A LEGO-like Metadata Architecture for Image Search&Retrieval

Ruben Tous, Jaime Delgado
Universitat Politècnica de Catalunya (UPC-Barcelona Tech)
Barcelona, Spain
rtous@ac.upc.edu

Abstract

This paper proposes a set of architectural patterns for the design of interoperable image search&retrieval systems on top of the latest standards in the area, ISO/IEC 15938-12 (MPEG Query Format) and ISO/IEC 24800 (JPSearch). The work introduces the notion of a "composable metadata model" as a means to allow the usage of multiple metadata ontologies together, along with mechanisms to specify semantic mappings among them. The paper also studies the process to choose a Core Metadata Ontology (based on the work taking place at the W3C Media Annotations Working Group), which plays the role of a minimal lingua franca among the different systems, providing a basic level of interoperability.

1. Introduction

Digital still and moving images are being generated, distributed and stored worldwide at an ever increasing rate. Therefore, Image Search&Retrieval tasks have become increasingly important in recent years. While multiple tools already exist, unfortunately almost every one provides a different search interface and multimedia metadata description format. This fact prevents users from experiencing a unified access to the repositories, and, which is probably worst, the impossibility to migrate the (costly) image metadata annotations between different systems. Therefore, several standardization efforts are trying to unify the way digital images are annotated, searched and retrieved. Two of the more relevant initiatives are the MPEG Query Format (MPQF) [1][2][3] and the JPEG's JPSearch project [4][5]. While MPQF has already reached its last standardization level, and offers a promising solution for the query interface interoperability, JPSearch (whose Part 3 makes use of MPQF) is still an ongoing work, and faces the difficult challenge to provide an interoperable architecture for images' metadata management. The coexistence of multiple metadata formats currently breaks the query uniformity provided by MPQF, and poses a controversial debate.

This paper aims contributing to solve some of the current problems of images' metadata management architectures by introducing the notion of a "*composable metadata model"* as a means to allow the usage of multiple metadata formats together, along with mechanisms to specify semantic mappings among the different formats, within an image search&retrieval system. The paper defends the idea that the choice of a unique metadata model is not sustainable in the mid-term, and that users should be able to snap together selected "building blocks" provided by different metadata standards, the "LEGO Metaphor".

2. Architecture overview

Figure 1 presents an overview of the proposed architecture. The scope and aims of the individual parts are highlighted in the following sections. The architecture exposes three main sets of functionalities through three different interfaces which we believe should be based on international standards. The first interface is the query format, for which we propose the usage of ISO/IEC 15938-12 (MPEG Query Format). The second interface is the one allowing the transfer of image files and their related metadata between different repositories. One possibility is the usage of ISO/IEC 24800-5 (JPSearch Part 5), which is now reaching its last standardization stages. The third main interface is the one allowing the registration of new metadata ontologies (which includes the registration of new ontology mappings). We are currently collaborating in the standardization of ISO/IEC 24800-2 (JPSearch Part 2), which will be probably finished before the summer of 2010.

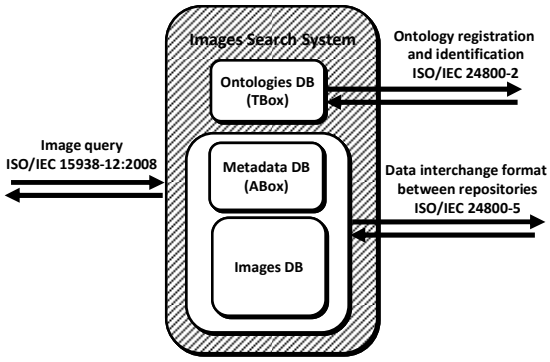

Figure 1 Architecture overview

However, selecting or designing the different interfaces is not enough, and could even drive to a useless artifact if done inappropriately. Each interface is designed assuming an underlying metadata model, so it is necessary first to clarify which is the overall metadata model of the architecture. Next sections will try clarifying this aspect by:

- Describing the *mTags* metadata model, allowing the usage of multiple metadata languages in the annotations and the queries.
- Specifying how new *mTags* vocabularies can be defined in the form of ontologies and registered into the system.
- Specifying how mappings between *mTag* ontologies can be defined and registered.

