SALERO Intelligent Media Annotation & Search

Wolfgang Weiss
(Institute of Information Systems & Information Management,
JOANNEUM RESEARCH Forschungsges. mbH, Graz, Austria
wolfgang.weiss@joanneum.at)

Tobias Bürger
(Semantic Technology Institute (STI), Innsbruck, Austria
tobias.buerger@sti2.at)

Robert Villa
(University of Glasgow, UK
villar@dcs.gla.ac.uk)

Punitha Swamy
(University of Glasgow, UK
punitha@dcs.gla.ac.uk)

Wolfgang Halb
(Institute of Information Systems & Information Management,
JOANNEUM RESEARCH Forschungsges. mbH, Graz, Austria
wolfgang.halb@joanneum.at)

Abstract: Currently the media production domain lacks efficient ways to organize and search for media assets. Ontology based applications have been identified as a viable solution to this problem, however, sometimes being too complex for non-experienced users. We present the *SALERO Intelligent Media Annotation & Search* system which provides an integrated view onto results retrieved from different search engines. Furthermore, it offers a powerful, yet user-friendly Web-based environment to organize and search for media assets.
Key Words: Information storage and retrieval, semantic media asset management
Category: H.3

1 Introduction

This paper presents the *Intelligent Media Annotation & Search*[1] (IMAS) system, an easy-to-use application based on the principles described in [Bürger 2008a], which is being developed within the European project SALERO[2]. The principles describe methodologies to support users in the process of manual semantic annotation including (i) selection of adequate ontology elements and (ii) extending

[1] `http://salero.joanneum.at/salero/`
[2] `http://www.salero.eu`

of ontologies during annotation time. An integral part of the work being done in SALERO is the management of media objects with semantic technologies which is addressed by the IMAS system by enabling their semantic annotation and retrieval. The use of semantic technologies reduces the problem of ambiguity in search by using existing, well-defined vocabularies, it allows us to do query expansions and to deal with multilinguality.

2 Related Work

The organisation, classification and retrieval of media objects is an ongoing challenge in games and media production. Semantic technologies have been identified as a viable solution to overcome some of the problems in this area [Bürger 2008]. A wide range of multimedia annotation tools [Obrenovic 2007, Simperl 2009] already offer functionality to attach ontological annotations to (parts) of the multimedia content and some offer reasoning services on top of them to semi-automatically create annotations based on existing annotations. The K-Space Annotation Tool [Saathoff 2008] provides a framework around the Core Ontology for Multimedia for efficient and rich semantic annotations of multimedia content. PhotoStuff [Halaschek-Wiener 2005] allows to use any ontology for the annotation of images and is available as a standalone desktop application. A web-based demonstrator for browsing and searching with very limited functionality is also available. Imagenotion [Walter 2007] already provides an integrated environment for the collaborative semantic annotation of images and image parts. User tests showed that the use of standard ontologies and tools is not generally suitable, which led to the development of a method where ontologies consist of imagenotions that graphically represent a semantic notion through an image.

As we experienced as well during prototypical iterations of our system, most paradigms applied in semantic annotation tools are not suitable for unexperienced users who are typically used to keyword-based tagging and suffer from information overload when confronted with complex interfaces.

3 System Description & Design Principles

The IMAS system architecture is shown in [Fig. 1]: The IMAS integrates two systems whose functionalities are offered as a set of Web services, i.e. the *Semantic Services* and the *Content-based Services*. The main aim of the IMAS is to allow easy annotation of media assets for later retrieval and reuse by users in media production. In order to support this, it has been built based on the following design principles:

1. **Designed for content creators.** The target users of the system are non-technically experienced content creators in the domain of media production.

