Ontological Infrastructure for a Semantic Newspaper

Roberto García, Ferran Perdrix, Rosa Gil

Departament d'Informàtica i Enginyeria Industrial,
Universitat de Lleida,
Jaume II 69, E-25001 Lleida, Spain
roberto@rhizomik.net
{fperdrix, rgil}@diei.udl.es

Abstract. In order to make news management in the journalism domain affordable it is necessary to make machines aware of a greater part of the underlying semantics. However, things can not change from one day to the next. Newspapers have made great investments in their current news management systems and there have been also great standardisation efforts in order to facilitate interoperation. The wish is to undertake a smooth transition. Moreover, it is also necessary to include more multimedia metadata as it is the main added value of digital newspapers.

Our proposal is to build an ontological framework based on journalism (NITF, NewsML and NewsCodes) and multimedia (MPEG-7) standards. These standards are based on XML technologies. Therefore, we have developed a generic XML Schema to OWL mapping. The previous mapping is complemented with an XML metadata instances to RDF mapping that completes a transparent transfer of metadata from the XML to the Semantic Web domain.

Once in a semantic space, data integration and news management and retrieval are facilitated enormously. The resulting ontological framework is now going to be applied in a journalism holding called Segre Media Group[1], which produces press, radio and television content.

1 Introduction

The introduction of information technologies in the news industry has marked a new age in journalistic activity. The digital news industry sector is one of the most eager surroundings for the last technologies consumption. Digital media allows an unprecedented dissemination, ease to access, promptly, cheap, and at last *à la carte* information delivery.

Moreover, digital media eliminate time and space restrictions, and changing editorial team routines. The conceptual architecture of the actual digital news industry involves the whole information lifecycle: beginning with journalists creating the news and ending with users' consumption of the news across several platforms.

[1] http://www.diarisegre.com

Next generation digital news industry is based on the assumption that consumers, actually mainly viewers, will become into participants [1]. This fact implies the need for interactive devices, content adaptation, and management for new distribution channels.

The need for a new architecture and structures for content management is important for the digital news industry. The aim is to improve knowledge management and information retrieval.

Content generation routines, a creative process that started its digitalization a few years ago, involve people and it is very important to track the impact that this change (to digital news) will have on them. Not to mention that the archive system can be used for search and retrieval functions.

The news archival process is done by the documentation department, according to the criteria of experts in this department, and the possibilities (and limitations) of the software platform. They classify news using a hierarchical thesaurus. This thesaurus is constantly growing as archivists are appending or changing its content.

Journalists search this information when they need to inform themselves on subjects, histories or events. This search can be performed in extreme situations, e.g., lack of time, or lack of knowledge in relation to the archive system. This is reflected in the way journalists formulate their queries. The gap between archivists' and journalists' mental models implies that we need a more flexible content categorization and search system. This trend is even bigger when we consider other more general end-users.

A wide margin remains yet for taking advantage of the possibilities offered by the digital medium to exploit a newspaper archive. Aspects that can be improved include:

- keyword search falling short in expressive power;
- weak interrelation between archive items: users may need to combine several indirect queries manually before they can get answers to complex queries;
- lack of a commonly adopted standard representation for sharing archive news across newspapers;
- lack of internal consensus for content description terminology between and among reporters and archivists;
- lack of involvement of reporters in the archiving process.

We believe that emerging Semantic Web technologies [2] provide a good approach to overcome these limitations. The size and complexity of the stored information, and the time limitations for cataloguing, describing and ordering the incoming information, make newspaper archives a relatively disorganised and difficult to manage corpus. In this sense, they share many of the characteristics and problems of the WWW, and therefore the solutions proposed in the Semantic Web vision are pertinent here.

The rest of this paper is organised as follows. Section 2 presents the state of the art of journalism and multimedia metadata, and then analyses the current situation. Section 3 presents the methodology that has been employed in order to produce an ontological framework based on existing standards. Section 4 shows the ontological infrastructure that has been generated, the architecture of the semantics processing

infrastructure and some examples of semantic integration mappings. Finally, there are the conclusions and a sketch of the future plans in section 5.

