Semantic multimedia search: the case of SMIL documents

CHKIWA Mounira*, JEDIDI Anis*

** Université de Sfax, MIRACL Multimedia, InfoRmation systems and Advanced Computing Laboratory, B.P 242 Sakiet Ezziet 3021 Sfax, Tunisie*

m.chkiwa@gmail.com, jedidianis@gmail.com

Abstract. Since the first implementations of its principles, the Semantic Web presented a field of free work to ensure its integration and adaptation to the various domains of research. The application of Semantic Web technologies into the process of search in a collection of SMIL documents appears a promising initiative seen the evolution of this language through its various versions. In this paper we propose a semantic search tool in a collection of SMIL documents; this tool adopts a procedure composed of three modules: description, interrogation and representation of the results. We employ for the first module metadata commonly used to annotate information semantically, and for the second we solicit languages of Semantic Web such as RDF, OWL and SPARQL and seen the importance of collaboration of ontologies in the semantic description of multimedia resources, we also present the technique of concepts connection allowing to extend an initial ontology.

Key words: Semantic Web, SMIL document, ontology, metadata, SPARQL, RDF.

1 Introduction

Faced with a variety of types of multimedia resources universally existing in the web, the importance of a SMIL document is that it provides a new vision more structured, organized and controlled of diversified multimedia contents. A SMIL document could enhance a classic multimedia presentation by integrating tags and attributes that provide a recommended overlap of resources in order to facilitate the understanding of a specific idea. Seen the valued prospect of this multimedia language and the spread of Semantic Web technologies come the need to generate a semantic tool to search multimedia contents in a collection of SMIL documents. By the contribution of its theoretical principles and standardized technologies, the Semantic Web offers a new vision for the search operation considering the meaning of things more than

its syntactic shapes. The reminder of this paper is as follows; in the three first sections we present the SMIL language and the Semantic Web framework. Section 5 describes the related works and section 6 presents our contribution in this topic followed by the technical view of the developed tool and its functionalities in section 7. Finally, section 8 underlines some conclusions and future research lines.

2 The SMIL multimedia presentation

Primarily, multimedia is everything dealing with the combination of two or more of the following media: image, sound, text and video. The presentation of multimedia contents is based on three fundamental axes:

- Time axis which defines the temporal ordering and synchronization of different objects by a script or scenario time predefined.
- Spatial axis which defines the spatial distribution of different media (with the exception of audio component).
- Logical axis considering hierarchical decomposition of a multimedia document into parts and subparts, with one or more media for each part.

SMIL (Synchronized Multimedia Integration Language) is a W3C specification which allows creating structured multimedia presentations. Another axis is considered in SMIL document: hypermedia axis; but this depend on the fact that such document offer or not a way to interact. This axis offers to the user the possibility to control the temporal, spatial and logical dimension making a personalized execution of a document according to the user preferences. The scope covered by the SMIL language is above websites and offers a range of possibilities such as:

- Collecting in a single presentation contents may come from different servers.
- Creating multimedia documents with very small size unlike the conventional multimedia presentations thanks to its simple textual structure.
- Insert controls events (play, stop, go to ...) to create customized presentations based on user interaction which allows many ways to present the same document.

A SMIL document is structured in two main parts: head and body; figure1 shows more details about these parts:

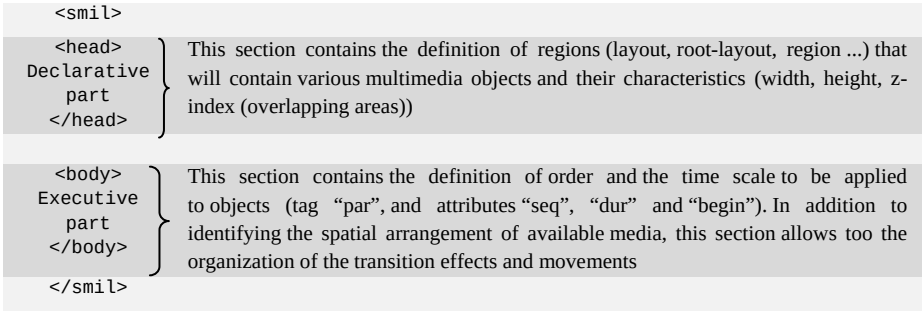

Fig.1. Structure of a SMIL document

3 Semantic Web

Accessible resources (images, text, audio, video ...) are at an earlier stage, formed by a set of documents, formatted in specific languages. These languages allow expressing the links between an object in the source document and another in a destination document. The Semantic Web is operated by software agents (browsers and search engines) browsing the links encountered. Metadata is the semantic descriptions of linked web contents; it's the global concept of Semantic Web which aims to yield semantic annotations to items accessible on the web even it is not a resource (i.e. image, text …); it can be persons or associations… The overall vision of Semantic Web could be summarized in three fundamental points:

- Identifying resources universally (URI: Uniform Resource Identifier): We use URIs to identify pieces of information across the Web. The URI includes the "Uniform Resource Locator" (URL), the "digital Object Identifier (DOI) and the "International Standard Book Number" (ISBN).
- Describing the relationship between resources (RDF [1]: Resource Description Framework). It is a model for describing data on the web making automatic the access to sense of contents available on the web. Development of RDF has been motivated for several perspectives such as handling and defining semantic relationships between data (unlike primitive source/destination relationship).
- Extending the description of the properties of relations (OWL [2]: Web Ontology Language). The OWL provides to the Semantic Web syntax and semantics for automated reasoning about the inferences and implications of knowledge. In brief it's used generally to structure, share and exchange knowledge in universal format.

A major characteristic of a SMIL document compared with the rest of multimedia presentations is that it offers a structure clearly decomposable: components of a SMIL document (text, image, audio…) are each identified by URIs. Note that the decomposition is a fundamental operation preparing the annotation issue. Hence, the

components of a SMIL document are distinct, even pretend homogeneous during its presentation. This decomposability ensures the annotation of each element separately and we get rid of the classic problem of intricate partition of multimedia objects.

4 Semantic SMIL

The integration of Semantic Web in the information retrieval process has seen a great success expressed by the user satisfaction to the relevance of information returned after a typical search. This justified success allows to this technology to be larger than laboratories research and seek achieving prospects of general public in different fields. Hence the classic information retrieval process has been changed: the use of metadata become fundamental to annotate searchable resources. Thanks to the Metadata module, the SMIL language, performed many changes ensuring its integration to the Semantic Web view:

- The 1.0 version [3]: the “meta” element is used to define document properties (i.e. author, expiration date, key word list…) and provide values to these properties.
- The 2.0 version [4]: SMIL 2.0 extend SMIL 1.0 functionalities by the new element “metadata” which allow the use of RDF statement and make easier and more general the processing of metadata seen the ability of RDF to combine several standards of annotations as FOAF [5] and DC [6] in a single presentation.
- The 3.0 version [7]: the metadata module could be included in the body section of a SMIL document instead to be limited in the head section (as the previous versions). By this innovation we could make the description of an element right close to the definition of that element.

In figure 2 we present a set of metadata annotating an exemplar SMIL document containing the sections of this paper. In addition to the evolution of the language side to consider the semantic side of objects, further improvements are essential to a full exploitation of the principles of the Semantic Web in the context of search of SMIL documents: the use of ontology as a base of concepts composing the metadata set.

```
<rdf:Description about=http://exemple.com/article.smi
   dc:Title="Semantic multimedia search: the case of SMIL documents"
   dc:Date="2011-11-04"
      dc:Format="text/smil">
       <dc:Creator>
          <rdf:Seq ID="wrriten_by">
             <rdf:li>CHKIWA Mounira</rdf:li>
             <rdf:li>JEDIDI Anis</rdf:li>
          </rdf:Seq>
       </dc:Creator>
</rdf:Description>
```

Fig.2. Example of metadata set

5 Related Works

In the context of integration of Semantic Web technologies in the multimedia search process, many contributions are presented. Audiovisual documents cover a large range of multimedia contents commonly available such as television programs. In this topic, [8] propose a way to annotate semantically audiovisual documents by using Semantic Web languages in different levels:

- Using RDF to produce descriptions like: "the TV program" could have a "presenter" and the presenter is a "person". These descriptions seem more adequate to describe the structure and the content than the general conventional image annotation using low level techniques restricted on shapes of objects of “key frames” in an audiovisual sequence.
- Using the ontology of the audiovisual in order to formalize knowledge form descriptions, to express document patterns and to reuse those patterns in the description of documents process.