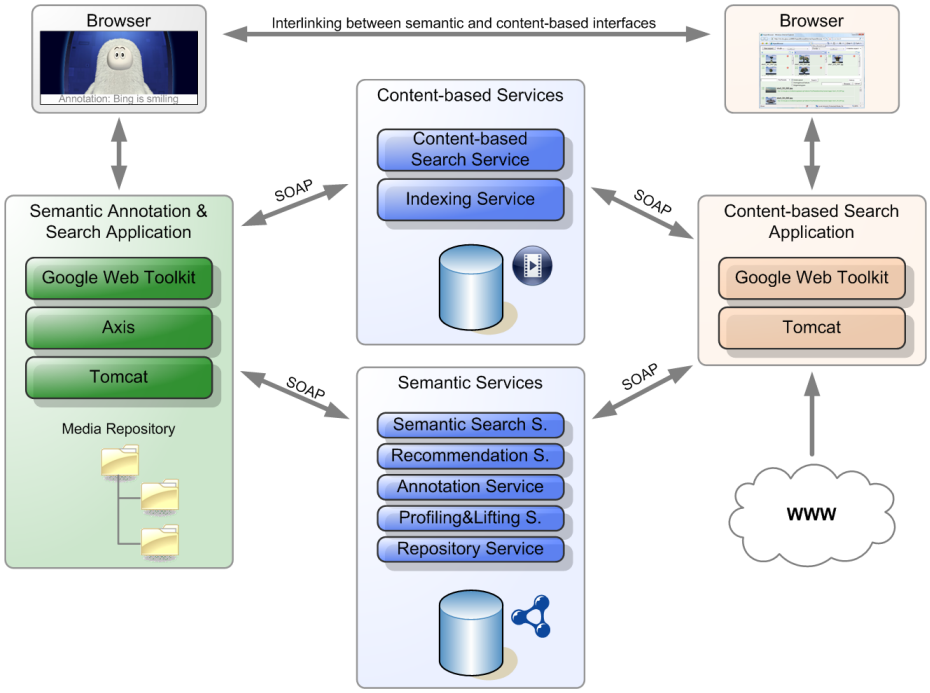

Figure 1: IMAS System Overview.

2. **Easy to use.** The interface provides Web 2.0 – based interaction mechanisms to make the annotation process as easy as possible.

3. **Global annotations.** To facilitate the annotation process, we only allow global annotations of media files instead of annotating parts of it.

4. **Statement-based annotation process.** We allow to create statements, which use ontological elements, to describe the content of media items.

5. **Ontology extension during use.** We allow users to easily extend the ontology during use based on principles described in [Bürger 2008a].

6. **Portability of the system.** In order to port the systems to other domains, only the underlying annotation ontology has to be adapted.

7. **Integration of semantic and content-based search.** The system provides an integrative view onto results from different search engines and by that provides a fallback solution which is able to retrieve objects without annotations too.

4 The Intelligent Media Annotation and Search Application

The IMAS end user application is an integrated Web-based application which can be used to annotate and search for media objects. As illustrated in [Fig. 1], it consumes functionality of (i) the *Semantic Services* which are used to add, update and delete semantic annotations as well as to search the annotation repository and (ii) the *Content-based Services* which are used to retrieve media items based on its intrinsic features such as colour, histograms or shapes.

Usage:

The application allows to annotate any arbitrary files which are stored in pre-configurable media repositories. In order to ease the annotation process for our target user group media items are annotated globally instead of region- or segment-based. Additionally, media items are annotated by creating statements which contain semantic elements which are defined in an domain specific ontology. The annotation statements are formalized according to the ontology defined in [Bürger 2009] and can be seen as sentences about the content of the media item and are in the form of

$$< Resource\ isRelatedTo\ \{Concept_1 ... Concept_n\} >.$$

Using statements with semantic elements is a compromise in complexity between loose and fully semantically described annotations. [Fig. 2] illustrates statements with an example. To create such statements, three different input options are

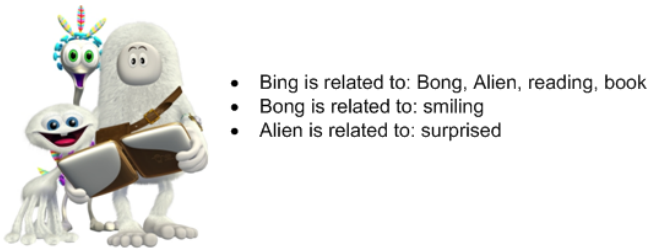

Figure 2: Example of Annotation Statements.

available as shown in [Fig. 3]: (1) combining concepts via drag-and-drop, (2) selecting concepts consecutively and (3) using the text box as a command line interface in the spirit of [Raskin 2008] with auto-completion. Input option three is optimally suited for frequent users and input options one and two are ideal for users who rarely create annotations.

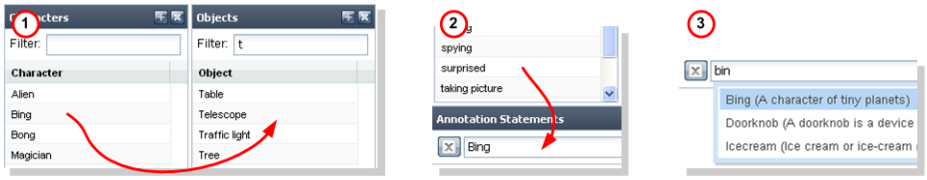

Figure 3: Creation of Annotation Statements.