2 State of the Art

In order to build an ontological infrastructure for the Semantic Newspaper, it is important to analyse the state of the art of the metadata initiatives in the journalism domain. Additionally, digital newspapers have stressed the requirements of multimedia management. Digital news are managed as multimedia packages that integrate text, images, video, audio, etc. Therefore, it is also important to analyse the current situation in the more general multimedia metadata domain. Both aspects are studied next, first journalism metadata and then multimedia metadata. This section ends with an analysis of the current situation, which identifies the keys points that have guided our approach to build the ontological framework.

2.1 Journalism Metadata

One of the main standardization frameworks in the journalism domain is the International Press Telecommunications Council[2] (IPTC), an international consortium of news agencies, editors and newspapers distributors. IPTC has developed standards like the Information Interchange Model[3] (IIM), NewsCodes[4] (formerly the Subject Reference System), the News Industry Text Format[5] (NITF) or NewsML[6].

Currently, almost all of them have evolved towards XML-based standards to represent and manage news along their whole lifecycle, including their creation, exchange and consumption.

For instance, NewsML is used to represent news as multimedia packages and NITF deals with document structure, i.e. paragraphs, headlines, etc. On the other hand, the Subject Reference System (SRS), now part of IPTC NewsCodes, is a subject classification hierarchy with three levels and seventeen categories in its first level.

2.2 Multimedia Metadata

All the previous initiatives are centred on the journalism specific aspects of a semantic newspaper. However, as it has been pointed out, newspapers are evolving towards the digital multimedia domain. Therefore, they stress more and more their multimedia management requirements.

In the multimedia metadata domain, as it is extensively shown in the literature [3,4,5,6], the MPEG-7 [7] standard constitutes the greatest effort for multimedia

[2] IPTC, http://www.iptc.org
[3] IIM, http://www.iptc.org/IIM
[4] NewsCodes, http://www.iptc.org/NewsCodes
[5] NITF, http://www.nitf.org
[6] NewsML, http://www.newsml.org

description. It is divided into four main components: the Description Definition Language (DDL, the basic building blocks for the MPEG-7 metadata language), Audio (the descriptive elements for audio), Visual (those for video) and the Multimedia Description Schemes (MDS, the descriptors for capturing the semantic aspects of multimedia contents, e.g. places, actors, objects, events, etc.).

2.3 Situation Analysis

The main standards that have been presented, both in the journalism and multimedia domains, are based on XML and specified by XML Schemas. The more significant case is the MPEG-7 one. It is based on a set of XML Schemas that define 1182 elements, 417 attributes and 377 complex types. NewsML and NITF are also very big standards, they define more than 100 elements, and the NewsCodes hierarchy of subjects defines more than one thousand different subjects.

The size of these standards makes it very difficult to manage them. Moreover, the use of XML technologies implies that a great part of the semantics remains implicit. Therefore, each time an application is developed, semantics must be extracted from the standard and re-implemented.

For instance, if we use XQuery in order to retrieve MPEG-7 *SegmentType* descriptions from an XML database, we must be aware of the hierarchy of segment types and implement an XQuery that has to cover any kind of multimedia segment, i.e. *VideoSegmentType*, *AnalyticClipType*, *AudiSegmentType*, etc.

Once the hierarchy of segments types is available in Web Ontology Language (OWL) form, semantic queries benefit from the, now, explicit semantics. Therefore, a semantic query for SegmentType will retrieve all subclasses without requiring additional developing efforts.

This is necessary because, although XML Schemas capture some semantics of the domain they model, XML tools are based on syntax. The captured semantics remain implicit from the XML processing tools point of view. Therefore, when an XQuery searches for a *SegmentType*, the XQuery processor has no way to know that there are many other kinds of segment types that can appear in its place, i.e. they are more concrete kinds of segments.

The previous example only illustrates one kind of difficulty derived from the use of just syntax-aware tools. MPEG-7 constitutes a valuable starting point for more specific developments, i.e. it can be seen as an “upper-ontology” for multimedia. However, the lack of explicit semantics makes MPEG-7 very difficult to extend in an independent way, i.e. third party extensions.

The same applies for the journalism standards. Moreover, standards from both worlds share many concepts so it would be possible, and easier, to integrate them once their implicit semantics are available from a computer processing point of view.

