The *K-Space* Network of Excellence: on the Way to 'Semantic' Multimedia

Vojtěch Svátek[1], Petr Berka[1], Jan Nemrava[1,2], Josef Petrák[1], Pavel Praks[1,3], Miroslav Vacura[4], Ebroul Izquierdo[5], Craig Stewart[5]

[1]Department of Information and Knowledge Engineering, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Praha 3, Czech Republic
{svatek, berka, nemrava, petrakj, prakspav}@vse.cz
[2]DFKI (German Research Center for Artificial Intelligence) GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
[3]Department of Applied Mathematics, VŠB – Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
[4]Department of Philosophy, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Praha 3, Czech Republic
vacuram@vse.cz
[5]Department of Electronic Engineering, Queen Mary, University of London, Mile End Road, London E1 4NS, UK
{ebroul.izquierdo, craig.stewart}@elec.qmul.ac.uk

Abstract. The *K-Space* project is an EU FP6 IST Network of Excellence aiming at bringing together research teams from the multimedia and semantics domains. There are fourteen partners, one of them being University of Economics, Prague. Project activities involved shared research as well as dissemination activities.

1 Introduction

Multimedia resources are nowadays abundant on the web as well as in diverse digital archives, and are massively used by various communitites, from journalists through natural scientists and physicians, to e.g. educational workers. However, the retrieval of a multimedia document (audiovisual stream or even a still image) with respect to its content is quite difficult. Similarly to large text collections, indexing has to be predominantly automatic, which however restricts the content of the index to low-level features such as dominant colours, textures, simple shapes, or voice intensity. On the other hand, end-users are interested in retrieving documents related to a certain event, person, location etc., which is often only slightly correlated with low-level features by themselves. This problem is usually called the *semantic gap*. It has recently been postulated that the semantic gap could only be overcome using (in additional to indispensable low-level analysis techniques) a sophisticated blend of metadata standards, well-founded ontological models, efficient metadata storage and retrieval tools, and reasoning engines running on the top of them.

© Václav Snášel (Ed.): Znalosti 2008, pp. 421–425, ISBN 978-80-227-2827-0.
FIIT STU Bratislava, Ústav informatiky a softvérového inžinierstva, 2008.

This was the starting point of the EU-funded FP6 IST project *K-Space*, or "Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content". The project has the status of Network of Excellence and falls under priority "Semantic Knowledge and Content Systems". There are fourteen partners (thirteen universities or public research institutes, and one content provider); the coordinator is the Multimedia & Vision Laboratory at Queen Mary College, University of London.

Due to limited space, in this material we focus on the pivotal tasks of the whole project and on those in which the Czech partner, University of Economics, Prague, has substantial participation[6]. The text is roughly divided to two parts: the first is devoted to research activities, and the second to integration and dissemination activities, which represent an equally important constituent of the vocation of a Network of Excellence. Much more information about the project can be found on the public website `http://www.kspace-noe.net`.

2 Research Activities

The bottom-up stream of research includes elaboration of various multimedia *content descriptors*, expressing the results of analysis ranging from colour and texture through simple shapes to relatively complex notions such as motions and faces. The workflow then contains activities such as shot detection, scene segmentation, genre recognition or event detection. In the audio modality, hot topics are related to music analysis and speech recognition.

Traditional way to provide semantics to the results of such analysis is to use an extensive metadata standard called MPEG-7, expressed in the XML Schema language. However, the semantics of MPEG-7 is inadequate in many aspects: the same information can often be expressed in different ways, there is lack of formal axioms, and the standard itself is sometimes too rigit to reflect the fast developments in the multimedia domain. This is why a collection of axiomatised multimedia ontologies, formalised in the OWL language, was devised as one of crucial K-Space activities. The ontology consists of a *core ontology for multimedia* – COMM [1] plus multiple extensions for different types of media. COMM imports and extends large parts of the DOLCE upper-level ontology [3], which should provide it with deep (philosophically well-grounded) semantics. The ontology has its own API (COMM API), which allows its interoperation with annotation and retrieval tools.

