Statement-based Semantic Annotation of Media Resources

Wolfgang Weiss[1], Tobias Bürger[2], Robert Villa[3], Punitha P.[3], and Wolfgang Halb[1]

[1] Institute of Information Systems,
JOANNEUM RESEARCH Forschungsges. mbH, Graz, Austria
firstname.lastname@joanneum.at
[2] Semantic Technology Institute (STI), Innsbruck, Austria,
tobias.buerger@sti2.at
[3] University of Glasgow, UK
{villar, punitha}@dcs.gla.ac.uk

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 a fast and easy to use approach to create semantic annotations and relationships of media resources. The approach is implemented in the *SALERO Intelligent Media Annotation & Search* system. It combines the simplicity of free text tagging and the power of semantic technologies and by that makes a compromise in the complexity of full semantic annotations. We present the implementation of the approach in the system and an evaluation of different user interface techniques for creating annotations.

Keywords: Semantic annotation, Information storage and retrieval, semantic media asset management

1 Introduction

The management of media resources in media production is a continuous challenge due to growing amounts of content. Because of the well-known limitations, manual annotation of media is still required. We present a statement-based semantic annotation approach which allows fast and easy creation of semantic annotations of media resources. The approach is implemented in the *Intelligent Media Annotation & Search*[4] (IMAS) system, which is being developed within the European project SALERO[5]. 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

[4] http://salero.joanneum.at/imas/
[5] http://www.salero.eu

in search by using existing, well-defined vocabularies, it allows us to do query expansions, and to deal with multilinguality.

During prototypical development iterations of our system we have experienced, that most paradigms applied in semantic annotation tools are not suitable for inexperienced users who are typically used to keyword-based tagging and suffer from information overload when confronted with complex annotation tasks and user interfaces. Our aim was thus to develop an approach which is faster and easier to use for our targeted user group, while making a compromise in complexity of full semantic annotations. Beside describing the content of each media resource, the approach allows to relate media resources to other media resources.

In the following we present the IMAS system and the implemented annotation approach. The system is based on the principles described in [1]. These principles describe methodologies to support users in the process of manual semantic annotation including (i) selection of adequate ontology elements and (ii) extending of ontologies during annotation time. Furthermore, we present an evaluation of different user interface techniques for creating annotations.

The remainder of this paper is organized as follows: Firstly we present the IMAS system (Section 2). Secondly we present the statement-based semantic annotation approach and its evaluation (Section 3). Then we are situating our work with related work in the area (Section 4) and conclude the paper with a summary and outlook to future work (Section 5).

2 System Description & Design Principles

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.
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 annotation of media resources 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 resources.
5. **Ontology extension during use.** We allow users to easily extend the ontology during use based on principles described in [1].
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.

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 architecture of the IMAS system is shown in [Fig. 1]:

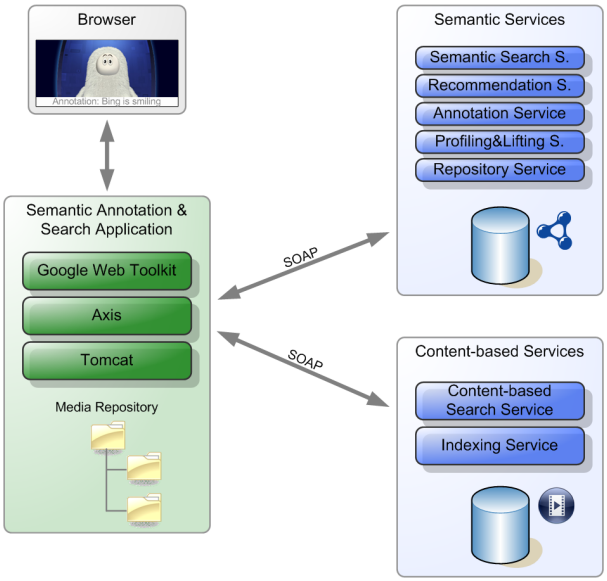

Fig. 1. IMAS System Overview.

2.1 Semantic Services

IMAS is realized on top of the *SALERO Semantic Workbench* (cf. [2]) 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. Most notably the services offer persistent storage of ontologies and annotations and the retrieval of stored information.

