federal Ministry of Economics and Labour and the Province of Salzburg. The project started in 2003 and ended in September 2006. It was coordinated by Salzburg Research: Further partners were ORF (the Austrian Broadcasting Corporation), X-Art ProDivision and Joanneum Research (all Austrian partners). LIVE (Live staging of media events, FP6-27312, [15]) is an integrated, multidisciplinary initiative that will contribute to the IST strategic objective "Semantic based knowledge and Content Systems" and "Exploring and bringing to maturity the intelligent vision". The project is funded by the European Commission within the 6th Framework Programme. The project started in 2006 with a duration of 45 months. It is coordinated by Fraunhofer IAIS. Further partners are Salzburg Research, the University of Ljubljana, the University of Bradford, the University of Applied Sciences in Cologne, the Academy of Media Arts Cologne, ORF, Pixelpark and ATOS Origin.

References

[1] BÜRGER, T.; GAMS, E.; GÜNTNER, G.: *Smart Content Factory - Assisting Search for Digital Objects by Generic Linking Concepts to Multimedia Content*. In: Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (HT '05), 2005.

[2] BLOEHDORN, S. et al.: *Semantic Annotation of Images and Videos for Multimedia Analysis*. In: Proceedings of the 2nd European Semantic Web Conference, ESWC 2005, Heraklion, Greece, May 2005.

[3] BÜRGER, T.; WESTENTHALER, R.: *Mind the gap - requirements for the combination of content and knowledge*. In: Proceedings of the first international conference on Semantics And digital Media Technology (SAMT), December 6-8, 2006, Athens, Greece.

[4] SMEULDERS, A. W. M. et al.: *Content-Based Image Retrieval at the End of the Early Years* In: IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 22 No. 12, December 2000.

[5] Smart Content Factory project web-site: http://scf.salzburgresearch.at/ – Last visited: 10.04.2007

[6] The Apache Jakarta Project: Lucene. From: http://jakarta.apache.org/lucene - Last visited: 10.04.2007

[7] Jena – A Semantic Web Framework for Java. From: http://jena.sourceforge.net/ - Last visited: 10.04.2007

[8] University of California, Santa Barbara: Alexandria Digital Library Feature Type Thesaurus, http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/ - Last visited: 12.10.2004

[9] University of California, Santa Barbara: *Guide to the Alexandria Digital Library Gazetteer Content Standard*, http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm - Last visited: 12.10.2004

[10] LingPipe: http://www.alias-i.com – Last visited: 10.04.2007

[11] Term frequency-inverted document frequency (tf/idf): http://en.wikipedia.org/wiki/Tfidf – Last visited: 10.04.2007

[12] International Press and Telecommunications Council. From: http://www.iptc.org/ - Last visited: 10.04.2007

[13] MILES, A. J. *SKOS Core - Guidelines for Migration* .http://www.w3.org/2001/sw/Europe/reports/thes/1.0/migrate/ - Last visited: 01.04.2007]

[14] University of Leipzig, Institute of Computer Sciences. Project "Deutscher Wortschatz" (German dictionary). From: http://wortschatz.informatik.uni-leipzig.de/ - Last visited: 28.04.2005

[15] LIVE – Live staging of media events; project web-site: http://www.ist-live.org – Last visited: 10.04.2007

[16] DOWMAN, M.; TABLIN, V.; URSU, C.; CUNNINGHAM, H.; POPOV, B.: *Semantically enhanced television news through web and video integration* In Proceedings of the Workshop on Multimedia and the Semantic Web at the European Semantic Web Conference (ESWC 2005), 2005.

[17] HENZINGER, M.; CHANG, B.-W.; MILCH, B.; BRIN, S.: *Query-Free News Search*. In Proc. of the 12th World Wide Web Conference, pp. 1-10, 2003.

[18] SNOEK, C. G. M. et al.: *MediaMill - Exploring News Video Archives based on Learned Semantics*. In: Proceedings of the ACM Multimedia Conference 2005, 2005

[19] Convera - http://www.convera.com/ - Last visited: 10.04.2007

[20] Autonomy - IDOL Server: http://www.autonomy.com/content/Products/IDOL/ – Last visited: 10.04.2007

[21] Virage: http://www.virage.com – Last visited: 10.04.2007

[22] FERGUS, R. ; PERONA, P.; ZISSERMAN, A.: *A Visual Category Filter for Google Images*. In: Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, 2004.

[23] Flickr: http://www.flickr.com – Last visited: 10.04.2007

[24] Last.FM: http://www.last.fm – Last visited: 10.04.2007

[25] Riya: http://www.riya.com/ – Last visited: 10.04.2007

[26] Vannotea: http://liris.cnrs.fr/advene/ – Last visited: 10.04.2007

[27] Muvino: http://vitooki.sourceforge.net/components/muvino/code/index.html – Last visited: 10.04.2007

[28] M-Ontomat-Annotizer : http://www.acemedia.org/aceMedia/results/software/m-ontomat-annotizer.html – Last visited: 10.04.2007

[29] Salzburg NewMediaLab (SNML), "Kompetenzzentrum für Neue Medien" (competence centre for new media technologies and digital content engineering): http://www.newmedialab.at/ – Last visited: 10.04.2007