The management of large amounts of metadata, appearing e.g. in the course of analysing a video stream, is, in turn, assured by an *RDF Store* ('storage and inference layer' – SAIL) on the top of Sesame[7]. In order to obtain sufficient performance, a collection of distributed stores operating on networked graphs[8] is envisaged [8].

[6] The bibliography should be understood as biased in this respect as well.

[7] `http://openrdf.com`

[8] `http://www.uni-koblenz.de/FB4/Institutes/IFI/AGStaab/Research/NetworkedGraphs`

On the other hand, metadata not only have to be stored but also processed by *reasoning engines* so as to reveal implicit information. Probably the most important application of reasoners currently is the *merging of semantically labelled domains*. Different tools are tested in combination, including the FiRE fuzzy description logics tool [9], the KAA fuzzy merging tool and the NEST soft-rule-based reasoner [2]. An example of their application is iterative merging and relabelling of image segments depicting a scene consisting e.g. of a sea, sand, sky, vegetation and persons, such that each segment's label influences the probability of its neighbours having a certain label.

In the analysis of images but also of video frames, *multimedia information retrieval* methods play an important role. Methods based on singular value decomposition are for example tested in the area of metallurgy, in cooperation with an external industrial partner [7].

Although *textual resources* are not the central target of multimedia analysis systems, they also play an important role: they serve as so-called *complementary resources* that help disambiguate entities in images or events in video. In the former case, image captions can be subject to NLP-based analysis, and the results can be stored either in an RDF repository or in the image itself [4]. In the latter case, time-stamped online textual reports on (especially sports) events, which are often available on the web, can be synchronised with corresponding video streams, and the event types detected in text (using ontology-based information extraction techniques) can be mapped on events detected using medium-level video/audio detectors: in the case of football, this can be for example field line or crowd detector [5]. The temporal concepts with different provenance (from text, audio and video) and their interrelationships are then captured using a uniform approach based on fuzzy description logics, which allows for derivation of high-level concepts from low level one or for retrieval of portions of video relevant to semantic notions [6].

Most of the mentioned activities eventually lead to back-end components available as plug-ins to an *end-user annotation and retrieval tool*, which is under development, with a working name KAT (K-Space Annotation Tool). KAT interacts with the multimedia ontology through its API, stores its intermediate as well as final results to the Sesame-based repository, and calls various multimedia analysis tools that correspond to end-user requirements.

3 Integration and Dissemination Activities

In terms of joint *research*, an example of large-scale 'ad hoc' intra-network collaboration is the *football broadcast analysis* initiative, carried out both via remote collaboration and via face-to-face visits and longer PhD student exchanges. Video and audio streams, provided by the British (University of Glasgow) and Irish (Dublin City University) partners, are temporally and conceptually aligned with multiple textual online reports from the web. The analysis of textual reports is mainly carried out by the Czech partner (University of Economics, Prague), using the football ontology and information extraction tools from one German

partner (German Research Centre for AI, Saarbrücken). Detection of overlaid text such as scoreboard is carried out by another German partner (Technical University of Berlin), while the actual OCR process is handled by the Austrian partner (Joanneum Research Centre, Graz). Finally, the logical consistency of concepts built across different modalities can be checked and further derivations can be made using the fuzzy reasoner developed by the Greek partner (ITI/CERTH Institute). The purpose of the prototype application under development is to provide the customer viewing (currently, offline) the broadcast with relevant information consistently extracted from both the A/V stream itself and from complementary textual resources, as well as to enable fast semantics-aware discovery of interesting portions of the broadcast.

Another, even larger, cooperation, is associated with *TRECVid*[9] – the annual evaluation benchmark of video analysis and video information retrieval techniques, run by National Institute of Standards and Technologies, USA. The joint K-Space team participated (in 2006 and 2007) in the feature extraction and video search tracks. Individual partners contributed e.g. to detection of concrete concepts such as vegetation or explosion, of genres such as 'sports' or 'indoor'/'outdoor', or of shot characteristics such as camera notion, as well as to measuring intra- and inter-video similarity and to automated speech recognition; the results were then fused and made accessible through an interactive query formulation interface [10].

