

Using Several Ontologies for Describing Audio-Visual Documents: A Case Study in the Medical Domain

Antoine Isaac^{1,2} and Raphaël Troncy³

¹ Institut National de l'Audiovisuel, Direction de la Recherche
4, Av. de l'Europe - 94366 Bry-sur-Marne, France
aisaac@ina.fr

² Université de Paris-Sorbonne, LaLICC

³ ISTI-CNR, Via G. Moruzzi 1, 56124 Pisa, Italy
raphael.troncy@isti.cnr.it

Abstract. Knowledge-driven audio-visual content description techniques bring promising solutions for improving the processing of video documents. In this paper, we present an experiment based on real use cases where Semantic Web technologies are used to describe deeply the content of a corpus of TV programs. We show that the combination of several formal ontologies and rules allow to better describe and retrieve the audio-visual sequences.

1 Introduction: Settling the Experiment

While digital video documents are more and more available on the Web, their effective processing is still problematic. In particular, retrieving relevant sequences broadcasted on TV on the basis of content criteria, or reusing some material from a given corpus inside a new document production is still a very difficult task. For such a purpose, knowledge driven audio-visual content processing techniques bring promising solutions. Hence, the Semantic Web community has recently standardized languages for representing ontologies [8] and annotations [9], and provided several tools for querying and performing reasoning on knowledge bases. We have already shown how multimedia content can benefit from these technologies, as well as pointed out some of the critical issue for such a success [12, 7]. In this paper, we detail an experiment based on real use cases where these technologies are used to describe deeply the content of a corpus of TV programs. We show that the combination of several formal ontologies and rules allow to better describe and retrieve the audio-visual sequences.

Our experiment focuses on medicine-related TV documentaries. From INA⁴ funds, we have selected about 30 documents, mostly health magazine programs, during almost one hour each. Amongst those, we extracted nearly half of them

⁴ The *French National Institute of Audio-visual* (INA) has been archiving and indexing the TV and radio programs broadcasted in France for thirty years.

which were linked to heart or heart surgery theme. We thus have a rather homogeneous collection to describe, which eases the search for a convenient theme-related ontology. Indeed, since the medicine theme has brought attention to numerous research works these last years, plenty of ontological resources in that field are available. Moreover, as these programs aim to be broadcasted on the TV, they are also good examples of how AV features are used to popularize complex scientific notions.

The applications that use audio-visual (AV) documents may be interested in different aspects. They have their own viewpoint on this complex media and usually they are just concerned with selected pieces of information corresponding to their needs. An institute like INA has to collect and describe an audio-visual cultural heritage. It is thus interested in both the form and the content of the documents, with a strong emphasis on a documentary archive viewpoint. Therefore, the descriptions should merge AV-oriented parts (*e.g.* stating that a document includes specific sequences like interviews, or techniques like animations) and notions found in a given scientific field. For example, knowing that a sequence shows several close-ups of a surgery action makes it valuable as a piece of educational material since this sequence, if thoroughly described, could be re-used in another production. Our experiment follows these guidelines established by the documentalists of INA, making hence a real use case of how semantic web technologies and traditional archivist practices can be used together.

In the next section, we detail the ontological resources necessary for describing our corpus of videos. In section 3, we show how these ontologies are used to design annotation patterns of both the structure and the content of the documents. In section 4, we discuss the kind of reasoning we are able to perform according to the ontology representation language and the use of additional rules. In section 5, we report some related work on knowledge driven multimedia content description. Finally, we give our conclusions and outline future work in section 6.

2 Ontological Resources

We have already shown in [12] the benefits from articulating an AV material-dedicated ontology with theme-specific ontologies so as to produce descriptions that would really fit specific documentary applications. Such a method allows to capitalize on our experience in AV description: the knowledge is modularized and can be easily adapted from one application to another. We have therefore proposed in [7] an audio-visual description core ontology useful for a wide range of applications using AV material.

This ontology mainly focuses on characterizing documentary elements: the main concept is the *AV production object*, which represents the core notion of an AV document. The first distinction occurs between a *program*, (a rather stand-alone entity from the points of view of production and broadcast), and a *sequence* (a part of a program or other sequences). These concepts are then specialized by means of form or content-linked differentiating features in order to obtain

the classification scheme that is common to all needs: their *genre*. For example, programs are divided into *heterogeneous* and *homogeneous*: the first is characterized by a sequence of autonomous elements in form and in content, unlike the second. They are then classified according to their length, and to their general content (fiction, informative, entertainment). After some further specialization, one can find the usual TV genres: *sitcom*, *tv show*, *documentary*, etc.

