Semantic Modeling of Digital Multimedia

Babak Akhgar, Jawed Siddiqi, Fazilatur Rahman, Nazaraf Shah
Faculty of ACES
Sheffield Hallam University, UK
Nahum Korda
Straight Technology UK

Raphael Attias, Norbert Benamou
ORT, France
Andrade Maria Teresa
INESC, Portugal),
Judy Dori
TECHNION University, Israel
Boas Hashavia
IDEA, Israel

Abstract **-** *The requirement for a commonly accepted efficient mapping between multimedia metadata standards and semantic web-ontology standards is a major issue recognized by semantic multimedia research community. Though there have been several attempts to translate MPEG-7 audio descriptions to ontology languages there is very little literature that addresses issues associated with streaming video contents. In this paper we outline our plan to develop a methodology and the corresponding software implementation of mapping techniques of MPEG-21 video items. The novelty of our effort lies in the fact that we address the complexity of video content's metadata descriptions and its integration with the well recognized ontological standard through transparent mapping from original XML to RDF description. The validity of the proposed method and implementation detail will be verified against the MOSAICA semantic framework and its use cases.*

Keywords: Multimedia, Semantic Web, Ontology, MPEG, Video.

1 Introduction

Content providers like *television broadcasters* and cultural *archives* are lagging behind with regard to ontology-integrated personalized systems. In order to provide personalized and context aware access to content (mostly digital multimedia contents) collected from different heterogeneous disjoint sources requires an understanding of the content as well as users using them. The *Semantic Web* technologies provide the means to achieve a common understanding of content and concepts needed to integrate and map content collections with enriched reasoning facilities to infer new knowledge to offer personalized service to the users. For this purpose, use of ontology-based approach has been commonly accepted by many researchers [1]. For a widespread use of ontologies in information integration and exchange, a prerequisite is the achievement of a joint standard for describing ontologies [2]. Standards activities for Semantic Web languages are mainly driven by working groups of the W3C; in particular the Semantic Web layer cake [3] proposed by Tim Berners-Lee shows the layering of the current state-of-the-art and future planned standards. While XML as a baseline allows for a syntactical description of documents, the layers RDF, Ontology and Logic are adding machine-process-able semantics - a necessary prerequisite for sharable web resources.

Currently the gap in the syntactic and semantic interoperability between semantic web technologies and existing MM annotation standards recognizes that a major issue is the problem of aligning semantic web-based approaches with MPEG metadata descriptions. Choosing vocabularies to use when annotating MM is a key decision to be made as we need more than a single vocabulary to cover the image's/object's different relevant aspects. Many vocabularies as described by standards were developed prior to the semantic web. So the task is to translate the vocabularies to RDF or OWL. The key ISO MM standard, MPEG-7, MPEG-21 are defined using XML schema. Presently there is no commonly accepted mapping [5, 6] from XML schema to RDF or OWL (semantic web standards).

This paper aims to present our plan to innovate a methodology and software developed for the interoperability of RDF with the complete MPEG-21 so that domain ontologies described in RDF can be transparently integrated with the MPEG -21 metadata. This allows applications that recognize and use the MPEG -21 constructs to make use of domain ontologies for applications like indexing, retrieval, filtering etc. resulting in more effective user retrieval and interaction with audiovisual material.

The idea described in this paper will be implemented in the context of the MOSAICA architecture [22] for semantic annotation & retrieval of video items. There are several works

([9], [7], [8], [10], [11]) concentrated towards interoperability issues of semantic descriptions of video elements and mapping techniques. But perhaps more significantly none of them considered the versatile characteristics of video contents and mapping issues specific to them. The novelty of our plan is to design a mapping method of MPEG-21 for video items to RDF metadata and develop software to demonstrate the performance of the implementation. Our implementation will further be tested and validated against the MOSAICA semantic architectural framework. The proposed approach will be based on the translation to RDF/OWL ontology, which fully captures the MPEG-21 video items and corresponding automatic mapping between XML metadata and the ontology described in RDF.