Our approach to the inconveniencies observed in this state of the art is presented in the next section.

3 Approach

We have undertaken the application of the Semantic Web proposals to the newspapers world by following a smooth transition strategy [8]. This strategy advises about keeping compatibility (at least initially) with the current technology: browsers, protocols, web and application servers, databases, and architectures.

The objective is then to design a platform that is an extension of previously working systems in mass media companies. The manual creation of semantic instances for news items, at a regular daily pace, is indeed a feasible goal in the future. The introduction of new semantic documentation tools requires, however, a careful work of analysis, design, testing and balancing of the additional burden that such tools may impose on archivists, journalists or end-users.

In order to put into practice the smooth transition strategy, the first step has been to reuse existing standards in the journalism and multimedia fields, which have been for long very active in standardization.

However, as it has been highlighted in the state of the art, all the more recent standards are based on XML but lack formal semantics that facilitate applying a Semantic Web approach. Therefore, in order to facilitate the transition from current standards and applications to the semantic world, we have applied the XML Semantics Reuse methodology, which is detailed in section 3.1.

There are other existing initiatives that try to move journalism and multimedia metadata to the Semantic Web world. In the journalism field, the Neptuno [9] and NEWS [10] projects can be highlighted. Both projects have developed ontologies based on existing standards (IPTC SRS, NITF or NewsML) but from an ad-hoc point of view. Therefore, in order to smooth the transition from the previous legacy systems, complex mappings should be developed and maintained.

The same can be said for the existing attempts to produce semantic multimedia metadata. Chronologically, the first attempts to make MPEG-7 metadata semantics explicit where carried out, during the MPEG-7 standardisation process, by Jane Hunter [11]. The proposal used RDF to formalise a small part of MPEG-7, and later incorporated some DAML+OIL construct to further detail their semantics [12]. However, at that moment, there were not mature technologies for Web-wide metadata semantics formalisation. Moreover, XML had already a great momentum, so it was the logical choice.

From this point, once Semantic Web has matured, there have been more attempts to relate MPEG-7 with Web ontologies. However, none of them has retaken the initial effort to completely move MPEG-7 to the Semantic Web. This initiatives range from separated modules for existing MPEG-7 tools that offer reasoning capabilities for concrete aspects of multimedia management [4], to a partial OWL modelling of the MPEG-7 Multimedia Description Schemes intended to facilitate MPEG-7 extensions [6]. Moreover, they are not systematic; they are applied on an ad-hoc basis, what makes them very costly to apply to the whole MPEG-7 standard.

All these initiatives have produced very interesting results, both in the journalism and multimedia fields, and they are complementary to our objective, i.e. to smoothly move towards a semantic newspaper based on the main journalism and multimedia standards.

The method we have used to perform this is detailed in the next section. It is a generic XML Schema to OWL mapper combined with an XML to RDF translator. It has already shown its usefulness with other quite big XML Schemas in the Digital Rights Management domain, such as MPEG-21 [13] and ODRL [14], and in the music metadata field [15].

3.1 XML Semantics Reuse Methodology

The main caveat of semantic multimedia metadata is that it is sparse and expensive to produce. The previously introduced initiatives are appropriate when applied to limited scopes. However, if we want to increase the availability of semantic multimedia metadata and, in general, of semantic metadata, we need methods that are more productive. The more direct solution is to take profit from the great amount of metadata that has been already produced by the XML community.

There are many attempts to move metadata from the XML domain to the Semantic Web. Some of them just model the XML tree using the RDF primitives [16]. Others concentrate on modelling the knowledge implicit in XML languages definitions, i.e. DTDs or the XML Schemas, using web ontology languages [17,18]. Finally, there are attempts to encode XML semantics integrating RDF into XML documents [19,20].

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. Or, on the other hand, they capture these semantics but they use additional ad-hoc semantic constructs that produce less transparent metadata.

Therefore, we have chosen the XML Semantics Reuse methodology [21] and the XML Schema to OWL and XML to RDF tools implemented in the ReDeFer project[7]. This methodology combines an 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 step in order to generate semantic metadata that makes XML Schema semantics explicit. Both steps are detailed next.

