SEMANTIC RELATIONSHIPS BETWEEN MULTIMEDIA RESOURCES

Mohamed Kharrat, Anis Jedidi and Faiez Gargouri

Multimedia, Information systems and Advanced Computing Laboratory, SFAX University, Sfax, Tunisia

Keywords: Semantic, Multimedia, Relationships.

Abstract: Existing systems or architectures does not almost provide a way to localize sub-parts of multimedia objects (e.g. sub regions of images, persons, events…), which represents semantics of resources. In this paper, we describe semantic relationships between resources, how to represent and create them via contextual schema. In our main system, we use languages such as XQuery to query XML resources. Using this contextual schema, the previously hidden query result can be reused to answer a subsequent query.

1   INTRODUCTION

Advances in multimedia technologies have made the storage of huge multimedia documents collections on computer systems possible. In order to allow an efficient exploitation of these collections, designing tools for accessing data resources is required. One of the biggest challenges is the exploitation of these collections as well as search and querying. But the lack of efficiency is perceived as the main obstacle for a large-scale deployment of semantic technologies. Following the same process, our system published already (Kharrat, 2011), proposes to retrieve multimedia documents using a multimodal approach. The main characteristic of our system is the use of two languages, XQuery and SPARQL to interrogate the description of multimedia resources. Performance of our system (Kharrat, 2011) can be significantly increased by applying a semantic relationship schema in a contextual document description.

The importance of discovering such links is essentially for retrieving relevant hidden resources in results.

The algorithm which is exposed here, allows the generation and publication of linked data from metadata. Any resource which is composed of many parts could have many relationships with other resources.

The aim of this paper is to present how to implement semantic relationships between data along with media resources.

The upcoming step will be the integration of this mechanism over XQuery language, which gives the possibility to add new relationship through queries. That means, concerning resources which are results of queries, we create new relationships based on these resources. XQuery will be used to build semantic relationships over queries based on functions and to harness them in second time.

The remainder of the paper is structured as follows. Section 2 provides an overview of semantic relationships. We present proposed semantic relationships and a complementary inference reasoning to build these relationships in section 3, finally we conclude in section4.

2   RELATED WORK

As far as we know, there is no research which exclusively deals with semantic relationships between multimedia resources, so we are going to present here some researches which are quite close to ours.

A number of researchers have proposed techniques to model video content. Some of these techniques are based on describing physical objects and spatial relationships between them. An approach that uses spatial relationships for representing video semantics is proposed in (Beretti, 2001).

In (Wang, 2009), the author identifies a number of semantic relations as well as artwork features and explores the use of their combinations. These

sequences of ratings allow users to derive some navigation patterns, which might enhance the accuracy of recommendations which can be reused for other recommender systems in similar domains. This system is for partially solve the cold-start and over-specialization problems for content-based recommender system.

In (Hassanzadeh, 2009), the author presents a framework to discover of semantic links from relational data. This framework supports rapid prototyping, testing and comparison of link discovery methods between different data items in a single data source. The author introduces LinQL, an extension of SQL that integrates querying with link discovery methods, which permits users to interleave declarative queries with interesting combinations of link discovery requests.

In (Erwin, 2010), the author presents a Framework (GUMF) that facilitates the brokerage of user profile information and user model representations. The goal is to allow Web applications to exchange, reuse, integrate, and enrich the user data using openly accessible data published on the Web as Linked Data with a new method that is based on configurable derivation rules that guide a new knowledge deduction process.

In (Aouadi, 2010), the author creates new links in precise region on image. This region represents the most relevant part in XML document of each image using hierarchical structure and adds weights for every link. The goal is to ameliorate the image retrieval in the semi-structured documents.

In (Murakami, 2010), the author describes the construction and evaluation of a prototype semantic relation identification system. He builds on the semantic relations {AGREEMENT, CONFLICT, CONFINEMENT, EVIDENCE, AGREEMENT and CONFLICT} that applies across facts and opinions, but that are simple enough to make automatic recognition of semantic relations between statements in Internet text possible. The author presents a system that identifies these semantic relations in Japanese Web texts using a combination of lexical, syntactic, and semantic information and evaluates the system against data that was manually constructed for this task.