The rest of the paper is organized as follows: Section 2 elaborates on the multi-media standardization activities and semantic interoperability issues. In section 3, we discuss contemporary research efforts in the area of modelling semantic multimedia. Section 4 shows our research plan for achieving a mapping method. Finally we conclude in section 5 drawing some conclusions on the progress.

2 Multi-Media Metadata Descriptions

2.1 Standardization and Standards

There are several standardization activities that have already taken place to use the standards to serialize metadata descriptions. The most important of these are [4]: Dublin Core ElementSet, MPEG-7, MPEG-21, NewsML, TV-Anytime, Virtual Resource Association (VRA) Core etc. Although several prominent activities have provided standards for describing MultiMedia (MM) content they are not widely used for several reasons. Firstly, it's difficult and/or expensive to manually annotate multimedia content. Secondly, the complexity of many standards makes multimedia annotation unnecessarily difficult. Thirdly, there's little incentive for organizations to provide multimedia metadata because there are insufficient applications that would benefit from its use [4].

2.2 Key representation issues

The information conveyed by MM documents may be represented with three different levels of abstraction [4]:

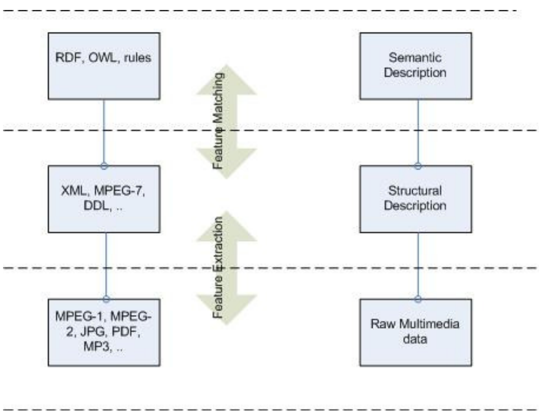

Figure 1: *The different levels of MM information and the type of annotation provided [4]*

The raw multimedia information represented in well known formats for video, image, audio and text metadata. These are binary formats, compressed for streaming delivery, not necessarily well-suited for further processing for meta information. The middle layer which lets us use feature detectors' output, (multicue) segmentation algorithms to provide a structural layer on top of the binary media stream. Information on this level is typically serialized in XML. The standards mainly operate in this middle layer of Figure 1. The problem with this structural approach is that the semantics of the information encoded in the XML are only specified within each standard's framework. For example, if we use the MPEG-7 standard, then it's hard to reuse this data in environments that aren't based on MPEG-7 or to integrate non-MPEG metadata (e.g. Dublin core, TV Anytime) in an MPEG-7 application. This conflicts with the interoperability of Web-based applications. To address this problem, a possible solution is to add a third layer for semantics and logical abstraction level. Logical abstraction level provides the semantics for the middle layer, actually defining mappings between the structured information sources and the domain's formal knowledge representation. An example of this is the Web Ontology Language (OWL). In this layer, we can make the implicit knowledge of the multimedia document description explicit and reason with it to derive new knowledge not explicitly present in the middle layer.

2.3 Semantic gap in Multi media

So, to take advantage of the existing MM metadata descriptions we need a formal mapping method that will translate those meta descriptions to semantic web description standards. Among the contemporary efforts in bridging the semantic gap in MM contents some of the researchers have used the RDF ontology definition language to partially describe the MPEG-7 content metadata structures, but not the complete MPEG-7, while others tried to provide methodologies but did it in an ad-hoc basis with the preconceived assumptions of specific domain ontologies. On the other hand few of them attempted to address the annotation and corresponding mapping issues but did it in the context of specific type of MM content such as still image and audio only.