4.1 Semantic Services

IMAS is realized on top of the *SALERO Semantic Workbench* (cf. [Bürger 2008]) which not only offers a graphical user interface to engineer ontologies but also a set of services which provide ontology management functionality to other applications. The services offered by the semantic workbench include:

- **The Repository Service** which offers an API for the persistent storage of WSML[3] ontologies and their elements. It supports management of elements and the execution of SPARQL queries. The services are realized on top of the *Ontology Representation and Data Integration (ORDI)* – framework[4].
- **The Annotation Service** is concerned with the management of semantic annotations and provides an API to manage and validate annotations against the ontologies stored in the repository.
- **The Semantic Search Service** offers an API to search for ontology elements and additionally offers keyword-based search for annotations which is expanded into full-text queries on a generated index and SPARQL queries.
- **The Recommendation Service** offers an API for retrieval of ontology elements which are prominently used for annotation and gives recommendations of related results during search.
- **The Profiling and Lifting Service** can be used to extract structural semantic information from existing MPEG-7[5] documents and for their semantic enrichment.

4.2 Content-based Services

The *Content-based Services* offer functionality for the indexing and retrieval of image, video and textual information. The aim of the service is to complement

[3] `http://www.wsmo.org/wsml/wsml-syntax`
[4] `http://www.ontotext.com/ordi`
[5] `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`

the semantic search services. As such their emphasis is on automatic indexing techniques, which can be used to retrieve images, text or video without manual annotation. This provides a fall-back system in the cases where material is not indexed by the semantic services, providing an alternative method of retrieval and bootstrapping. Architecturally, the content-based services are split into a backend indexing and retrieval system which provides a SOAP interface to other software components and a frontend search interface for end users. Textual information is indexed using standard Information Retrieval techniques (the Terrier system is used [Ounis 2006]); image and video data is indexed by extracting low-level visual features based on the MPEG-7 standard, as currently implemented in the ACE toolbox [O'Connor 2005]. Searches can be carried out on locally uploaded and indexed data or the system can also use third party Web services to search remote collections, such as Yahoo BOSS[6] or YouTube. Both local collections and remote services can be searched via the same search interface. The interface allows searching based on an optional textual query and zero or more image examples which may come from already indexed collections or be ad-hoc serialized images. The AspectBrowser [Villa 2009], a flexible environment for search and search result organisation, is the end user interface to the content-based services.

4.3 Search System

To search for media objects, the following input options are available to create a query: (i) free text, (ii) semantic concepts, (iii) statements, and (iv) images. Free text search is both executed in the *Semantic Services* and the *Content-based Services*. The concept-based search is expanded in the *Semantic Services*. Via the exemplary images a query is submitted to the *Content-based Services*. The results of both systems are merged based on the respective rank values being returned.

4.3.1 Semantic Search Algorithm

The *Semantic Services* support two different types of queries: The first option are *concept-based* queries which are translated into SPARQL queries and which are evaluated via the *Repository Service* of the *Semantic Services* in the repository. The second option are *keyword-based* queries which are executed in a Lucene[7] index which is generated based on the ontologies and instance data in the repository. The Lucene index preserves the annotation triples and also stores subsumption information to enable efficient retrieval. Keyword and concept queries can be mixed to increase the precision of keyword-based querying in the system.

[6] `http://developer.yahoo.com/search/boss/`
[7] `http://lucene.apache.org`

Furthermore selected WordNet relations of the concepts can be expanded to increase the result set while preserving precision. Once the queries are answered, the results are ranked based on the semantics of the annotations attached to the returned results and their degree of match of the query. The ranking is based on the popular tf/idf measure which is customized for semantic annotations to take into account triples instead of just terms and the subsumption hierarchies.

4.3.2 Search Result Fusion

The core problem of generating the final list lies in devising a mechanism of combining the results from each system to produce effective results through the IMAS engine. The problem of combining results is widely studied, both by information retrieval communities as well as multimedia retrieval groups at various levels: combining results from many query examples, combining results originating from multiple features or results obtained from different search engines. Multimedia information retrieval communities refer to this as multimodal fusion (cf. [Mc Donald 2005], [Hoi 2008], [Qi 2005]) and querying with multiple examples (cf. [Westerveld 2004], [Kludas 2008]).

In our system a *round robin* mechanism combined with a *polling-based result fusion* mechanism is adopted to fuse results obtained from semantic and content-based search systems. A keyword-based search is employed to obtain the initial list of n results from the semantic search system. The top $m << n$ results are then used by the content-based search to retrieve more visually similar images. Since it is also important to select the best feature or re-order the features according to their preference for the query examples favouring to obtain more diverse results, the knowledge embedded in queries are exploited by computing the correlation and t-distribution in the queries. The feature which has less correlation and high t-distribution is chosen as the most useful feature.