2.2 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 the semantic search services and as a fall-back system in the cases where material is not indexed by the semantic services. As such their emphasis is on automatic indexing techniques, which can be used to retrieve images, text or video without manual annotation. Textual information is indexed using standard Information Retrieval techniques (the Terrier system is used [3]); 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 [4].

2.3 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 executed in both the *Semantic Services* and the *Content-based Services*. The concept-based and statement-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 integrated on a late fusion fashion. In our system a round robin mechanism combined with a polling-based result fusion technique is adopted to fuse results. Further details of the backend services can be found in [5–7].

3 Statement-based Annotation

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 resources based on its intrinsic features such as colour, histograms or shapes.

3.1 Usage

The application allows to annotate arbitrary resources which are stored in pre-configurable media repositories. In order to ease the annotation process for our target user group, media resources are annotated globally instead of region- or segment-based. Media resources are annotated by creating statements and by relating them to other media resources. Annotation statements contain semantic elements which are defined in the annotation ontology (see also section 3.2). The annotation statements are formalized according to the annotation ontology and represent natural language-like statements about the content of the media resource. Statements are in the form of

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

which are triples where a concept can be set in relation to other concepts. Using statements with semantic elements is a compromise in complexity between loose and fully semantically described annotations. [Fig. 2] illustrates statements with an example Image from the tiny planets[6] universe.

To create such statements, three different input options are 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 [8] 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.

[6] `http://www.mytinyplanets.com/`

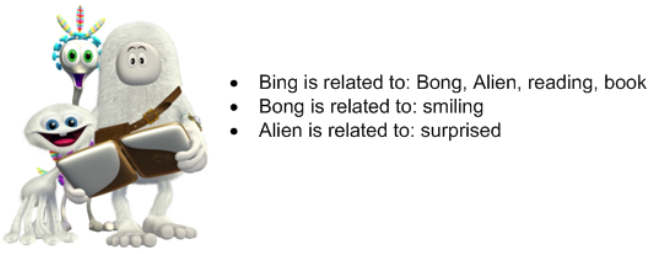

Fig. 2. Example of Annotation Statements.

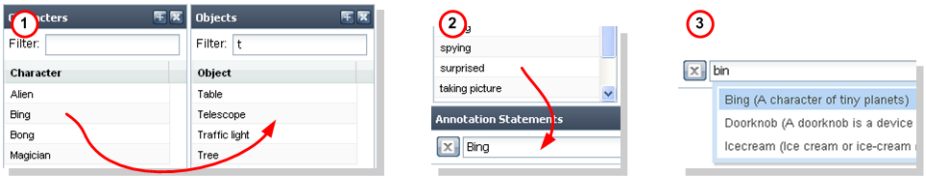

Fig. 3. Creation of Annotation Statements.

An additional possibility to annotate the content of media resources in the IMAS system is to relate them to each other. Hereby, we can describe that one media resource is, for instance, a *part*, a *revision*, or a *transformation* of another media resource. This allows us to use statements of the related media resources, to keep track of revisions of the media resources or to suggest alternatives in the search result. The relationship ontology with its properties is described subsequently in the section 3.2. To create relationships (see also [Fig. 4]) of selected media resources (1), the user drags a source media resource from the file browser (2) and drops it on the desired relationship (e.g. the *revision*) of the relationship panel (3). This action creates the following annotation (4):

$$< \textit{SourceResource} \ \textbf{\textit{is a revision of}} \ \textit{TargetResource} >.$$

3.2 Ontologies

To represent the semantic annotations we developed two ontologies according our needs. The first one is the SALERO annotation ontology (c.f. [Fig. 5]), for describing media resources, authors of media resources, projects in which they are used, and annotations. The property *annotea:related*, derived from the Annotea Annotation Schema[7], is used to describe the annotation statements. The subclasses of *Annotation* and *AnnotationConcept* form the domain specific parts of the ontology, which have to be adapted if someone wants to use the annotation tool in a different domain. The subclasses of *Annotation* describe media specific annotation types such as annotations of audio, image or video.