3 Current works bridging the semantic gap in MM

A key contribution in the area of multimedia semantic annotation has been presented in [8]. The less explored issue of machine-generation of semantic descriptions of audiovisual information requires attention to combine standards for multimedia content description with recent advances in semantic web technologies, to develop systems that maximize the potential knowledge which can be mined from large heterogeneous multimedia information sets on the internet. The described idea were confined within the FUSION (Fuel cell Understanding through Semantic inference, ontologies and Nanotechnology) project which attempts to develop an architecture for such combination and describe their innovation into the inference of high-level, domain-specific, semantic descriptions of multimedia content from low-level, automatically-extracted (MPEG-7) features, using ontologies and pre-defined inference rules. Such semantic descriptions enable sophisticated semantic querying of multimedia resources in terms familiar to the user's domain and ensure that the information and knowledge within the multimedia content has a much greater chance of being discovered and exploited by services, agents and applications on the web. Though the architecture is at a very preliminary stage it provides insight on how to mediate the gap between the semantic web's layered architecture and it requires further research to improve its potential for more complex multidimensional video contents.

A generic methodology for XML semantics reuse is based on mapping from XML Schema constructs to the OWL [11] ones that are semantically more appropriate. The previous mapping is complemented with a XML instance metadata to RDF instance metadata mapping. The latter makes possible to take existing XML metadata to the Semantic Web space. There are many attempts to make XML metadata semantics explicit. Usually, they translate it to Semantic Web languages that facilitate the formalization. Some of them just model the XML tree using the RDF primitives while others concentrate on modelling the knowledge implicit in XML languages definitions, i.e. DTDs or the XML Schemas, using web ontology languages. Finally, there are attempts to encode XML semantics integrating RDF into XML documents. However, none of them facilitates an extensive transfer of XML metadata to the Semantic Web in a general and transparent way. Their main problem is that the XML Schema implicit semantics are not made explicit when XML metadata instantiating this schemas is mapped. Therefore, they do not take profit from the XML semantics and produce RDF metadata almost as semantics-blind as the original XML. Alternatively, they capture this semantics but they use additional ad-hoc semantic constructs that produce less transparent metadata. Therefore, Gonzalez [10] have chosen the XML Semantics Reuse methodology that combines a XML Schema to web ontology mapping, called XSD2OWL, with a transparent mapping from XML to RDF, XML2RDF. The ontologies generated by XSD2OWL are used during the XML to RDF mapping in order to generate semantic metadata that makes XML Schema semantics explicit. But the main problem with this mapping is that its XML to RDF translation scheme requires the existing prior mapped OWL ontologies. There are different translation mechanisms: structure-mapping approach [23], model-mapping approach [24]. However, when the objective is semantic metadata that can be easily integrated, it is better to take a more transparent approach. Transparency is achieved in structure-mapping models because they only try to represent the XML metadata structure, i.e. a tree, using RDF. The RDF model is based on the graph so it is easy to model a tree using it.

These mappings [10] have been validated in different ways. First, OWL validators have been used in order to check the resulting OWL ontologies. Moreover, the two mappings have been tested in conjunction. Testing XML instances have been mapped to RDF, guided by the corresponding OWL ontologies from the used XML Schemas, and then back to XML. Then, the original and derived XML instances have been compared using their canonical version in order to correct mapping problems. Examples of both the steps of XSD2OWL and XML2RDF have been provided in [25].

4 Towards a semantic architecture for MM

4.1 Work plan

To achieve the final objective of designing a formal mapping technique of MM metadata to the corresponding semantic models we are currently doing a literature survey. After we are done with the background study we will start modeling the MM metadata of Jewish cultural archives those are defined for use as a repository in the MOSAICA architecture. Using MOSAICA cultural contents as a use case we will further try to generalize the mapping technique for any other MM contents. We plan to conduct our research and development through evolutionary approach.