5 Conclusion and Outlook

In this paper we have presented the *Intelligent Media Annotation & Search*[8] system. The system is designed for the content creators in the realm of media production and uses a semantic enabled statement-based approach for fast and easy annotation of media files. The system also integrates semantic and content-based search to provide a fall-back and an alternative retrieval system. It is planned to explore a more sophisticated fusion mechanism with the automatic computation of the preference of the semantic system or the content-based system for a topic. A thematic browsing through the content, based on the ontologies, would also be feasible. Furthermore we are currently in the process of evaluating the integrated system; results are expected in the coming months.

[8] `http://salero.joanneum.at/salero/`

Acknowledgements: The research leading to this paper was partially supported by the European Commission under contract "IST-FP6-027122" (SALERO).

References

[Bürger 2008] Bürger, T.: The need for formalizing media semantics in the games and entertainment industry. Journal for Universal Computer Science (JUCS) (June 2008).

[Bürger 2008a] Bürger, T., Ammendola, C.: A user centered annotation methodology for multimedia content. In: Poster Proceedings of ESWC 2008. (2008).

[Bürger 2009] Bürger T., Kienast G., Weiss W., Villa R., Swamy P., and Hannah D.: D3.1.6 - Embedding of Semantic Metadata in Content Formats. SALERO Public Deliverable D3.1.6, (2009).

[Halaschek-Wiener 2005] Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B., Hendler J.: PhotoStuff - An Image Annotation Tool for the Semantic Web. In: Poster Proceedings of the 4th ISWC (2005).

[Hoi 2008] Hoi S. C. H., Lyu M. R.: A multimodal and multilevel ranking scheme for Large-Scale Video Retrieval. In: IEEE transactions on Multimedia. Vol. 10, No. 4, 607 -619, (2008).

[Kludas 2008] Kludas, J., Bruno, E., Marchand-Maillet S.: Can Feature Information Interaction help for Information Fusion in Multimedia Problems?. In: Proc. of the first International Workshop on Metadata Mining for Image Understanding (MMIU), Funchal, Madeira, (2008).

[Mc Donald 2005] Mc Donald, K., Smeaton A. F.: A comparison of score, Rank and Probability-Based Fusion Methods for Video Shot Retrieval. In: In Proc of Intl. Conf on Image and Video Retrieval. 61-70, (2005).

[Obrenovic 2007] Obrenovic, Z., Bürger, T., Popolizio, P., Troncy, R.: Multimedia semantics: Overview of relevant tools and resources. `http://www.w3.org/2005/Incubator/mmsem/wiki/Tools_and_Resources` (September 2007).

[O'Connor 2005] O'Connor, N., Cooke, E., le Borgne, H., Blighe, M., Adamek, T.: The ace-toolbox: low-level audiovisual feature extraction for retrieval and classification. In: 2nd IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies. (2005) 55–60.

[Ounis 2006] Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Lioma, C.: Terrier: A High Performance and Scalable Information Retrieval Platform. In: Proc. of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR). (2006).

[Qi 2005] Qi, X., Han, Y.: A novel fusion approach to content based image retrieval. In: Pattern Recognition. Vol.38, 2449 -2465, (2005).

[Raskin 2008] Raskin, A.: The Linguistic Command Line. In: ACM Interactions. (January and February 2008) 19–22.

[Saathoff 2008] Saathoff, C., Schenk S., Scherp A.: KAT: The K-Space Annotation Tool. In: Proccedings of the SAMT 2008 Demo and Poster Session, (2008).

[Simperl 2009] Simperl, E., Tempich C., Bürger, T.: Methodologies for the creation of semantic data In: Miguel-Angel Sicilia (ed.) Handbook of Metadata, Semantics and Ontologies, World Scientific Publishing Co., 2009 (forthcoming).

[Villa 2009] Villa, R., Cantador, I., Joho, M., Jose, J.: An aspectual interface for supporting complex search tasks. In: Proceeding of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Boston, USA: ACM (in press). (2009).

[Walter 2007] Walter, A., Nagypal, G.: IMAGENOTION - Collaborative Semantic Annotation of Images and Image Parts and Work Integrated Creation of Ontologies. In: Proceedings of 1st Conference on Social Semantic Web (CSSW). (2007) 161–166.

[Westerveld 2004] Westerveld, T., de Vries A.P.: Multimedia Retrieval using Multiple Examples. In: LNCS Image and Video Retrieval. Vol. 3115, (2004) 344-352.