The notions used to characterize the AV objects are also defined in the ontology. First, we have introduced a hierarchy of the roles that people can play in a program, whether as authors (*producer*, *director*) mentioned because of their importance in the program production, or participants (*host*, *actor*), being part of the description since they are visible or audible in the document. Then, we can find a large set of AV properties that mirror a given production or broadcast preoccupation or mode. They are organized according to their belonging to the production world (way of filming, such as *camera motion*, editing or post-producing, such as *text insertion*) or to the broadcasting one (*periodicity*, *intended audience*, etc.). A typology of the general *themes* that a document can refer to completes the ontology. We can thus state that *program* are *broadcasted* at some *broadcastTime*, that there are people that play different production roles in the production process, etc.

We mention that this ontology has been built according to methodological principles, as explained in [1, 7]. The concepts are linked to upper-level ontological patterns in order to increase the re-usability of the *core* notions. Such a design enables to extend quite easily the AV ontology to fit additional application needs. For example, we have added for this experiment some basic document-content relations which denote interpretative statements regarding the way thematic content is presented by documents: *clarifies*, *exemplifies*, *demonstrates*, etc. Finally, we propose conceptual relations to link the AV objects to external themes, which allows the description of *content*.

Considering the chosen corpus of video, we have explored some of the existing medical terminologies. One of them is the MENELAS ontology which describes the domain of coronary pathologies [14]. In particular, this ontology contains a lot of concepts dealing with heart surgery, the theme of our corpus. The general medical concepts of this ontology overlap also some of the knowledge heavily formalized in other famous medical ontologies such as GALEN⁵ (*General Architecture for Language and Nomenclatures*), a system dedicated to the development of ontology in all medical domains including surgical procedures. Rather than trying to strictly align these two ontologies, we have put some equivalent classes definitions when necessary. In the same way, the articulation between the AV ontology and the thematic-specific ontology is managed with *equivalent class* definition (e.g. `av:person` and `menelas:human_being`). In the next section, we show how these ontological resources are used to describe the corpus of videos.

⁵ Ontology GALEN. <http://www.opengalen.org>, 2001.

3 Annotating the Videos

3.1 Annotation Mechanism

Describing audio-visual document amounts to consider documentary aspects (*i.e.* identify the elements that constitute the logical structure of the document) as well as thematic aspects (*i.e.* state that these document elements are *about* something). Distinguishing the AV ontology from other theme-specific ontologies allow us to consider both aspects.

Concepts and relations from the AV ontology are manually introduced in the description in order to specify the document-content links at the knowledge level. Afterwards, this conceptual description can be easily linked to strictly documentary metadata, expressed in a language such as MPEG-7. This step can be done using the layered architecture described in [12], or the whole description can be directly represented in our audio-visual description language proposal that supports both ontological and documentary considerations [13]. The documentary items are therefore described as resources, classified under AV concepts. However, in this experiment, we were less interested in the structure validation facilities than in a comprehensive help concerning the way the concepts have to be used in the query and annotation process. We have thus preferred to provide the potential users with some description relational patterns rather than pure documentary structures as explained in the next section.

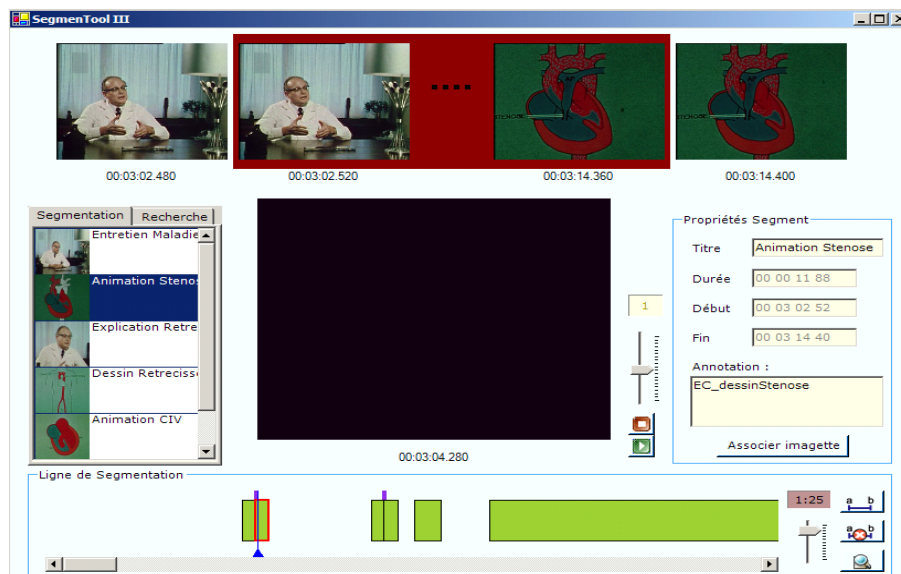


Fig. 1. Video segmentation with SegmenTool

Finally, we have used **SegmenTool**⁶ to decompose temporally the audio-visual documents and to create the audio-visual specific part of the descriptions (Figure 1). The description is then completed by hand by human indexers.