4.2 Investigation

In an attempt to formulate ideas on how to create a translation mechanism from low-level XML/HTML to high-level semantic concepts in OWL ontologies we concentrated our effort on Jewish Encyclopedia [31] where each page contain plain text, images and video links. One assisting circumstance was that the Jewish Encyclopedia is a well-organized lexical resource that maintains a recognizable lexico-syntactic pattern used for narration. Accordingly, the articles in it could be easily recognized as belonging to several distinct categories, such as articles on personalities, geographic locations, historical periods, events, artifacts, etc. The lexical information in the electronic lexical resources is dense and their lexico-syntactic structure is reasonably uniform, which guarantees efficient and precise machine processing.

In order to build a framework for semantic mapping we initially started handling the plain textual contents and gathering semantics from it. Our plan is to extend this primary framework on textual contents to handle images and videos as well.

4.2.1 Terminology Adopted

Before elaborating on detail of the proposed framework we would like to clarify few terminologies [31] we adopted: we want to distinguish between the intension and the extension of conceptualization[1]. The distinction between the two comes from the early days of the formal logic [27], and expresses two different aspects of the semantics: connotation and denotation. However, for the purpose of discussion in this paper, we shall use the following definitions:

Intension of a concept is its formal definition, which includes all properties that are required of all possible instances of the concept.

Extension of a concept is the set that includes all instances to which the concept applies.

[1] Not to be confused with the intensional and extensional conceptualization introduced in [31].

Accordingly, the intension of the concept “book”, for example, would be “physical object consisting of a number of pages bound together”, while its extension are all existing books.

By analogy, the database theory distinguishes between the intension of a database, which is its definition in the form of a schema, and the extension of a database, which are the actual values in the database [29]. We shall further extend this analogy to any conceptualization, and thus define the intension and the extension of conceptualization as follows:

Intension of conceptualization is the theoretical construct (model) that represents some phenomenon, data or theory, with a set of variables and a set of logical and quantitative relationships between them.

Extension of conceptualization is the set of all statements (assertions) in the domain of discourse to which it is applied.

Accordingly, the intension of conceptualization is actually the conceptual model or “explicit specification”, and is consequently ontology in the sense of the Gruber’s definition.

Thematic content sources on Jewish cultural heritage could be viewed as the extension of the conceptualization of the Jewish cultural heritage – they contain numerous assertions about individual concepts declaring logical and quantitative relationships between them. Nevertheless, these sources do not provide the corresponding intension of that conceptualization, i.e. they do not explicitly specify the underlying conceptual model.

We define distinctive feature as follows:

Distinctive feature is the characteristic shared among some individual items that clearly distinguish them as a group (class) from other items.

According to this definition distinctive feature is actually class belonging. This is not accidental, since assuming existence of an underlying conceptual model in an extension we also assume that the elements is such extension are instances of classes formed by conceptualization. Accordingly, the procedure for extracting distinctive features in fact coincides with the methodological approaches for automatic acquiring of vocabulary from free-text documentation. Such vocabulary extraction is typically the first step in developing formal model “from scratch”, and is an established research field.

4.2.2 Developed Methodology

The methodology we developed to achieve our goal have the following steps:

1. Extraction of the distinctive features from extensions,

2. Formalization of these distinctive features in ontology,
3. Identification of instances that share same distinctive features, and
4. Consolidation of a formal model.

In step1, we applied NLP using the platform named SANDRA [2], which was previously successfully applied to vocabulary extraction for conceptual modeling [30]. Search AND Retrieval Application (SANDRA) was also integrated in the implementation of the MOSAICA[3] methodology. Extraction of semantic information from free text documentation is evidently a well-established research field with significant achievements.