All these works treat mono-media documents. They propose descriptions which allow establishing of relations between annotated concepts, resources and parts of resources. These works does not take into account multimedia resources and the set of sturdy relations which are between resources or between parts of a same resource. In addition, most of them, do not consider the semantic side. This lead

us to multi-modal querying considering multimedia features existing in each resource. We have published and presented our multi-modal system for multimedia resources retrieval. In this paper we introduce relations and rules which allow us to extract semantic from resources. We demonstrate that semantic relationships are a solution for multimodal querying.

3 SEMANTIC RELATIONSHIPS

In this work, we introduce a contextual schema which constitutes formalism for semantic relationships representation. It expresses meaning in a form that is both logically precise and humanly readable. This schema is implemented in our multimodal system.

However, not all semantically related concepts are interesting for end users. In this paper, we have identified a number of semantic relations.

Media fragments are really parts of a parent resource. The use of identifiers seems therefore appropriate to specify these media fragments. As for any identifier, access to the parent identifier shall be possible in order to inspect its context. Providing an agreed upon way to localize sub-parts of multimedia objects (e.g. sub regions of images, temporal sequences of videos or…) is fundamental.

In NewsML the metadata itself comes in bewildering variety. There are specific terms to describe every type of media. We harness them to extract contextual relations to be used in semantic and contextual recognition. Each resource in our collection is described with NewsML which is a standard for news documents. Most visual and audio features (motion, speech, text) will be used to describe each part. For example, in order to describe the content of video news, we apply concepts to describe scenes like meeting, speech, interview, live reporting or events/topics like sports, politics and commercials. Notably, we also apply the identities of persons that can be recovered from the visual flow (person who appears on the screen), form audio or from textual information.

Additional links are possible if one considers the existence of links between the locations of patients and the presence of clinical trials in these locations.

Our goal is to make semantic search based on content and structure at the same time. We do not propose to use existing links between resources, but we create our own links.

Our algorithm takes in input a resource and generates a new relationship if links exit with some

other resources.

We could extend this schema in the future with spatial and temporal relations or other types.

```
<?xml version="1.0"?>
<resources>
<resource id="IMAGE01" type="image">
<link name= "AI">
<resource id="IMAGE02"
type="image"></resource>
<resource id="VIDEO01"
type="video"></resource>
</link>
</resource>
<resource id="VIDEO07" type="video">
<link name="SH">
<resource id="IMAGE03"
type="image"></resource>
                                    …
</link>
```

Figure 1: Sample Contextual Schema.

3.1 Relationships Mechanism

T: Talk

This type of relationships describes links between resource R which contains {person, organization, team…} talking. This relation must be between image and another type of document.

TA: Talk About

This type of relationships describes links between resource R which represents {Document, reportage documentary…} and another resource R'.

S: Speak

This type of relationships describes links between resource R which contains only person and another resource R'.

SA: Speak About

This type of relationships describes links between resource R which contains only person peaking, and another resource R'.

SH: Show

This type of relationships describes links between a resource R {documentary, event, interview…} which show {person, organization, team, place…} and another resource R'.

AI: Appear In

This type of relationships describes links between a resource R which represents {person, organization, team, place…} and appears in another resource R' which represents {event, scene, sequence…}.

In the following we briefly explain the mechanism via rules that must be used to create

these links. The main goal of this mechanism is generating semantic relationships over multimedia resources.

Algorithm

```
Input: Xml resource r
Output: relation between two or more
resources in contextual schema CS
For all r   {R} do
{Extract  metadata from r and r'

If any verified module
{
If   in CS
Add new relation to CS
End If
}
Else
Execute module inference
End If
End For
Return  r ↔ r'
```

Rule 1:

$$Talk(\exists R \supset \{image\} \wedge R' \supset \{video, audio, text\}$$
$$\wedge \exists \{‹objet›, ‹person›, ‹organization›\} \supset R \supset \quad (1)$$
$$R' \wedge \exists \{‹interview›, ‹report›\} \supset R .$$

To add new relationship **Talk** the resource origin **R** must be an image and the destination **R'** could be any type (video, audio, text). Secondly metadata like ‹object›, ‹person› or ‹organization› must exist in the two resources, in addition, ‹interview› or ‹report› must be present in the destination resource.