3. Annotating images with multiple metadata ontologies

Structured metadata annotations are crucial in advanced image retrieval systems. Metadata-based search is a relevant feature in addition to keyword-based retrieval and CBIR techniques. The quest for the definitive metadata model has been a "white whale" in the multimedia community during the last years. Among dozens of relevant attempts, MPEG-7 emerged as (probably) the most accurate and complete collective effort to deliver a universal metadata description suite. However, in practice, no metadata model has reached enough adoption to be considered a success, apart from those which are coupled to popular media formats (e.g. ID3, iTunes). The proliferation of new formats, along with the social tagging phenomenon (including new user-defined tags), seems driving to the conclusion that the choice of a unique metadata model is not sustainable in the mid-term (see Figure 2).

With the Semantic Web initiative in the background, a new paradigm for interoperability revises the traditional conception. Instead of sharing a common metadata superset, users from different contexts will be able to snap together selected description "building blocks" provided by different metadata standards.

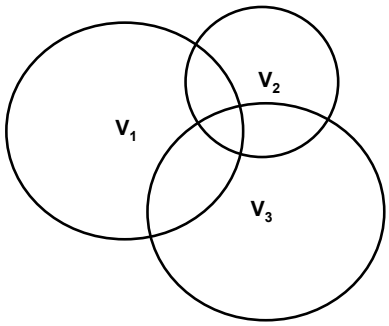

Figure 2 Graphical representation of multiple overlapping metadata ontologies (V_1, V_2, V_3)

3.1 Composable metadata models, the LEGO Metaphor

A helpful metaphor is to think of metadata annotations as a assembly of LEGO® pieces, being each one of the pieces an individual metadata statement (e.g. "format = image/jpeg") and being each metadata format a kit of several LEGO pieces. In real LEGO, individual pieces from different kits can be easily combined to build complex artefacts because each individual piece has a clearly defined interface (self-explanatory or well documented) and can be used independently. Coming back to the metadata world, some metadata formats are based in data model theories that do not facilitate the usage of individual metadata statements separatedly (e.g. XML). The Semantic Web data model theory, the RDF model, already allows the individual usage of metadata statements. However, as pointed out in [6], this is not enough if we do not ensure that the syntax and semantics of each "metadata piece" will be always clearly specified (e.g. specifying that the "format" property, when used for an image, expects an image-based MIME type). Current Semantic Web technologies allow to freely specify the syntax and semantics of properties, but the absence of compulsory canonical ways of doing it impedes the achievement of the LEGO Metaphor in practice.

3.2 Architecture's metadata model

Any image search&retrieval system has an underlying metadata model which determines an information space comprising the available metadata structures about images and other elements in the

system. The proposed reference architecture is based in what we call the *mTags* model (from"machine" tags, inspired in Flickr's Machine Tags [11]). The *mTags* model determines which information can be stored (independently or embedded in image files as described in ISO/IEC 24800-4), exchanged (as described in ISO/IEC 24800-5), queried (as described in ISO/IEC 15938-12), and also how to interpret this information. The proposed model is independent of the internal formats used for storing and managing metadata. The model acts just as a logical view giving to the user the illusion that there's a metadata catalogue of all the images and other entities accessible through the architecture.

3.3 An statement-based model

The proposed *mTags* metadata model allows the management of metadata annotations at the statement level. This fine granularity facilitates the achievement of the "LEGO Metaphor" mentioned before, and facilitates the definition of ontologies of tags and also ontologies of semantic mappings.

The suggested approach specifies a metadata model built on work undertaken by the World Wide Web Consortium (W3C) on the Semantic Web[1], but provides interfaces which are independent of particular implementations (such as RDF, RDFS or OWL). The approach is also inspired by current practices, especially by the Dublin Core Metadata Initiative [6] and the Flickr's Machine Tags initiative [11].

The result allows overcoming the limitations of the traditional approach for managing metadata using only XML schemas. Metadata elements defined with XML schemas do not have an individual identity, and consequently they cannot be used or understand alone, outside their surrounding structure. They cannot be individually combined with other metadata languages and cannot be referenced from ontologies to define mappings to other languages.

3.4 The *mTags* model

A "machine tag" (abbreviated *mtag*) is basically an independent property-value pair (for example, "*dc:format=image/jpeg*"). When omitted, the subject of the mtag is assumed to be the image being described in the context of the mtag, however, other subjects can be specified (e.g. users, collections, etc.).