[7] `http://www.w3.org/2000/10/annotation-ns`

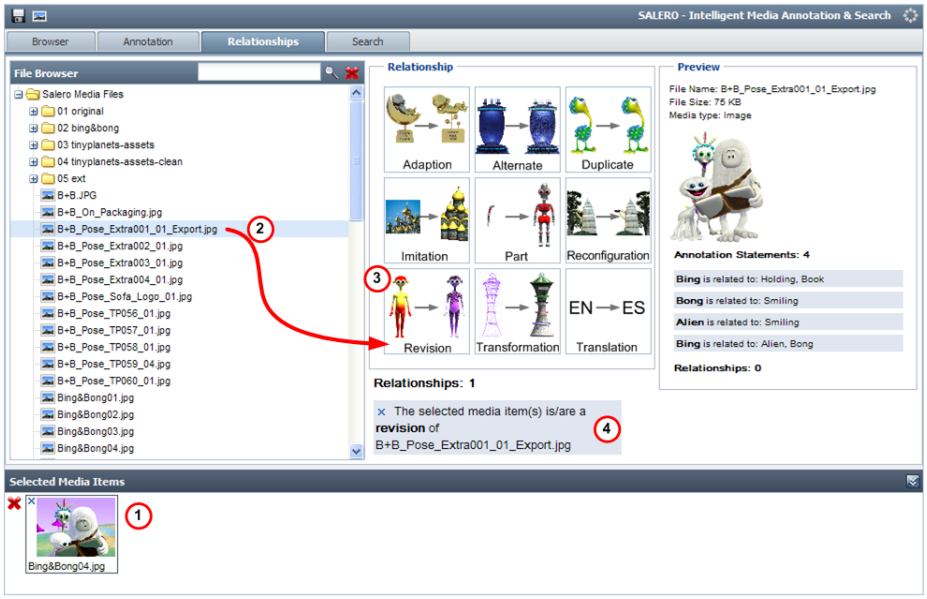

Fig. 4. Creation of relationships.

The subclasses of *AnnotationConcept* represent the domain-specific part of the annotation ontologies and are used to describe the content of media resources. They currently include the following classes:

- **Character:** The actors of a movie, e.g. *Bing and Bong.*
- **CharacterParts:** Parts of the body of the characters, e.g. *Hand, Nose, or Toe.*
- **Expression:** Includes verbs and adjectives to describe the behaviour of characters, e.g. *smiling, dancing, open, or wet.*
- **Location:** A geographical or imaginary location, e.g. *Barcelona or Outer Space.*
- **Object:** A tangible or visible entity, e.g. *Balloon, Umbrella, or Cake.*

The scope of the second ontology is to describe relationships between media resources and how they are most probably derived or based on each other. The relationships of the Functional Requirements for Bibliographic Records (FRBR)[8] model provide a solid ground to describe the relationships of media resources in the media production domain in general and in the SALERO project in particular. The relationships are supposed to enhance the browsing experience of media professionals in their media collections.

The relationships of our ontology were chosen based on a general analysis of the domain based on related work, an analysis of images from the MyTinyPlanets