3.2 Conceptual Annotation

In order to ease the description process and make its results more coherent, we use a relational *pattern* giving advice on the way concepts and relations are to be used. For our experiment, we have adapted the one proposed in [7]. As hinted in the Figure 2, one has to describe AV elements by assigning them AV parameter values (*i.e.* the way they are produced) and decomposing them from a documentary standpoint. Their content has also to be indexed by asserting relationships with domain concepts, whether strictly representational – what is shown in the videos – or more interpretation-flavored ones – what is the use of such representations.

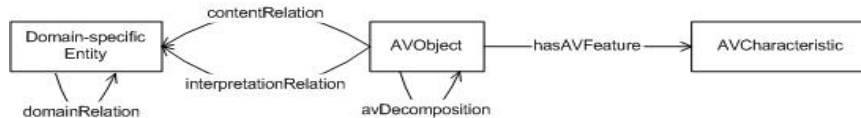


Fig. 2. The description relational pattern

This very simple structure can lead to very rich descriptions, since it brings some sort of recursiveness, as introduced by domain and AV decomposition relations. The example of the Figure 3 gives an idea of the descriptions we are looking for. We have highlighted here the distinction between the two kind of knowledge we are interested in.

4 Query and Perform Reasoning on the Knowledge Base

The aim of our experiment is also to demonstrate the interest of using inference in a content retrieval scenario. Explicit facts in descriptions such as the one we have previously shown can be augmented with derived assertions, as illustrated in Figure 4. For instance, if a sequence has, as a subsequence, another one that explains a specific aspect, we must deduce that it also explains this aspect. We can therefore retrieve audio-visual objects that refer to many themes even if they were only explicitly mentioned in the items they contain. Here, the system will find the documentary described in figure 3 relevant for the query “retrieve the programs that explain a disease and show visually one of its causes”.

⁶ SegmenTool is developed by the DCA team of INA and has been funded by the CHAPERON project under a PRIAMM grant from the French Ministry of Industry.

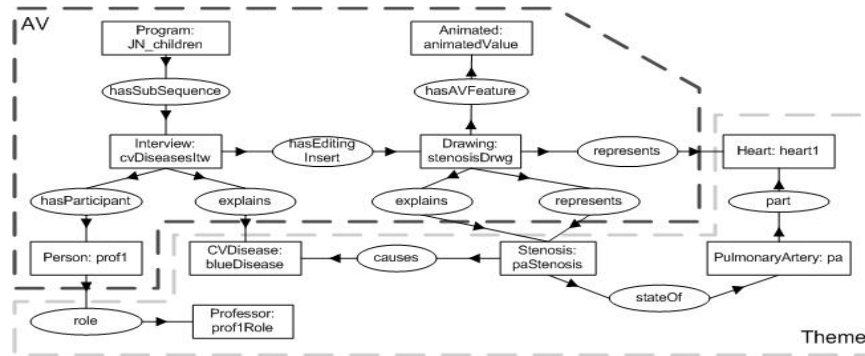


Fig. 3. An example of description specializing the description pattern. Graphical notation for concepts and instances is `Concept:instance`.

In order to benefit from the real power of semantic web technologies, we have to explore the way such reasoning knowledge can be encoded and articulated with the knowledge base. In the following, we show what kind of reasoning we can perform, depending on the specific languages we turn to. We use the **Sesame** architecture [2] for storing and querying the ontologies and the statements. Currently, this architecture supports RDF Schema as ontology specification language and offers the reasoning services specified in the RDF Model Theory [9]. That enables basic inferences, like following specialization paths for concepts and relations. In the example we present, it is therefore possible to find the *Interview* we described in the results of a query for *Sequences* that explain the blue disease.