Rule 2:

$$Speak(\exists R \supset \{image\} \wedge R'$$
$$\supset \{video, audio\} \wedge \exists \{‹person›\} \supset R \supset \quad (2)$$
$$R' \wedge \exists \{‹interview› \vee ‹speech›\} \supset R' .$$

To add new relationship **Speak** the resource origin **R** must be an image and the destination **R'** could be (video or audio). Secondly metadata only ‹person› must exist in the two resources, in addition, ‹interview› or ‹speech› must be present in the destination resource.

Rule 3:

$$TalkAbout(\exists R, R' \supset$$
$$\{image, video, audio, text\} \wedge R \equiv R' \wedge type(R) \neq \quad (3)$$
$$type(R')$$

To add new relationship **TalkAbout** the resource origin **R** and the destination **R'** could be of any type of media (image, video, audio, text). Secondly there must be a similarity between metadata of both resources. The type of related resources must be

different, (e.g. we could not relate two images or two videos).

Rule 4:

$$SpeakAbout(\exists\,R \supset \{video, audio\} \land R' \supset \{image, video, audio, text\} \land \exists\{\langle person\rangle\} \supset R \land R \equiv R' . \quad (4)$$

To add new relationship **Speak About** the resource origin R must be a video or an audio resource and the destination **R'** could be any type (image, video, audio, text). Secondly metadata ‹person› must exist in the origin resource. Finally, there must be a similarity between metadata of both resources.

Rule 5:

$$Show(\exists\,R \supset \{video, image\} \land R' \supset \{image, video, audio, text\} \land \exists\{\langle documentary\rangle, \langle event\rangle, \langle interview\rangle\} \supset R \land R \equiv R' . \quad (5)$$

To add new relationship **Show** the resource origin R could be only a video or an image and the destination **R'** could be any type (image, video, audio, and text). Secondly metadata {documentary, event, interview…} must exist in the origin resource. Finally, there must be a similarity between metadata of both resources.

Rule 6:

$$AppearIn(\exists\,R \supset \{image\} \land R' \supset \{video, image\} \land$$
$$\exists\{\langle objet\rangle, \langle person\rangle, \langle organization\rangle\} \supset R \quad (6)$$
$$\supset R'$$

To add new relationship **Appear In** the resource origin R must be only an image and the destination **R'** could be (image or video). Secondly metadata like ‹object›, ‹person› or ‹organization› must exist in the two resources.

3.2 Inference Reasoning

Since XML does not support or suggest reasoning mechanisms, we have to rely on an underlying logical formalism.

We define here some rules to deduce new relationships from existing relationships.

Case1:

$$\exists\,link(R1, R2) \land proximity(R3, R2) \Rightarrow link(R1, R3) . \quad (7)$$

R1→R2 are two related resources, so if the new resource R3 **have proximity with** R2 then
R1→R3
See Figure 2 below.

Case2:

$$\exists(link(R1, \{R\}) \equiv link(R2, \{R\})) \Rightarrow R1 = R2. \quad (8)$$

Figure 3 display two resources R1 and R2 which are related indirectly and they have many semantic relations with other resources. If these types of relationships are the same, then R1 and R2 are similar.

Case3:

$$\exists(link(R1, R2) \land link(R1, R2) \land link(R1, R3)) \Rightarrow R2 \equiv R3 \equiv R4 . \quad (9)$$

If there is semantic relationship between resource R1 and other resources as follow:
R1→R2
R1→R3
R1→R4
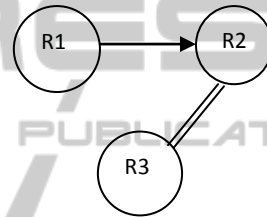So, there will be similarity between R2, R3 and R4

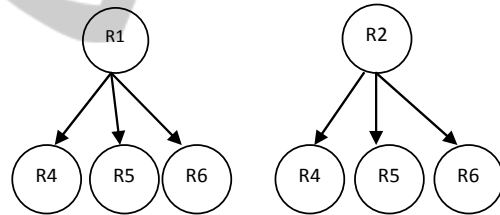Figure 2: Reasoning mechanism (case1).

Figure 3: Reasoning mechanism (case2).

We define *link (R,R')* as an existing semantic relation between R and R' and *Proximity (R,R')* as the similarity between R and R' calculated by the measure below.

3.3 Similarity Measure

We use this computation in any case we need a similarity measure which is composed of three steps.
First Step:
Pre-processing: this module is concerned with pre-processing operations preparing the input resource to be linked, it check if a resource R is typed.
Second Step:
- Comparing ‹keyword› of R and R' are equal or similar
- Comparing ‹title› of R and R' are equal or similar

345

Third Step:

Similarity is defined by some functions:

The Jaccard index is a statistic used for comparing the similarity and diversity of sample sets.