[8] `http://www.ifla.org/VII/s13/wgfrbr/index.htm`

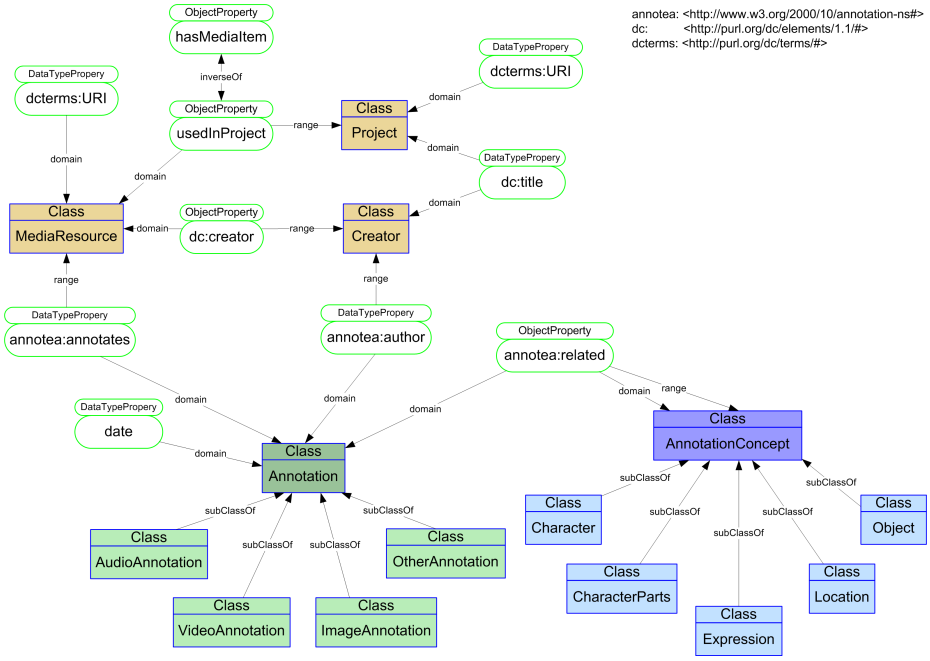

Fig. 5. The SALERO annotation ontology.

collection, and a set of expert interviews. It contains a subset of the FRBR core ontology[9] and two additional relationships, which were not covered by the FRBR core ontology (frbr: http://purl.org/vocab/frbr/core#) (c.f. [9, 10]):

- **frbr:adaptation** A media resource, which is based on another media resource, exchanges parts of the original content, e.g. in a new image a tree is exchanged by a space ship.
- **frbr:alternate** An alternative (file) format of the same content, e.g. jpg instead of png.
- **frbr:imitation** A real world scene is imitated in a cartoon, e.g. a scene of "Star Wars" is imitated by Bing & Bong in "Tiny Planets".
- **frbr:part** A media resource contains a part of another media resource, e.g. the hands of the character Bing are a *part of* the final animation of Bing. Supposing that the hands are modelled in a different file.
- **frbr:reconfiguration** A rearrangement of the content of a media resource, e.g. a new scene is based on an existing scene with the same content such as trees, space ships, characters. In the new scene the content is locally rearranged.
- **frbr:revision** A newer or other version of a media resource, e.g. a new version of a space ship, which is based on an existing space ship, uses a different texture for its surface.

[9] `http://vocab.org/frbr/core.html`

- **frbr:transformation** From a sketch to a (3-d) model, e.g. the first step in creating a 3-d model is to draw a sketch on a sheet of paper. The next step is to create a 3-d model in the computer, which is a *transformation* of the sketch.
- **frbr:translation** A translation into a different language, e.g. the embedded text and audio of a clip are translated from English to French.
- **duplicate** For duplicated media resources, e.g. the same file is stored at a different location.
- **version** The super property for *frbr:adaption*, *frbr:reconfiguraiton*, *frbr:revision* and *frbr:transformation*.

The property *relationships* is the super property for all relationship properties, its range and domain is *MediaResource*.

3.3 Usability Test

An initial evaluation of the annotation aspect of the IMAS is already done. Our aim was (i) to find major obstacles in the annotation user interface as well as in the annotation process and (ii) to compare the semantic statement-based approach with other existing approaches.

Evaluation Methodology: We recruited 9 participants of our institute who were not involved in the project or in the realm of semantic annotation. The subjects ranged in age from 25 to 40 and all are software developers. We created two groups. The first group of five participants compared the annotation process of the IMAS application with the desktop application of Google Picasa[10]. In the second group we did chronometries of creating annotations with (i) IMAS, (ii) a free text tagging approach, similar to flickr[11] and (iii) creating fully semantic annotations with PhotoStuff [11]. For later analysis we did screen captures of the user actions and conspicuous behaviour was noted by the observer. Furthermore, the usability test consisted of a short user questionnaire where the participants had to answer following questions:

- What are the positive and negative impressions of each tool?
- What was the most difficult task?
- Which tool would you use, if you have to annotate 1000 images?

