Media Semantics

Werner Haas, Werner Bailer and Michael Hausenblas

JOANNEUM RESEARCH
Institute of Information Systems & Information Management
Steyrergasse 17, 8010 Graz, Austria

firstname.lastname@joanneum.at

Abstract

Our institute's scope is put on the integration of content-based and semantic technologies into multimedia applications. We therefore highlight the current state-of-the-art in dealing with the Semantic Gap and present our approach based on the experience gained from projects focusing on real-world media data.

Introduction

The key research in media semantics still is focused on how to bridge the Semantic Gap. Following [1], "The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation". We subscribe to this view, and discuss in the following section different approaches how to actually deal with this situation in applications for multimedia content retrieval and production.

So far our standard methodologies to describe and search specific content, for example an image, video or piece of music, are mostly utilizing "piggy-back" text technologies based working on metadata. Text and metadata may be manually entered, gained from optical character recognition (OCR) or from automatic speech recognition (ASR). Content Based Image Retrieval (CBIR) methods extract meaning directly from multimedia objects. While this is relatively straightforward for low level features like colour, texture, pitch or volume, it is extremely difficult to extract objects or genres, to name just a few real world concepts. At this point we have to resort to knowledge. Semantic Web technologies are offering a way to formalize the knowledge available and help us in describing – and later on in finding – our content in a much more user oriented way. Even more, single content objects knowing about their meaning will on the long run be able to combine themselves on the fly into meaningful sequences, according to domain needs, following established drama rules.

Approaches

Low-level feature based. Throughout the 1980s and 90s, low-level based approaches have dominated the work on content-based image retrieval (CBIR). Using the query-by-example (QBE) paradigm, similarity between multimedia documents is defined in terms of low-level features that can be directly derived from the multimedia data, such as colour, texture and shape in the visual domain, or pitch and frequency spectrum in the audio domain.

A huge number of feature descriptors have been developed throughout this period, along with efficient matching and indexing approaches, some of these feature descriptors have been standardized in the MPEG-7 standard [2]. The advantage of this approach is that the problem of interpreting the multimedia data is avoided, with the drawback that queries can only be formulated by presenting a signal representation of the query example.

Model-based. This class of approaches emerged from the application of computer vision and image understanding research to multimedia indexing and retrieval. Generally speaking, the concepts (objects, events, etc.) to be detected in the content are modelled and connected to their low-level feature representations by training classifiers using supervised learning approaches. If the domain of the multimedia content is known, these approaches yield satisfying results in practical applications (e.g. content-based description of sports video). However, in general applications the quality of the results depends crucially of the grounding of the concept in the low-level features. As can be seen from concept detection benchmarks (cf. TRECVID, http://www-nlpir.nist.gov/projects/trecvid), the performance for abstract concepts is very poor.

Semantic Web. Grounded on the Semantic Web vision (http://www.w3.org/2001/sw), the approach is to use RDF (http://www.w3.org/RDF) and Description Logic-based languages (cf. OWL http://www.w3.org/2004/OWL) to model audio-visual features and domain semantics. To overcome the problem that comes along with multimedia metadata standards, viz. the missing formal basis [3] and [4], this seems to be a good idea. This purist Semantic Web approach solves interoperability issues and allows for sound retrieval operations. Though it solves some of the problems, it introduces new ones: the lack of support for basic multimedia requirements as time-based descriptions, weak data typing, and scalability issues, just to mention a few.

Hybrid. Instead of either using a bottom-up (multimedia data based) or top-down (knowledge based) approach, hybrid approaches aim at integrating both worlds, as both can mutually benefit from each other. How to practically realise this integration is still an active research topic and different approaches have been proposed recently, differing in the metadata representations and technologies being used. As this approach brings together previously separated communities, it is a very promising, yet not mature one.

On the way to Media Semantics

Projects. Our institute focuses its research on the integration of content-based and semantic technologies into applications and workflows for the production, and distribution of digital content, as well as media understanding. We head for environments in the realm of professional movie and film productions, and likewise market research. To pursue these goals, the institute participates in national and European projects such as SALERO, Semedia, MediaCampaign, UAd, NM2, IP-RACINE, and K-Space. Whereas K-Space is an example from the foundational research, with NM2 and IP-RACINE we have two media production projects that elaborate on the question: "*How to use semantic technologies together with standard media metadata to produce content*". The three latter mentioned projects are discussed in detail below.

K-Space (http://www.k-space.eu) integrates leading European research teams to create a Network of Excellence in semantic inference for semi-automatic annotation and retrieval of multimedia content. The aim is to narrow the semantic gap by researching content analysis, annotation and description techniques, knowledge representation and extraction from multimedia content and complementary sources as well as multimedia mining and user relevance feedback.

IP-RACINE (http://www.ipracine.org) aims at improving digital cinema production from “scene to screen” by establishing a better essence and metadata workflow throughout the process. A key challenge is the heterogeneity of metadata types, formats, standards and tools involved. Today the final product of the production is mainly the audiovisual essence, while valuable metadata are lost at some stage in the workflow due to a lack of interoperability. The tools developed by our institute use semantic technologies to facilitate conversion between different metadata formats and standards and to automatically apply the editing decisions taken for the audiovisual essence to the metadata descriptions of the content.