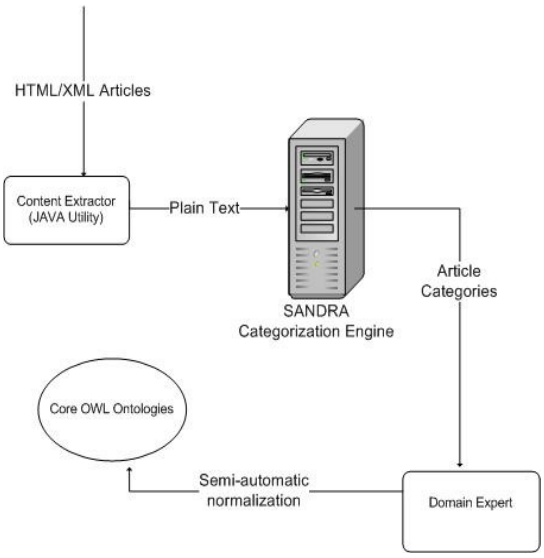

Figure 2: Distinctive Feature Extraction

As SANDRA accepts only plain natural text for categorization purpose we needed to device a utility to extract the plain text from HTML/XML content of the articles and the specially designed *Content Extractor* [figure 1] that converts the articles in the required format acceptable by SANDRA. The output produced by SANDRA is also a plain text file showing extracted categories (distinctive features).

[2] SANDRA is an NLP tool which is contributed by one of MOSAICA consortium members
[3] Our current investigation will contribute to the semantic layer of the MOSAICA application stack which has been described in section 4.3

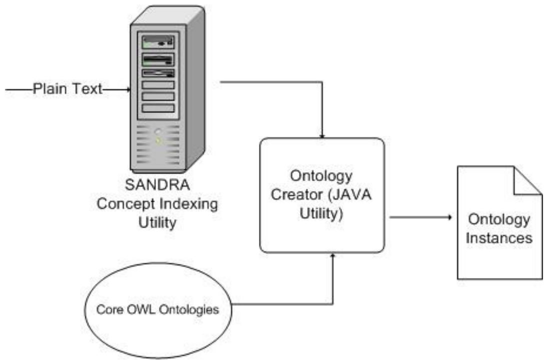

Figure 3: Identification of instances

In step 2, the vocabularies acquired from the encyclopaedia were manually cross-referenced, in order to determine overlaps, and resolve morpho-syntactic variations. For this type of manual analysis we took help from domain experts. This analysis must establish the core set of classes (OWL Core Ontology) meaningful and significant in the underlying domain of discourse as depicted in figure 2.

The extracted core set of classes were used as reference ontology in order to categorizing each article in the encyclopaedia in step 3. We developed an application called *Ontology Creator* that automatically assigned each article under the category defined in the base ontology with the help of SANDRA (using concept based indexing instead of keyword matching) as showed in figure 3. Currently we are working on formalizing the proposed model to adapt it to any kind of content e.g. image, video etc. Also, further effort is underway to improve the core ontology definitions.

4.3 MOSAIA test-bed

MOSAICA has been envisioned as a showcase for demonstrating how already existing digital cultural resources can effectively be put together into a well-defined conceptual framework. To achieve that purpose MOSAICA has been planned to utilize two cutting edge technologies. First, semantic web & ontology engineering and second, distributed content management.

One of the basic objectives of MOSAICA [22] is providing solution for access and accumulation of multimedia content irrespective of its format, availability locations, device and network connections suitable for user requirements. As a result MOSAICA requires that the contents already exist in required format or the platform has the built-in capabilities to dynamically adapt the content on request. To meet such requirements the content management system must identify the characteristics of both content and its usage

context, to decide the kind of adaptation needs to be applied and then should apply adequate adaptation mechanisms. In order to do it efficiently it is necessary to generate, store, convey and use meaningful shared descriptions of the identified characteristics.
The storage and access of the video content requires further research and development effort. Specifically live contents do not have an end point in time without actual time of request for consumption. Each requested unit may contain a number of resources of different media types and their associated descriptions may be available on different networked peers. The MPEG-21 [19] standard seems to be of suitable solution for the declaration of the parts that make up each item and its location. To provide universal and transparent access to multimedia content MOSAICA will incorporate MPEG-21 concepts into a distributed content management system. To describe the characteristics of the content and capabilities of terminals a further adaptation will be needed by combining RDF with MPEG-21. We scope our mapping method to be validated against this adaptation frame.