Before the test began, each subject got an introduction in the usability test and into the different annotation applications. Then the users had to annotate the test data with each application. The test data consisted of ten images from the tiny planets[12] universe together with descriptive text for each image. For example, the descriptive text for [Fig. 2] is: *"Bing has a bag on his shoulder and a book in his hands. Bong and alien are smiling."*

[10] `http://picasa.google.com/`
[11] `http://www.flickr.com/`
[12] `http://www.mytinyplanets.com/`

Results: The success rate to complete the annotation tasks with IMAS is 93%. The reasons why not all participants where able to successfully complete all tasks are (i) user interface and design problems and (ii) that the semantic statement-based approach was not clear for every participant and thus produced wrong or incomplete annotations. A successful example, done by a participant with IMAS, includes following statements for the image [Fig. 2]:

- Bing *is related to:* Shoulder bag, Shoulder
- Bing *is related to:* Book, Hand
- Bong *is related to:* smiling
- Alien *is related to:* smiling

and following tags created with Google Picasa: *"alien, bing, bong, book, hand, shoulder bag on his shoulder, smiling"*. These examples describe the content according the descriptive text of the image well and fulfils our quality requirements for a successful annotation.

[Fig. 6] shows the time measurements for completing the tasks where the statement-based approach is compared to a free text tagging approach and to creating fully semantic annotations with PhotoStuff. Creating annotations with PhotoStuff was the most time consuming approach (median 60s; mean 69.9s for creating the annotations of a single media resource). The subjects complained that they have to do long-winded recurrent tasks, such as selecting concepts and manually creating instances. The fastest approach was to use the simple free text tagging approach (median 19s; mean 18.6s), although the subjects claimed to use a system with auto completion feature. Task completion time using IMAS with the statement-based approach ranks between the two other systems with a median of 30s and a mean of 36.3s.

We observed that most subjects firstly used the concept tables (see also [Fig. 3] and the section 3.1) and after annotating approximately three media resources the subjects tended to use only the command line interface with auto completion. The questionnaire revealed following facts: The users liked the auto completion feature which demonstrates following user statement:

> *"I highly appreciate that the system suggests concepts to me I can use."*

In the subjects opinion this feature helps efficiently to create specific suitable annotations. Furthermore, this was a crucial reason why for 8 out of 9 participants IMAS is the first choice for annotating 1000 media resources. One participant prefers to use a simple free text tagging approach. On the other hand, the participants also revealed new suggestions to improve the annotation process:

> *"A copy and paste functionality of concepts and statements would be fine."*

> *"Existing annotations should be editable."*

The users also complained that the text box does not work as expected in some situations, e.g. when pressing the space bar to select a concept.

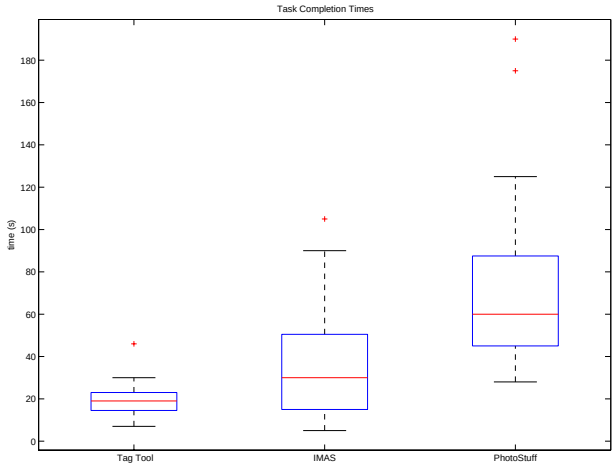

Fig. 6. Comparison of task completion times.

4 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 [2]. A wide range of multimedia annotation tools [12, 13] 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 [14] provides a framework around the Core Ontology for Multimedia for efficient and rich semantic annotations of multimedia content. PhotoStuff [11] 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 [15] 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.

A first step in closing the semantic gap, between low level signal processing and high level semantic descriptions, is to use a multimedia ontology infrastructure. The Core Ontology on Multimedia (COMM) [16] aims to provide a high quality multimedia ontology that is compatible with existing (semantic) Web technologies. The MPEG-7 based multimedia ontology enables to describe algorithms and digital data, decompositions, region locations, content and media annotations, semantic annotations and example annotations. The objective

of the AIM@SHAPE [17] ontology for virtual humans is to provide a semantic layer to reconstruct, stock, retrieve and reuse content and knowledge of graphical representations of humans. The ontology allows to represent, amongst other things, shapes, parts of bodies, animations and emotions. The aim of the W3C Annotea [18] project is to enhance collaborations via shared metadata based on Web annotations, bookmarks, and their combinations. The project encourages users to create Annotea objects, including annotations, replies, bookmarks and topics. Therefore, a RDF schema can be used which defines all necessary properties.