[8] uses also MPEG-7 describing technically the audiovisual resources to enrich semantic descriptions. Adopting MPEG-7 is suitable in this approach seeing its event-for features i.e., it can give details of the moment where something happens, people and even relations between objects in an emission.

In [9] we find an approach which aims to integrate a multimedia ontology into structured rich multimedia presentations such as SMIL, SVG, and Flash. The Multimedia Metadata Ontology M3O bases on Semantic Web technologies for representing sophisticated multimedia annotations. This ontology is represented in OWL; the annotations can therefore be represented in RDF, which can be directly embedded within formats such as SMIL or SVG. Note that these formats already provide appropriate means for embedding XML-based metadata.

The integration of SMIL documents in a “semantic” framework seems to [10] a way to present these type of multimedia content according to the user’s preferences. Indeed, the semantics discussed in this context is the adaptation of SMIL documents in order to respect the limitations of the hardware platform display. [10] treated separately the spatial and temporal adaptation for SMIL documents whose textual structure allows any kind of software manipulation . Thus we can redistribute the components of a multimedia document in order to change their spatial arrangements or their moments display. We can say that the semantics discussed in this context seems more user-oriented than system-oriented: the "multimedia product" is packaged according to user preferences whereas the Semantic Web technologies promote the role of engines to automatically treat semantic information.

Although the studied reflections are close to our context (semantic search of SMIL documents), multimedia documents handled in some studies are unstructured unlike SMIL documents. In the topic of semantic multimedia search, some contributions [11] treat the multimedia issue as a vague item “collection of multimedia documents” whereas some others deals with multimedia types as distinct components such as the semantic search of images or semantic search of audio sequences. In the context of processing SMIL documents, other reflections remain restricted to a technical level as in the case of spatial/temporal adaptation of SMIL documents.

6 Our contribution

In the context of semantic search in a collection of SMIL documents, we propose a search procedure composed of three modules: the description of multimedia components, querying and reporting the results. In figure 3 we describe from a technical perspective, the proposed research process.

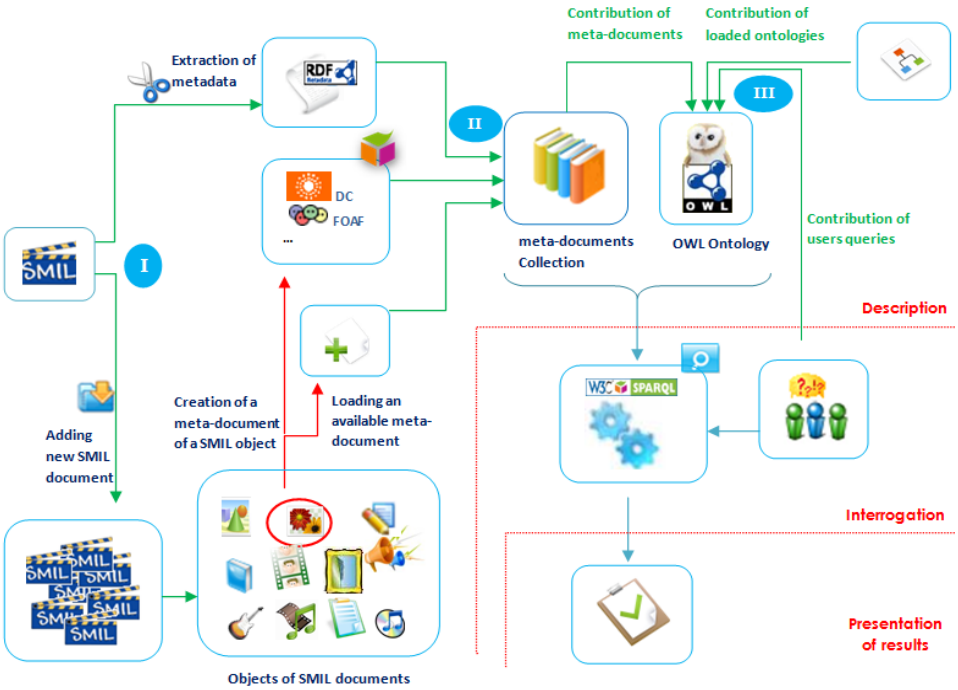

Fig. 3. Overall application architecture

The description of multimedia components is based on the transfer of a new SMIL documents to the collection, this operation is followed by an automatic process of control to check the document type and the eventually existence of an integrated

metadata. If a metadata part is found, it is extracted and assigned to a separated structure called a meta-document [12] describing the SMIL document or its components. The add operation (Part I of Figure 3) of the new SMIL is along with:

- Enumerating components existing in the SMIL document and the identification of its technical information (format, size, duration...).
- Duplicating components of the SMIL document using its URIs (specified in the code of SMIL document) and its transfer to the multimedia collection.
- A creation of a text copy of the SMIL document for further treatments.
- A record of all such information in the database for use in subsequent operations (description and the composition of a result of a query).

After adding a new SMIL document, the basic operation in the phase of description is the assignment of meta-documents [12] to the various components of SMIL documents. A meta-documents (part II of Figure 3) consists of a set of metadata, each bringing a different indication concerning the same multimedia component, take for example: the title, subject, creation date and creator of a piece of text existing in a SMIL document, all of this indications are encapsulated within the same structure : the meta-document. Compared to a traditional search process, the use of meta-documents in the phase of description brings several clear advantages:

- The description is selective: only significant items expressing the general idea of the document are described: (i.e. links images or background music are omitted).
- Provide to different types of multimedia components the same chance to be described and in this way the videos, sounds and images have the same level of expressiveness as a media text.
- A created meta-document could describe the same multimedia component existing in two or more different SMIL document which ensures its reuse.
- Offer a unified structure to annotate all multimedia components regardless their types.

The interrogation is the procedure triggered when submitting a query; user queries are categorized into three types:

- Simple query: a set of keywords query designed for the non-expert users.
- Advanced query: a set of parameters to be selected designed for more specific details and restrictions concerning the results.
- Experts query: Queries using SPARQL language oriented to the users knowing to use such language.

The interrogation allows extracting relevant information from a metadata set by comparing the query and the collection of annotations in meta-documents. The interrogation also aims to formulate and classify well the result satisfying a user need specified by the query. The classic interrogation way can consider items which are not reprehensive of a multimedia component for example when the description step extracts all the multimedia objects regardless of their value (sky, street, trees… in the image). The interrogation of SMIL documents set requires a unified structure describing multimedia components in order to perform fairly the same research process on the mixed contents. The use of meta-documents gives the privilege of querying only useful data strictly may reflect what a given component wants to express. The match meta-documents/query is performed thanks to a retrieval algorithm which takes into account the query regardless its type, turn it in SPARQL language, interrogates all of meta-documents written in RDF, retrieve relevant multimedia components (through its meta-document), assign to them a relevance score, rank multimedia items based on these scores, and finally show results.

Obtaining results starts with the selection of components / documents matching a query and followed by the classification and representation of these entities in an interactive way making easy the access to all of them. SMIL documents set presented in a given result have necessarily multimedia objects which respond to existing needs expressed in the user query. This relevance explains the degree of similarity between a query and multimedia components annotated by meta-documents. Representation of the results is the last part in the search procedure of SMIL documents. The way to display a given output could be set by the user when submitting the query: the user can choose the type of multimedia components to display [image, piece of text…] and how to display it, thus the result could be:

- Result composed by the same type of multimedia object (i.e. images only)
- Result composed by SMIL documents.
- Result grouping the two already mentioned types.
- Result composed by the same type of multimedia object grouped by SMIL document (i.e. all pieces of text in each SMIL document responding to a given query)

In our context we use ontology to retrieve relations between terms in the querying phase and to propose new queries to the user considering those relations. Independently to the progress of the three fundamental search modules, the extension of ontology is a continuous phase which aims to enrich ontology by concepts already used in the description module. For the enrichment of ontologies we propose a semi-automatic method of connecting concepts to extend an initial ontology with consideration of its meaning. The connection process (Part III of Figure 3) aims to choose a given term,

give it a type (class, property or individual in OWL), find a proper relationship with an existing term in the ontology and join the two by this relationship. The connection technique that we propose to enrich an initial ontology is based on three sources:

- From the meta-document annotating a multimedia component or a SMIL document, an automatic extraction of concepts is done using the anti-dictionary structure which removes not-meaningful terms, such as possessive pronouns or demonstrative Pronouns. Manual selection from the resulting concepts is performed in order to enrich the ontology base. After selecting a concept, we can set the connection parameters such as the type of the new concept, the relation of an existing concept in order to join the new concept to the ontology.
- From loaded ontology: the tool can automatically extract and categorize from an ontology file the constituent concepts, this extraction may drive the connection technique. To end the process of connection we should specify parameters concerning the new concept. By this type of connection we can connect even a complete OWL sub-arborescence to our initial ontology.
- From the user queries, a quantification frequency of occurrence of terms is carried out and a cloud of words based on these frequencies is established grouped by domain; the size of a term in a cloud is depending of the number of its occurrences in users' queries, finally, a selection of candidate concepts and an ordinary connection procedure could be applied.

7 Functionality

In our work we deal with a collection of SMIL documents and ontology concerning the LMD (License, Master and Doctorate) domain. The LMD Reform started in Tunisia in 2006. It aims to create flexible and efficient trainings, both fundamental and applied, offering to students wider opportunities for professional integration. We choose this domain in order to clarify some intricate notions to students using a semantic search engine based on standards of annotation which could be combined in an RDF code such as DC, FOAF and others. The functionality of our application becomes accessible through its interfaces. In this section we choose four basic interfaces among many others. The first application interface is shown at Figure 4 and it consists of three main parts designed as a flower. The first petal (blue) designed to add a new SMIL document to the collection, the second (green) is designed to trigger the search process, by the last part (orange) we can begin an annotation process in order to annotate a multimedia component.

Fig. 4. A screenshot of the application's first interface

In The orange part of figure 4 we select the SMIL document in order to annotate one of its multimedia components. This leads us to a new interface which is composed into 9 zones as we see in the figure 5. Those zones are explained subsequently:

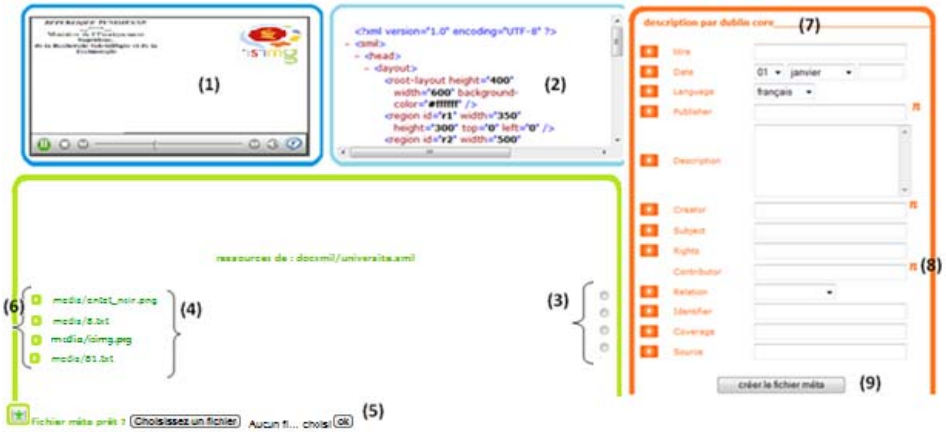

Fig. 5. form of annotation process

(1) In this area the SMIL document is played to make an idea about the overall presentation and the temporal/spatial position of the multimedia component to annotate.

(2) In this area, we find the source code of SMIL document from which the user could make a different kind of idea as the technical features of the multimedia component to annotate (the time, format, durations … are picked up automatically).

(3) Radio buttons for selecting the component to be described.

(4) List of existing components in the SMIL document and which have not yet meta-document describing them: i.e. in the document "universite.smil" There are four types of components (two text and two images).

(5) The user could load a file (.rdf or.txt extension) as a meta-document (instead of filling the form).

(6) When clicking on a green squares a window appears displaying or playing the correspondent multimedia object (image, video, animation swf, text, textstream, audio sequence).
(7) The following form contains the elements of the DC to fill in order to annotate the selected media. (Other forms could be displayed according the chosen namespace [orange petal of the previous figure] here we use DC to annotate the component).
(8) The orange "n" ensures multiple descriptions for only one item; it could create RDF sequences i.e. several authors of a single text.
(9) The check of information filled in the form and the generation of a new meta-document are done by pressing this button.