XSD2OWL Mapping

The XML Schema to OWL mapping is responsible for capturing the schema implicit semantics. This semantics are determined by the combination of XML Schema constructs. The mapping is based on translating these constructs to the OWL ones that best capture their semantics. These translations are detailed in Table 1.

The XSD2OWL mapping is quite transparent and captures a great part XML Schema semantics. The same names used for XML constructs are used for OWL ones, although in the new namespace defined for the ontology. XSD and OWL constructs names are identical; this usually produces uppercase-named OWL

[7] ReDeFer project, http://rhizomik.net/redefer

properties because the corresponding element name is uppercase, although this is not the usual convention in OWL.

Therefore, XSD2OWL produces OWL ontologies that make explicit the semantics of the corresponding XML Schemas. The only caveats are the implicit order conveyed by *xsd:sequence* and the exclusivity of *xsd:choice*.

For the first problem, *owl:intersectionOf* does not retain its operands order, there is no clear solution that retains the great level of transparency that has been achieved. The use of RDF Lists might impose order but introduces ad-hoc constructs not present in the original metadata. Moreover, as it has been demonstrated in practise, the element ordering does not contribute much from a semantic point of view. For the second problem, *owl:unionOf* is an inclusive union, the solution is to use the disjointness OWL construct, *owl:disjointWith*, between all union operands in order to make it exclusive.

Table 1. XSD2OWL translations for the XML Schema constructs and shared semantics with OWL constructs

XML Schema	OWL	Shared informal semantics
element | attribute	rdf:Property owl:DatatypeProperty owl:ObjectProperty	Named relation between nodes or nodes and values
element@substitutionGroup	rdfs:subPropertyOf	Relation can appear in place of a more general one
element@type	rdfs:range	The relation range kind
complexType|group |attributeGroup	owl:Class	Relations and contextual restrictions package
complexType//element	owl:Restriction	Contextualised restriction of a relation
extension@base | restriction@base	rdfs:subClassOf	Package concretises the base package
@maxOccurs @minOccurs	owl:maxCardinality owl:minCardinality	Restrict the number of occurrences of a relation
sequence choice	owl:intersectionOf owl:unionOf	Combination of relations in a context

To conclude, one important aspect is that the resulting OWL ontology may be OWL-Full. This is due to the fact that, in some cases, the XSD2OWL translator must employ *rdf:Property* for those *xsd:elements* that have both data type and object type ranges.

XML2RDF Mapping

Once all the metadata XML Schemas are available as mapped OWL ontologies, it is time to map the XML metadata that instantiates them. The intention is to produce RDF metadata as transparently as possible. Therefore, a structure-mapping approach has been selected [16]. It is also possible to take a model-mapping approach [22].

XML model-mapping is based on representing the XML information set using semantic tools. This approach is better when XML metadata is semantically exploited for concrete purposes. 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. Moreover, we do not need to worry about the semantics loose produced by structure-mapping. We have formalised the underlying semantics into the corresponding ontologies and we will attach them to RDF metadata using the instantiation relation *rdf:type*.

The structure-mapping is based on translating XML metadata instances to RDF ones that instantiate the corresponding constructs in OWL. The more basic translation is between relation instances, from *xsd:elements* and *xsd:attributes* to *rdf:Properties*. Concretely, *owl:ObjectProperties* for node to node relations and *owl:DatatypeProperties* for node to values relations.

However, in some cases, it would be necessary to use *rdf:Properties* for *xsd:elements* that have both data type and object type values. Values are kept during the translation as simple types and RDF blank nodes are introduced in the RDF model in order to serve as source and destination for properties. They will remain blank for the moment until they are enriched with semantic information.

The resulting RDF graph model contains all that we can obtain from the XML tree. It is already semantically enriched thanks to the *rdf:type* relation that connects each RDF properties to the *owl:ObjectProperty* or *owl:DatatypeProperty* it instantiates. It can be enriched further if the blank nodes are related to the *owl:Class* that defines the package of properties and associated restrictions they contain, i.e. the corresponding *xsd:complexType*. This semantic decoration of the graph is formalised using *rdf:type* relations from blank nodes to the corresponding OWL classes.