The “New Media for a New Millennium" (NM2) project (http://www.ist-nm2.org) aims at developing tools for the media industry that enable the efficient production of non-linear, interactive broadband media. To provide non-linearity, NM2 productions are not final edited pieces of media, rather they consist of a pool of small media units to be recombined at run-time. Central units in NM2 are so called media items that refer to some multimedia essence (a video or audio clip, etc.), and provide a machine processable description of the essence to make its semantics explicit. Semantics of a media item is made explicit in two steps: Firstly, by attaching MPEG-7 description of the essence; secondly, through the manual annotation with logical entities defined in production specific ontologies. The major task lies in bridging the Semantic Gap, viz. to map media intrinsic information (captured within MPEG-7 descriptions) to logical entities (represented formally in a production’s ontology).

MediaCampaign's (http://www.media-campaign.eu) scope consists of discovering, inter-relating and navigating cross-media campaign knowledge and extensively automating the detection and tracking of media campaigns on television, Internet and in the press. For the pilot system developed within the project, the focus is on a concrete example for a media campaign: advertisement campaigns. The currently manual process used to acquire the media campaign data will be significantly accelerated by means of the MediaCampaign. The key research objectives are (i) creation of a knowledge model for semantic description of media campaigns in general, (iI) identification & tracking of new media campaigns in different media, and (iii) modeling of domain specific ontologies which relate media campaigns over different media and countries.

Standardization Activities. We participate in the **W3C** Multimedia Semantics Incubator Group (MMSEM-XG, http://www.w3.org/2005/Incubator/mmsem); there the mission is to "show how metadata interoperability can be achieved by using the Semantic Web technologies to integrate existing multimedia metadata standards." This is accomplished by a series of use cases that describe typical interoperability problems along with proposed solutions. One trailblazing use case that is in line with Web 2.0 issues is the *Collaborative Tagging* use case (http://www.w3.org/2005/Incubator/mmsem/wiki/Tagging_Use_Case).

In a recent keynote (http://tomgruber.org/writing/ontology-of-folksonomy.htm), Tom Gruber stated:

> *Ontologies are enabling technology for the Semantic Web. They are a means for people to state what they mean by formal terms used in data that they might generate or consume. Folksonomies are an emergent phenomenon of the social web. They are created as people associate terms with content that they generate or consume. Recently the two ideas have been put into opposition, as if they were right and left poles of a political spectrum. This piece is an attempt to shed some cool light on the subject, and to preview some new work that applies the two ideas together to enable an Internet ecology for folksonomies.*

This approach is a promising one that could help to bridge another gap: The one between emerging Web 2.0 and Folksonomy community on the one side, and the academic-driven Se-

mantic Web development on the other. Again we contribute by best practice in the realm of metadata deployment [5].

The scope of **MPEG** has been extended from only signal coding to multimedia metadata, processes and applications. The MPEG-7 standard [1] specifies the description of multimedia content, integrating content structure (e.g. shots of video, regions of image), low-level visual and audio features and high-level descriptions (e.g. production information, content semantics). The high-level descriptors allow linking external thesauri or knowledge bases and thus the integration of media oriented content descriptions with semantic web technologies. MPEG-7 profiles have been proposed as subsets for certain application areas to reduce the interoperability problem caused by the comprehensiveness and generality of the MPEG-7 standards. Our institute has proposed the Detailed Audiovisual Profile (DAVP, http://mpeg-7.joanneum.at), the first MPEG-7 profile with a description of the semantics of the description elements in the context of the profile in order to solve the interoperability problem.

Applying the Hybrid Approach. As mentioned above, we focus on real-world multimedia applications that typically deal with large media collections. Due to the requirements of these applications, such as scalability, multi-modality, and heterogeneity in terms of context, we follow the hybrid approach. Features automatically extracted by content analysis tools are represented using MPEG-7, while domain semantics are formalized in terms of OWL. Different methodologies of integrating these two representations (e.g. formal-driven, feature-oriented [6]) are utilized depending on the project's needs.

Conclusion

In this paper we have described a number of approaches how to overcome the Semantic Gap. We have shown how solutions could look like in a practical setup, and have pointed out main activities and further directions. Based on the experience gained from our projects, we are convinced that the integration of multimedia content analysis and Semantic Web technologies is necessary in order to build next generation multimedia applications. The currently ongoing standardization activities will provide the basis for these technologies in the near future.

References

[1] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, Ramesh Jain. Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1349–1380, 2000.

[2] ISO/IEC 15938, Multimedia Content Description Interface, 2001.

[3] Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web, Part 1. IEEE MultiMedia, 11(4):38–48, 2004.

[4] Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web, Part 2. IEEE MultiMedia, 12(1):54–63, 2005.

[5] Ben Adida and Michael Hausenblas. RDFa Use Cases: Scenarios for Embedding RDF in HTML. Editor's draft, W3C Semantic Web Deployment Working Group, 2007.

[6] Peter Schallauer, Werner Bailer and Georg Thallinger. A Description Infrastructure for Audiovisual Media Processing Systems Based on MPEG-7. Journal of Universal Knowledge Management, 1(1): 26-35, 2006.