The importance of concept connection technique is that it allows making richer an ontology so we present in the next figure an example of this technique. Figure 6 shows the common window appearing when we choose a concept in order to connect it to the ontology. In our case we present a connection technique based on users' queries. The cloud of terms behind the window represents the candidates terms of connection, those terms are the most frequently used in queries concerning the LMD domain.

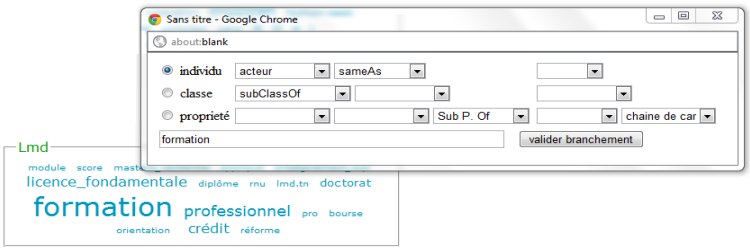

Fig. 6. Common window of connection technique

Our last chosen interface shows a typical presentation of results. Here, the type of multimedia picked in order to be searched is image, the form in the top of this figure represents two types of queries (advanced and expert query) while the other type of query (simple one) is presented in the green part of the first application interface (figure 4). Icons in the right side of this interface represent links to others interfaces of the application (clouds of queries terms, extending and loading ontologies, turn back to simple query interface …). Small blue icons right on the bottom of each image shows more details about the annotation and the rank of the correspondent image.

Fig. 7. Typical screenshot of results presentation

8 Conclusion

In this article we developed a tool for the semantic search in a SMIL documents collection. Based on a simple text, SMIL allows creating rich interactive multimedia presentation where the multimedia components are uniquely identified by URIs ensuring an easily decomposition usable in the annotation issue. The metadata annotating web resources are fundamental to join the Semantic Web principles. We use meta-documents to annotate SMIL multimedia components by a unified structure. In addition to the use of meta-documents structure in the querying module, we use also ontology which is primary in a "semantic" context. In order to extend ontology, we develop a semi-automatic connection technique considering the user queries, meta-documents and ontologies loaded to this purpose. For our short-term outlook, we wish to extend our work to be usable in a collection of multimedia documents as HTML or PDF. As for long-term prospects we hope to restrict semantic results by exploiting in deep ontology's relationships.

References

1. Resource Description Framework (RDF) http://www.w3.org/RDF/
2. Web Ontology Language (OWL) http://www.w3.org/2004/OWL/
3. Synchronized Multimedia Integration Language (SMIL) 1.0 Specification W3C Recommendation 15-June-1998 http://www.w3.org/TR/REC-smil/
4. Synchronized Multimedia Integration Language (SMIL 2.0) - [Second Edition] W3C Recommendation 07 January 2005 http://www.w3.org/TR/2005/REC-SMIL2-20050107/
5. The Friend of a Friend (FOAF) project http://www.foaf-project.org/
6. Dublin Core Metadata Initiative http://dublincore.org/
7. Synchronized Multimedia Integration Language (SMIL 3.0) W3C Recommendation December 2008 http://www.w3.org/TR/smil/
8. Troncy R. « Nouveaux outils et documents audiovisuels : les innovations du web sémantique » in « Documentaliste - Sciences de l'information » DSI 05, vol. 42, n°6. 2005.
9. Saathoff, Carsten ; Scherp, Ansgar. « Unlocking the Semantics of Multimedia Presentations in the Web with the Multimedia Metadata Ontology », Raleigh, North Carolina, USA. ACM 978-1-60558-799-8/10/04. April 26–30, 2010.

10. Sébastien Laborie, Antoine Zimmermann. « A Framework for Media Adaptation using the Web and the Semantic Web ». The Second International Workshop on Semantic Media Adaptation and Personalization (SMAP), Londres. 17-18 Decembre 2007
11. Laborie S, Manzat A., Sèdes F. « Création et utilisation d'un résumé de métadonnées pour interroger efficacement des collections multimédias distribuées » in 27th « Informatique des Organisations et Systèmes d'Information et de Décision » (INFORSID 2009), pages 227-242, Toulouse France. 26-29 May 2009.
12. Jedidi A. « modélisation générique de documents multimédia par des métadonnées : mécanismes d'annotation et d'interrogation » Thesis of « Université TOULOUSE III Paul Sabatier », France. July 2005.