5 Conclusion and Outlook

In this paper we have presented a *semantic statement-based* approach for fast and easy annotation of media resources in the realm of media productions. This concept is implemented in the *Intelligent Media Annotation & Search*[13] application which also allows to create relationships of media resources. Creating annotations using statements with semantic elements is a compromise in complexity and expressiveness between loose and full semantical descriptions. We have developed two ontologies to store annotations and relationships. The system integrates semantic and content-based search to provide a fall-back and an alternative retrieval system. The initial usability test has shown that the approach of semantic statement-based annotation is not as fast as free text tagging but much faster than creating full semantic annotations with PhotoStuff.

It is planned to explore a more sophisticated fusion mechanism for the search system. A thematic browsing through the content, based on the ontologies, would also be feasible. An important item is a further evaluation of IMAS, especially the search system in combination with the annotations, to determine the limitations of this approach.

Acknowledgements.

The research leading to this paper was partially supported by the European Commission under contract "IST-FP6-027122" (SALERO). Images from "My Tiny Planets" are by courtesy of Peppers Ghost Productions[14].

The authors would like to thank their colleagues who contributed to the implementation of the components described here, especially Georg Thallinger, Gert Kienast, Philip Hofmair, Roland Mörzinger and Christian Ammendola.

References

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

[13] `http://salero.joanneum.at/imas/`
[14] `http://www.peppersghost.com`

2. Bürger, T.: The need for formalizing media semantics in the games and entertainment industry. Journal for Universal Computer Science (JUCS) (June 2008)
3. 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)
4. 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
5. Punitha, P., Jose, J.M.: Topic prerogative feature selection using multiple query examples for automatic video retrieval. In: Proceeding of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Boston, USA: ACM (in press). (2009)
6. Punitha, P., Jose, J.M.: How effective are low-level features for video retrieval (submitted for review). (2009)
7. Weiss, W., Bürger, T., Villa, R., Swamy, P., Halb, W.: SALERO Intelligent Media Annotation & Search. In: Proceedings of I-SEMANTICS 09 - International Conference on Semantic Systems. Graz, Austria. (2009)
8. Raskin, A.: The Linguistic Command Line. In: ACM Interactions. (January and February 2008) 19–22
9. IFLA Study Group: Functional requirements for bibliographic records, final report. `http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf` (February 2009)
10. Riva, P.: Introducing the functional requirements for bibliographic records and related ifla developments. ASIS&T The American Society for Information Science and Technology (August/September 2007)
11. Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B., Hendler, J.A.: PhotoStuff - An Image Annotation Tool for the Semantic Web. In: Poster Proceedings of the 4th International Semantic Web Conference. (2005)
12. 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)
13. Simperl, E., Tempich, C., Bürger, T.: Methodologies for the creation of semantic data. In Sicilia, M.A., ed.: Handbook of Metadata, Semantics and Ontologies. World Scientific Publishing Co. (2009)
14. Saathoff, C., Schenk, S., Scherp, A.: KAT: The K-Space Annotation Tool. In: Proccedings of the SAMT 2008 Demo and Poster Session. (2008)
15. Walter, A., Nagypál, 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
16. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.: COMM: Designing a Well-Founded Multimedia Ontology for the Web. In: 6th International Semantic Web Conference (ISWC'2007), Busan, Korea
17. Garcia-Rojas, A., Vexo, F., Thalmann, D., Raouzaiou, A., Karpouzis, K., Kollias, S.: Emotional body expression parameters in virtual human ontology. In: In Proceedings of 1st Int. Workshop on Shapes and Semantics, Matsushima, Japan, June 2006, pp. 63-70. (2006)
18. Koivunen, M.R.: Annotea and semantic web supported collaboration. In: Proceedings of UserSWeb: Workshop on End User Aspects of the Semantic Web. Heraklion, Crete. (2005)