The Jaccard coefficient measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (10)$$

We use in addition, term frequency. This count is usually normalized to prevent a bias towards longer documents (which may have a higher term count regardless of the actual importance of that term in the document) to give a measure of the importance of the term t_i within the particular document d_j. Thus we have the term frequency, defined as follows:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (11)$$

where $n_{i,j}$ is the number of occurrences of the considered term (t_i) in document d_j, and the denominator is the sum of number of occurrences of all terms in document d_j, that is, the size of the document $| d_j |$.

A threshold parameter is used which changes during evaluation.

In the main system, queries attempt to find semantic contents such as specific people, objects and events in a broadcast news collection.

We define the following classes according to intent of the queries:

1. Find videos of president OBAMA speaking.
2. Find shots of archaeological sight of Carthage.
3. Find shots of the Fukushima earthquake.

Named Person: queries for finding a named person, possibly with certain actions, e.g., "Find shots of president OBAMA speaking ".

Named Object: queries for a specific object with a unique name, which distinguishes this object from other objects of the same type. For example, "Find shots of archaeological sights of Carthage".

General Object: queries for a certain type of objects, such as "Find shots of Fukushima earthquake ". They refer to a general category of objects instead of a specific one among them, though they may be qualified by adjectives or other words.

Our retrieval system needs to go through the following steps to find relevant multimedia resources for content-based queries without any user feedback and manual query expansion.

4 CONCLUSIONS

Metadata provides rich semantic relationships that can be used for retrieval purposes. This paper has presented our proposition of a contextual schema for interlinking multimedia resources semantically via XML. The goal of this schema is firstly is to be used in the main multimodal retrieval system; secondly, to provide more efficiency and recover previously hidden query result. An initial evaluation of the algorithm has shown good results.

The next step is to integrate the process of building these semantic relationships process to XQuery language, which gives the possibility to add new relationship over queries. Based on resulting resources, we could build new relationships that will be used in second time. Furthermore, we plan to investigate the weights for different semantic relations based on their relevance. More investigations are still present.

REFERENCES

Aouadi, H., Torjmen, M., 2010. Exploitation des liens pour la recherche d'images dans des documents XML. In *Conférence francophone en Recherche d'Information et Applications –CORIA.*

Beretti, S., Del Bimbo, A., Vicario, E., 2001. Efficient Matching and Indexing of Graph Models in Content-Based Retrieval. In *IEEE Transactions on Pattern Analysis and Machine Intelligence - Graph algorithms and computer vision 23(10):* 1089-1105.

Erwin, L., Fabian, A., Dominikus, H., Eelco, H., Jan, H., Geert-Jan, H., 2010. A Flexible Rule-Based Method for Interlinking, Integrating, and Enriching User Data**.** In the *Proceedings of the 10ᵗʰ ICWE 2010, Springer Verlag,* Vienna, Austria, July.

Hassanzadeh, O., Kementsietsidis, A., Lim, L., Miller, R J., Wang, M., 2009. A framework for semantic link discovery over relational data. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management,* CIKM, Hong Kong, China, November 2-6, 2009, 1027-1036.

Kharrat, M., Jedidi, A., Gargouri, F., 2011. A system proposal for multimodal retrieval of multimedia documents. In Parallel and Distributed Processing with Applications Workshops (ISPAW), *2011 Ninth IEEE International Symposium on Parallel and Distributed Processing with Applications- Busan-Korea*, pages 177 -182.

Murakami, K., Nichols, E., Mizuno, J., Watanabe, Y., Goto, H., Ohki, M., Matsuyoshi, S., Inui, K., Matsumoto, Y., 2010. Automatic Classification of Semantic Relations between Facts and Opinions. In *Proceedings of the Second International Workshop on*

NLP Challenges. In the Information Explosion Era (NLPIX 2010). Beijing, China.

Wang, Y., Stash, N., Aroyo, L., Hollink, L., Schreiber, G., 2009. Semantic relations in content-based recommender systems. In *5th International Conference on Knowledge Capture (K-CAP 2009)*, September 1-4, 2009, Redondo Beach, CA, pages 209-210.