5 Conclusion and Remarks

In this paper we have outlined our plan to innovate a mapping between MPEG-7/MPEG-21 multi-media contents and corresponding semantic descriptions. In an effort to do that a brief summary of contemporary research work related to ontology-integrated multimedia contents, tools, techniques and architectures have been presented. As part of our implementation strategy a simple overview of agile evolutionary development process has been described that we will be following to conduct our research. From the investigation we carried out we learned very useful lessons of developing ontological frameworks from scratch. Currently the proposed model handles only the plain textual contents and valid for the distinctive features identified in the core ontology only. We are in the process of formalizing it in a way that can easily adapt to plain XML meta-data descriptions of image and video. Finally, we described our test-bed MOSAICA upon which the performance and validity of our method will be tested further.

6 References

[1] Lora Aroyo, Pieter Bellekens, Martin BjÄorkman, Jeen Broekstra, Geert-JanHouben. Ontology-based Personalization in User-Adaptive Systems. Available online from: http://www.doc.ic.ac.uk/~gu1/WPRSIUI/paper7.pdf , accessed on October 2, 2006

[2] ONTOWEDSS - AN ONTOLOGY-BASED ENVIRONMENTAL DECISION-SUPPORT SYSTEM FOR THE MANAGEMENT OF WASTEWATER TREATMENT PLANTS. A PhD thesis by Luigi Ceccaroni. UNIVERSITAT POLIT`ECNICA DE CATALUNYA BARCELONA, SPAIN, OCTOBER 2001

[3] KTWeb Fact Sheet on Semantic Web. November 2003. Available Online at: http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/semanticweb.pdf, accessed on October 10, 2006

[4] Giorgos Stamou, Jacco van Ossenbruggen, Jeff Pan and Guus Schreiber. Multimedia Annotations on the Semantic web. 2006 IEEE. v 13, n 1, January/February, 2006, p 86-90.

[5] Decker, S. Melnik, S. van Harmelen, F. Fensel, D. Klein, M. Broekstra, J. Erdmann, M. Horrocks, I. The Semantic Web: the roles of XML and RDF. Internet Computing IEEE. Sept/Oct 2000.

[6] F. Nack, J. van Ossenbruggen, and L. Hardman, "That Obscure Object of Desire: Multimedia Metadata on the Web (Part II)," *IEEE MultiMedia*, vol. 12, no. 1, Jan.– Mar. 2005, pp. 54-63.

[7] SWBPD MM Task Force Description: Multimedia Annotation on the Semantic Web (MM) http://www.w3.org/2001/sw/BestPractices/MM/, accessed on October 15, 2006

[7] Giovanni Tummerello et al. Facing the hard problem: Automatic RDF annotations from MPEG 7 streams.

[8] Hunter Jane, Little Suzanne. A framework to enable the semantic inferencing and querying of multimedia content. International Journal of Web Engineering and Technology, v 2, n 2-3, 2005, p 264-281

[9] Interoperability support for ontology-based video retrieval applications. Tsinaraki, C. (Lab. of Distributed Multimedia Inf. Syst. & Applications, Tech. Univ. Crete, Greece); Polydoros, P.; Christodoulakis, S. Source: Image and Video Retrieval. Third International Conference, CIVR 2004. Proceedings (Lecture Notes in Comput. Sci. Vol.3115), 2004, p 582-91

[10] Roberto García González. A Semantic Web Approach to Digital Rights Management. PhD thesis Submitted to Department of Technologies, Universitat Pompeu Fabra, Barcelona, Spain. XML Semantics Reuse. Available online from http://rhizomik.net/~roberto/thesis/html/Methodology.html, accessed on October 25, 2006.

[11] B. Liu, A. Gupta and R. Jain "A Live Multimedia Stream Querying System", *2nd Int. Conf. Computer Vision Meets Databases (CVDB)*, June. 2005. Best Student Paper Award.