However, it is not sufficient for the precise concept-and-relation index exploitation we aim at. The opportunity to use the full expressivity of OWL languages [8] – or at least the decidable OWL-DL subset – was thus appealing. With OWL, one can specify that an *ExpertInterview* is exactly defined as an interview where *some* participant plays an expert role. The *ExpertRole* concept will be defined by the means of a logical class equivalence enumerating the theme-specific roles⁷ that can be labelled as expert ones. Thus, we can encode the reasoning knowledge grounding the inferences shown in shaded plain characters in the Figure 4. To implement such inferences, one can turn to specific OWL reasoners, like **BOR** [10], which has been integrated with Sesame. Actually, the BOR reasoner implements the semantics of DAML+OIL, which is very close to what could be achieved with OWL inference engines.

Yet this solution does not completely match our needs. As we focus on rich relational indexes, we would like to deal with reach relational reasoning knowledge, too. OWL-DL allows to specify algebraic properties for relations, which is obvi-

⁷ In our specific case, we picked from the *role* concept the specializations *academic*, *professional* and *hospital*, excepted the role of the hospital institution itself, which was dealt with a OWL *complementOf* expression.

ously useful. However, it lacks the ability to encode more general compositional knowledge. For example, we desperately need to assert rules like $\text{hasSubSequence}(x,y) \cap \text{represents}(y,z) \Rightarrow \text{represents}(x,z)$ in order to infer the assertions shown in dotted shaded lines in the Figure 4.

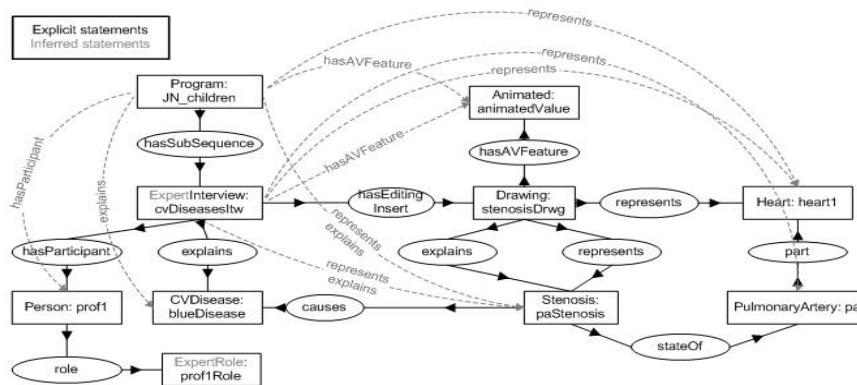


Fig. 4. Index augmented with inferred assertions

This kind of concerns is growing in the Semantic Web community, and begins to be addressed with dedicated languages and tools. Amongst them, SWRL [6] brings an improvement towards the gathering of OWL language family with logical rules. Some of the considerations brought there can be implemented via decidable logical grounding like OWL-DLP [4] which restricts OWL-DL while enabling to turn to some logic programs features. Losing part of OWL-DL expressivity – especially the existential restrictions in necessary conditions – may reveal harmful, but the richness of possible relational rules makes us clearly prefer this choice. In Sesame, such an opportunity is implemented in a *custom inference* module, where the RDFS axioms and rules are augmented with OWL-DLP ones and, further, with ontology-specific rules. This is a rather static way of proceeding – rules are encoded at the level of the inferencer specification, not in the ontology itself – but it is at least a workable one. Table 1 summarizes finally the number of triples (explicit and inferred) contained in the Sesame repository of our experiment. The saturation of the knowledge base is obtained by using the OWL-DLP rules completed by about 20 application-specific rules (mainly compositional).

5 Related Work

Part of the work presented here has already been tested during the French OPALES Project⁸ which provides a general framework for manually describing the

⁸ <http://opales.ina.fr/public/>

content of educational AV documents using conceptual graphs, and for searching amongst these descriptions using an appropriate inference engine. However, the ability to explicitly refer to several ontologies to produce descriptions was especially missing in this architecture. The experiment proposed here is then an evolution of some of these ideas towards Semantic Web related languages and tools, thus benefiting from the extensive research effort lead in this area. The annotations are expressed in RDF, the ontologies are modeled in OWL(DL), and all these resources can then be distributed and reused in other applications.