At this point we have obtained a semantics-enabled representation of the input metadata. The instantiation relations can now be used to apply OWL semantics to metadata. Therefore, the semantics derived from further enrichments of the ontologies, e.g. integration links between different ontologies or semantic rules, are automatically propagated to instance metadata thanks to inference.

However, before continuing to the next section, it is important to point out that these mappings have been validated in different ways. First, we have used OWL validators in order to check the resulting ontologies, not just the MPEG-7 Ontology but also many others [13,14]. Second, our MPEG-7 ontology has been compared to Jane Hunter's one [12]. This comparison has shown that our mapping captures the same semantics as those captured by hand, by Jane Hunter, using RDF Schema and DAML+OIL. Finally, 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.

4 Ontological Infrastructure

As a result of applying the XML Semantics Reuse methodology, we have obtained a set of ontologies that reuse the semantics of the underlying standards, as they are formalised through the corresponding XML Schemas. All the ontologies related to journalism standards, i.e. NewsCodes NITF and NewsML, are available from the Semantic Newspaper site[8]. The MPEG-7 Ontology is available from the MPEG-7 Ontology site[9].

The ontologies that are going to be used as the basis for the info-structure of the semantic newspaper are:

- **NewsCodes Subjects Ontology**: an OWL ontology for the subjects' part of the IPTC NewsCodes. It is a simple taxonomy of subjects but it is implemented with OWL in order to facilitate the integration of the subjects' taxonomy in the global ontological framework.
- **NITF 3.3 Ontology**: an OWL ontology that captures the semantics of the XML Schema specification of the NITF standard. It contains some classes and many properties dealing with document structure, i.e. paragraphs, subheadlines, etc., but also some metadata properties about copyright, authorship, issue dates, etc.
- **NewsML 1.2 Ontology**: the OWL ontology resulting from mapping the NewsML 1.2 XML Schema. Basically, it includes a set of properties useful to define the news structure as a multimedia package, i.e. news envelope, components, items, etc.
- **MPEG-7 Ontology**: The XSD2OWL mapping has been applied to the MPEG-7 XML Schemas producing an ontology that has 2372 classes and 975 properties, which are targeted towards describing multimedia at all detail levels, from content based descriptors to semantic ones.

4.1 System architecture

Based on the previous XML world to Semantic Web domain mappings, we have built up a system architecture that facilitates journalism and multimedia metadata integration and retrieval. The architecture is sketched in Fig. 1. The MPEG-7 OWL ontology, generated by XSD2OWL, constitutes the basic ontological framework for semantic multimedia metadata integration and appears at the centre of the architecture. In parallel, there are the journalism ontologies. The multimedia related concepts from the journalism ontologies are connected to the MPEG-7 ontology, which acts as an upper ontology for multimedia. Other ontologies and XML Schemas can also be easily incorporated using the XSD2OWL module.

Semantic metadata can be directly fed into the system together with XML metadata, which is made semantic using the XML2RDF module. For instance, XML MPEG-7 metadata has a great importance because it is commonly used for low-level visual and audio content descriptors automatically extracted from its underlying

[8] Semantic Newspaper, http://rhizomik.net/semanticnewspaper
[9] MPEG-7 Ontology, http://rhizomik.net/mpeg7ontos

signals. This kind of metadata can be used as the basis for audio and video description and retrieval.