The *dissemination* and *sharing* of results with the larger research community is assured through K-Space having a central role in establishing a series of conferences—Semantics and Multimedia Technologies (SAMT)—as well as of the annual Summer School on Semantic Multimedia (SSMS). The Network also issues a six-monthly electronic Newsletter. Finally, the persistence of the research and dissemination infrastructure created within K-Space is guaranteed by a newly-formed non-profit association called SMaRT, for "Semantic MultimediA Research and Technology", having as starting point an identically named cluster of similar EU projects[10].

4 Conclusions

The K-Space project, currently entering the third and last year of its lifespan, has already given rise to several practically usable tools for multimedia management and exploitation, set aside dozens of research publications. Numerous fruitful collaborations between partners have been established, which would be hard to achieve otherwise, and provisions for long-term research framework have been laid, most notably, in the form of the SMaRT society.

The research leading to this paper was supported by the European Commission under contract FP6-027026, "Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content" - K-Space.

[9] `http://www-nlpir.nist.gov/projects/t01v/`
[10] `http://www.smart-network.eu/`

References

1. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.: COMM: Designing a Well-Founded Multimedia Ontology for the Web 2007. In: 6th International Semantic Web Conference, ISWC 2007, Springer LNCS.
2. Berka, P., Laš, V., Svátek, V.: NEST: re-engineering the compositional approach to rule-based inference. *Neural Network World*, 2004, Vol. 14, No. 5, 367379. ISSN 1210-0552.
3. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE. In: Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, 13th International Conference, EKAW 2002, Springer LNCS.
4. Kolínsky M., Nemrava J., Svátek V.: Ukladanie výsledkov extrakcie informácií do IPTC hlavičiek športových obrázkov. In this volume.
5. Nemrava J., Buitelaar P., Svátek V., Declerck T.: Event Alignment for Cross-Media Feature Extraction in the Football Domain. In: Int'l Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'07), Santorini, Greece, June 6-8, 2007.
6. Nemrava J., Buitelaar P., Simou N., Sadlier D., Svátek V., Declerck T., Cobet A., Sikora T., O'Connor N., Tzouvaras V., Zeiner H., Petrák J.: An Architecture for Mining Resources Complementary to Audio-Visual Streams. In: Workshop on Knowledge Acquisition from Multimedia Content (KAMC) at SAMT 2007, Genova.
7. Praks, P., Grzegorzek, M., Moravec, R., Válek, L., Izquierdo, E.: Wavelet and Eigen-Space Feature Extraction for Classification of Metallography Images. In: H. Jaakkola, Y. Kiyoki, and T. Tokuda, editors: 17th European-Japanese Conference on Information Modeling and Knowledge Bases (EJC 2007), pages 193–202, Pori, Finland, June 4-8, 2007. Juvenes Print-TTY, Tampere.
8. Schenk S., Petrák J.: Sesame RDF Repository Extensions for Remote Querying. In this volume.
9. Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J.Z., Horrocks, I.: The Fuzzy Description Logic f-SHIN. In: International Workshop on Uncertainty Reasoning for the Semantic Web (2005).
10. Wilkins, P., Adamek, T., Byrne, D., Jones, G., Lee, H., Keenan, G., Mc Guinness, K., O'Connor, N., Smeaton, A.F., Amin, A., Obrenovic, Z., Benmokhtar, R., Galmar, E., Huet, B., Essid, S., Landais, R., Vallet, F., Papadopoulos, G., Vrochidis, S., Mezaris, V., Kompatsiaris, I., Spyrou. E., Avrithis, Y., Morzinger, R., Schallauer, P., Bailer, W., Piatrik, T., Chandramouli, K., Izquierdo, E., Haller, M., Goldmann, L., Samour, A., Cobet, A., Sikora, T., Praks, P.: K-Space at TREVid 2007. NIST TRECVid 2007 - Text REtrieval Conference TRECVid Workshop, Gaithersburg, MD, 5-6 November 2007.