	Explicit triples	Inferred triples	All triples
RDF Model			129
AV Ontology	5231	10810	16041
Menelas Ontology	10534	26637	37171
Instances	276	1507	1783
Total	16041	38954	54995

Table 1. Number of triples (explicit and inferred) in the *Sesame* repository

There is relatively little work on semantic annotation of multimedia documents. One of the few examples is the work of [5] that shows that linking a number of diverging thesauri to an annotation application for images of paintings can improve both the semantic annotation process for human annotators and the search process. We share with this work the possibility to use various ontological resources. The prototype called *Vannotea* has been developed for enabling the collaborative indexing, annotation and discussion of audio-visual content over high bandwidth networks [11]. However, this tool focuses on the description of the documentary elements composing a document, representing this structure in MPEG-7. The annotation of the content remains mainly in free text and the inferences on the knowledge base are quite limited with respect to the ones detailed in this paper. Finally, the MIAKT (*Medical Imaging and Advanced Knowledge Technologies*) Project⁹ aimed to apply knowledge representation and intelligent analysis techniques to collaborative medical problem solving in the domain of breast cancer screening and diagnosis [3]. It focuses on static medical images which restricts the range of possible descriptions: as such, they propose a rather static description frame, which largely differs from our flexible, possibly recursive, indexing pattern for audio-visual objects.

6 Conclusion and Future Work

We have presented here an experimentation that consists in describing AV documents using Semantic Web languages and tools. We have explored the way audio-visual and domain-related ontologies can be articulated so as to obtain relevant descriptions, and investigated how reasoning knowledge can be encoded and used to make really helpful systems. It is important to note that the observations made here are based on realistic use-cases. We have now to conduct a

⁹ <http://www.aktors.org/miakt/>

complete evaluation of the system, but the first results show already that it is perfectly feasible – and even desirable – to use the semantic web technologies for describing audio-visual documents, keeping in mind that a trade-off between expressivity and workable computation has to be fixed according to the needs of the application targeted.

References

1. B. Bachimont, A. Isaac, and R. Troncy. Semantic Commitment for Designing Ontologies: A Proposal. In *Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, LNAI 2473, p. 114-121, Sigüenza, Spain, 2002.
2. J. Broekstra, A. Kampman, and F. van Harmelen. Sesame: a Generic Architecture for Storing and Querying RDF and RDF Schema. In *Proc. of the 1st International Semantic Web Conference (ISWC'02)*, LNCS 2342, p. 54-68, Sardinia, Italia, 2002.
3. S. Dasmahapatra S., D. Dupplaw, B. Hu, H. Lewis, P. Lewis and N. Shadbolt. Facilitating multi-disciplinary knowledge-based support for breast cancer screening. *International Journal of Healthcare Technology and Management*, 2004.
4. B. N. Grosf, I. Horrocks, R. Volz and S. Decker. Description Logic Programs: Combining Logic Programs with Description Logic. In *Proc. of the 12th International World Wide Web Conference (WWW'03)*, p. 48-57, Budapest, Hungary, 2003.
5. L. Hollink, G. Schreiber, J. Wielemaker and B. Wielinga. Semantic Annotation of Image Collections. In *Workshop on Knowledge Markup and Semantic Annotation*, Sanibel Island, Florida, USA, 2003.
6. I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. N. Grosf and M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML W3C Member Submission, 21 may 2004. <http://www.w3.org/Submission/SWRL/>
7. A. Isaac and R. Troncy. Designing and Using an Audio-Visual Description Core Ontology. In *Workshop on Core Ontologies in Ontology Engineering*, CEUR-WS Vol. 118, Whittlebury Hall, Northamptonshire, UK, 2004.
8. OWL, Web Ontology Language Reference Version 1.0. W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/owl-ref/>
9. RDF, Ressource Description Framework Primer W3C Recommendation, 10 February 2004. <http://www.w3.org/TR/rdf-primer/>
10. K. Simov, and S. Jordanov. BOR: a Pragmatic DAML+OIL Reasoner. Deliverable 40, On-To-Knowledge Project, 2002.
11. R. Schroeter, J. Hunter and D. Kosovic. Vannotea - A Collaborative Video Indexing, Annotation and Discussion System For Broadband Networks. In *Workshop on Knowledge Markup and Semantic Annotation*, Sanibel Island, Florida, USA, 2003.
12. R. Troncy. Integrating Structure and Semantics into Audio-visual Documents. In *Proc. of the 2nd International Semantic Web Conference (ISWC'03)*, LNCS 2870, p. 566-581, Sanibel Island, Florida, USA, 2003.
13. R. Troncy and J. Carrive. A Reduced Yet Extensible Audio-Visual Description Language: How to Escape From the MPEG-7 Bottleneck. In *Proc. of the 4th ACM Symposium on Document Engineering*, Milwaukee, Wisconsin, USA, 2004.
14. P. Zweigenbaum and MENELAS Consortium. MENELAS: An access system for medical records using natural language. *Computer Methods and Programs in Biomedicine*, 45:117-120, 1994.