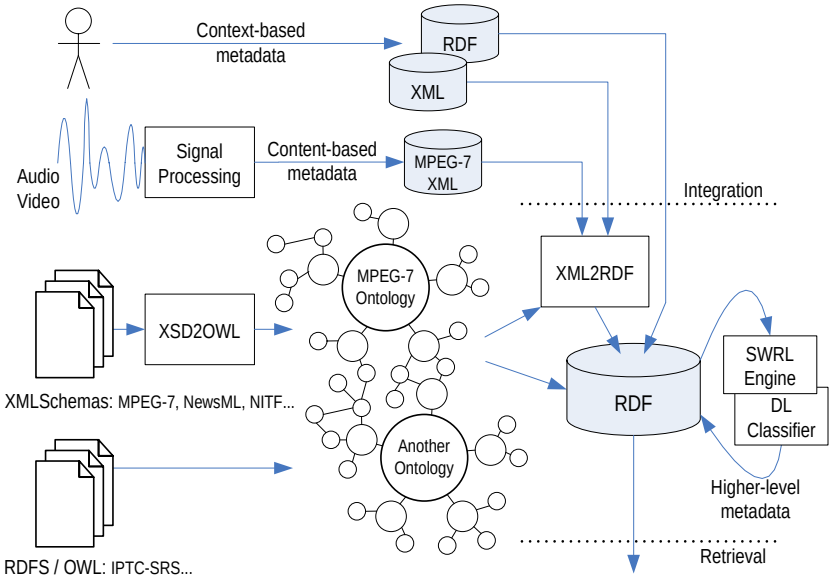

Fig. 1. Metadata integration and retrieval architecture

In addition to content-based metadata, there is context-based metadata. This kind of metadata higher level and it usually, in this context, related to journalism metadata. It is generated by the system users (journalist, photographers, cameramen, etc.). For instance, there are issue dates, news subjects, titles, authors, etc.

This kind of metadata can come directly from semantic sources but, usually, it is going to come from legacy XML sources based on the standards' XML Schemas. Therefore, in order to integrate them, they will pass through the XML2RDF component. This component, in conjunction with the ontologies previously mapped from the corresponding XML Schemas, generates the RDF metadata that can be then integrated in the common RDF framework.

This framework has the persistence support of a RDF store, where metadata and ontologies reside. Once all metadata has been put together, the semantic integration can take place, as it is exemplified in section 4.2.

4.2 Semantic Integration Outline

As mentioned in the introduction, one of the main problems in the Wold Wide Web and intraweb domains is that of heterogeneous data integration. Even within a single organization, data from disparate sources must be integrated. Our approach to solve this problem is based on Web ontologies and, as the focus is on multimedia and journalism metadata integration, our base ontology is the MPEG-7 and the journalism ontologies.

In order to take profit from the system architecture presented before, when semantic metadata based on different schemes has to be integrated, the XML Schemas are first mapped to OWL. Once this first step has been done, these schemas are easily integrated into the ontological framework using OWL semantic relations for equivalence and inclusion: *subClassOf*, *subPropertyOf*, *equivalentClass*, *equivalentProperty*, *sameIndividualAs*, etc.

These relationships capture the semantics of the data integration. Then, once metadata is incorporated into the system and semantically-decorated, the integration is automatically performed by applying inference. Table 2 shows some of these semantics mappings, performed once all metadata has been moved to the semantic space, plus the preliminary mappings performed during the XML to RDF mapping. The later are necessary in order to recognise implicit identifier, i.e. attributes used to identify instances that are not explicitly used as identifiers in XML.

Table 2. Journalism and multimedia metadata integration mapping examples

Semantic Mappings
∀ nitf:tobject.subject . subj:Subject
nitf:tobject.subject.detail ≡ subj:explanation
nitf:body ⊆ newsml:DataContent
newsml:Subject ≡ subj:Subject
XML2RDF Mappings
tobject.subject.refnum → rdf:ID

5 Conclusions and future work

This research work has been guided by the need for a semantic journalism and multimedia metadata framework that facilitates semantic newspaper applications development. It has been detected, as it is widely documented in the bibliography and professional activity, that IPTC standards and MPEG-7 are the best sources for a ontological framework that facilitates a smooth transition from legacy to semantic web era systems. MPEG-7 and most of the IPTC standards are based on XML Schemas and thus they do not have formal semantics.

Our approach contributes a complete and automatic mapping of the whole MPEG-7 standard to OWL and from the main IPTC standard schemas (NITF, NewsML and NewsCodes) to the corresponding OWL ontologies. The resulting set of ontologies is being aligned in order to facilitate the integration of semantic instance metadata. Instance metadata is automatically imported from legacy systems through a XML2RDF mapping, based on the ontologies previously mapped from the standard XML schemas. Once in a semantic space, data integration, which is a crucial factor when several sources of information are available, is facilitated enormously.

The future plans are now centred on testing this ontological framework in the professional context of the Segre Media Group. This media group includes television, press and radio. Therefore, it is going to be possible to test this framework from the journalism and multimedia points of view. Moreover, a semantic system has been already developed in this media group in order to make its newspaper archive semantic [9].