[12] Jonathan GRUDIN and David BARGERON. Multimedia Annotation: An Unsuccessful

Tool Becomes a Successful Framework. *Microsoft Research, One Microsoft Way, Redmond, Washington, USA*
[13] Yannis Avrithis et al. Using a Multimedia Ontology Infrastructure for Semantic Annotation of multimedia content.
[14] Vangelis Karkaletsis , Georgios Paliouras, and Constantine D. Spyropoulos. A. A Bootstrapping Approach to knowledge acquisition from Multimedia content with Ontology evolution. Available online from: www.iit.demokritos.gr/~paliourg/papers/ICCS2005.pdf, accessed on September 18, 2006
[15] Towards a common multimedia ontology framework. Available online from: http://www.acemedia.org/aceMedia/reference/multimedia_ontology/, accessed on October 15, 2006
[16] A.G. Hauptmann. Towards a large scale concept ontology for broadcast video. In Proc. of the 3rd int conf on Image and Video Retrieval (CIVR'04), 2004.
[17] VERL: an ontology framework for representing and annotating video events. Francois, A.R.J. (Southern California Univ., USA); Nevatia, R.; Hobbs, J.; Bolles, R.C.; Smith, J.R. Source: IEEE Multimedia, v 12, n 4, Oct.-Dec. 2005, p 76-86
[18] FAQs on MPEG-2: Available online from http://www.chiariglione.org/mpeg/faq/mp21-mds/mp21-mds.htm#E12E9 , accessed on October 25, 2006.
[19] L. Chiariglione. The MPEG (Moving Pictures Experts Group) generation – new information-age elements ISO Bulletin, 2000/09. Available online from http://www.chiariglione.org/leonardo/publications/index.htm, accessed on October 27, 2006.
[20] Evolutinary developement: Available online from http://www.agiledata.org/essays/evolutionaryDevelopment.html, accessed on October 25, 2006.
[21] Model based testing in evolutionary software development. Pretschner, A. (Institut fur Informatik, Technische Universitat); Lotzbeyer, H.; Philipps, J. Source: Proceedings of the International Workshop on Rapid System Prototyping, 2001, p 155-160
[22] STREP (Specific Targeted Research or Innovation Project) (IST-034984) - Annex I - "Description of Work". And MOSAICA STREP Proposal.
[23] Klein, M. Interpreting XML Documents via an RDF Schema Ontology. Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA 2002). pp. 889-894, 2002
[24] Tous, R.; GarcÃa, R.; RodrÃguez, E. & Delgado, J.: "Architecture of a Semantic XPath Processor. Application to Digital Rights Management (DRM)".In Bauknecht, K.& Werthner, H. "E-Commerce and Web
[25] MPEG-7Ontots Project. Available online from: http://rhizomik.net/content/mpeg7ontos , accessed on October 17, 2006
[26] Building a visual ontology for video retrieval Hollink, L. (Vrije Universiteit Amsterdam, Netherlands); Worring, M.; Schreiber, Ath. Source: 13th Annual ACM International Conference on Multimedia, 2005, p 479-82
[27] F.L.G. Frege (1892). Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik 100, 25-50.
[28] N. Guarino, P. Giaretta (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. N. J. I. Mars (ed.), Towards Very Large Knowledge Bases, IOS Press 1995, pp. 25-32.
[29] C. Batini, M. Lenzerini, S. B. Navathe (1986). A Comparative Analysis of Methodologies for Database Schema Integration. ACM Computing Surveys, 1986.
[30] D. Dori, N. Korda, A. Soffer, S. Cohen (2004). SMART: System Model Acquisition from Requirements Text, in J. Desel, B. Pernici, and M. Weske (Eds.): BPM 2004, LNCS 3080, pp. 179 194, 2004. Springer Verlag Berlin Heidelberg 2004.
[31] D2.1 Methodology for Aligning Multiple, Distributed Ontologies v.1.0.doc
[32] Jewish Encyclopedia. Available online: http://www.jewishencyclopedia.com/ , accessed on January 21, 2007