References

1. McDonald, N.: Can HCI shape the future of mass communications? Interactions, Vol. 11 Issue 2, pp. 44-47, 2004
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American, 2001
3. Doerr, M.; Hunter, J. & Lagoze, C.: "Towards a Core Ontology for Information Integration". Journal of Digital Information, Vol. 4 Issue 1, April 2003
4. Troncy, R.: "Integrating Structure and Semantics into Audio-visual Documents". Proceedings of the 2nd International Semantic Web Conference, Florida, USA, 2003
5. Hunter, J.: "Enhacing the Semantic Interoperability of Multimedia through a Core Ontology". IEEE Trans. on Circuits and Systems for Video Technology, February 2003
6. Tsinaraki, C.; Polydoros, P. & Christodoulakis S.: "Integration of OWL ontologies in MPEG-7 and TVAnytime compliant Semantic Indexing". Proc. 16th International Conference on Advanced Information Systems Engineering, 2004
7. Salembier, P. & Smith, J.: "Overview of MPEG-7 multimedia description schemes and schema tools". In Manjunath, B.S.; Salembier, P. & Sikora, T. (ed.) "Introduction to MPEG-7: Multimedia Content Description Interface". John Wiley & Sons, 2002
8. Haustein, S., Pleumann, J. Is Participation in the Semantic Web too Difficult? Proceedings of the International Semantic Web Conference. Sardinia, Italy, 2002
9. Castells, P.; Perdrix, F.; Pulido, E.; Rico, M.; Benjamins, R.; Contreras, J. & Lorés, J.: "Neptuno: Semantic Web Technologies for a Digital Newspaper Archive". Springer, LNCS Vol. 3053, pp. 445-458, 2004
10. Fernández-García, N. and Sánchez-Fernández, L.: "Building an Ontology for NEWS Applications". Poster at the International Semantic Web Conference, 2004
11. Hunter, J.: "A Proposal for an MPEG-7 Description Definition Language", MPEG-7 AHG Test and Evaluation Meeting, Lancaster, 15-19 Feb 1999
12. Hunter, J.: "Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology", International Semantic Web Working Symposium (SWWS), 2001
13. Gil, R.; García, R. & Jaime Delgado: "An interoperable framework for IPR using web ontologies". LOAIT'05. IAAIL Workshop Series, Wolf Legal Pub., pp. 135-148, 2005
14. García, R.; Gil, R.; Gallego, I. & Jaime Delgado: "Formalising ODRL Semantics using Web Ontologies". Open Digital Rights Language Workshop, ODRL'2005
15. García, R. & Celma, O.: "Semantic Integration and Retrieval of Multimedia Metadata". Knowledge Mark-up and Semantic Annotation Workshop, Semannot'05. CEUR, 2005
16. Klein, M.C.A.: "Interpreting XML Documents via an RDF Schema Ontology". Proc. 13th Int. Workshop on Database and Expert Systems Applications, pp. 889-894, 2002
17. Amann, B.; Beeri, C.; Fundulaki, I. & Scholl, M.: "Ontology-Based Integration of XML Web Resources". Proc. 1st International Semantic Web Conference, pp. 117-131, 2002
18. Cruz, I.; Xiao, H. & Hsu, F.: "An Ontology-based Framework for XML Semantic Integration". 8th Int. Database Engineering and Applications Symposium, Portugal, 2004
19. Lakshmanan, L. & Sadri, F.: "Interoperability on XML Data". Proc. 2nd International Semantic Web Conference, 2003
20. Patel-Schneider, P.F. & Simeon, J.: "The Yin/Yang web: XML syntax and RDF semantics". Proc. 11th World Wide Web Conference, pp. 443-453, 2002
21. García, R.: XML Semantics Reuse. Chapter 7, A Semantic Web Approach to Digital Rights Management, PhD Thesis (2006) http://rhizomik.net/~roberto/thesis
22. Tous, R.; García, R.; Rodríguez, E. & Delgado, J.: "Arquitecture of a Semantic XPath Processor". In Bauknecht, K.; Pröll, B. & Werthner, H. (ed.): "E-Commerce and Web Technologies, EC-Web 2005". Springer, LNCS Vol. 3590, pp. 1-10, 2005