Subject	Property (predicate)	Value (object)
(this image)$_{implicit}$	*dc:format*	*image/jpeg*

Table 1. Example *mtag*

The *mTags* model builds on work undertaken by the World Wide Web Consortium (W3C) on the Resource Description Framework (RDF). The model is based upon the idea of making statements about resources, in particular, images, in the form of subject-predicate-object expressions, called triples in RDF terminology. The subject denotes the resource being described, and the predicate expresses a relationship between the subject and the object. An image can be annotated with multiple mtags belonging to different metadata ontologies (e.g. Dublin Core or EXIF).

dc:creator=*"John Smith"*
dc:title=*"San Francisco"*
geo:long=*100.3*
exif:make=*"Canon"*
exif:model=*"Canon IXY DIGITAL 30"*
exif:xResolution=*"180/1"*
exif:yResolution=*"180/1"*
exif:resolutionUnit=*"inch"*

Table 2 Example *mtags*

All the *mtags* in the previous example have as implicit subjects the image being described. However, a different resource can be used as subject.

dc:creator=*<urn:name:john:smith>*
<urn:name:john:smith> ***foaf:name***=*"John Smith"*
<urn:name:john:smith> ***foaf:homepage***=*"http://johnsmith.org"*
exif:IFD=*<urn:anonymous:aux1>*
<urn:anonymous:aux1> ***exif:exposureTime***=*"1/400"*
<urn:anonymous:aux1> ***exif:exifVersion***=*"2.20"*

Table 3 Example mtags with explicit subjects

4. The choice of a Core Metadata Ontology

As explained in the previous sections, the possibility to choose a unique superset metadata vocabulary has been discarded in our architectural design. We adhere to the new interoperability paradigm in which multiple metadata vocabularies are used together, weaving semantic relationships among them when necessary. However, following a usual practice in the industry (e.g. the Flickr's API or the YouTube API), we propose providing a subset of metadata terms which will be present in any system, and which will provide a minimum level of interoperability among the different contexts. This pragmatic design pattern is based in the "pidgin" concept. In linguistics, a pidgin (originally

[1] http://www.w3.org/2008/01/media-annotations-wg.html

used to describe Chinese Pidgin English) is a simplified language that develops as a means of communication between two or more groups that do not have a language in common, in situations such as trade. We call our simplified language "core metadata ontology" (see Figure 3).

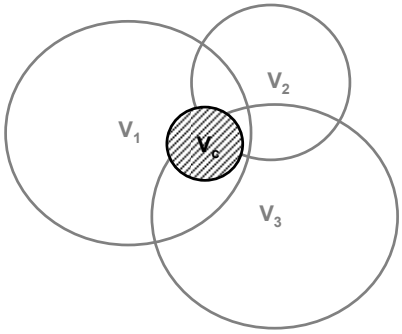

Figure 3 Graphical representation of the Core Metadata Ontology (V_c)

The process of selection of the metadata terms to be included in the core ontology falls outside the scope of this paper. In fact, the selected set of terms is still open, because it comes from the ongoing work within the W3C Media Annotations Working Group (MAWG). Among other expected outcomes, the group is working in the specification of an ontology and a client-side API for cross-community data integration of information related to media objects on the Web. The purpose of the ontology and the API is to help circumventing the proliferation of multimedia metadata formats by providing a basic set of metadata access functions and also some partial mappings between the existing formats. A very accurate work has been done within MAWG to select the basic set of access functions, which implies also the selection of a core set of metadata terms. We adopt this set (which is still subject to modifications) as our core metadata ontology. Currently, the group has taken parts of Adobe XMP as the core ontology, trying to map them to the other metadata formats. The selected parts of XMP are: The totality of the XMP Dublin Core Schema, some elements of the XMP Basic Schema, and some elements from the XMP Media Management Schema.

5. Declaring and registering metadata ontologies and semantic mappings between them

The simplest form of a *mTag* Ontology is a namespace, a namespace prefix, and an enumeration of properties related to an image with their respective labels and textual definitions. MTags Ontologies are defined using the *mTag* Ontology Definition Language (*m*ODL), which is defined to allow its direct mapping to the Typed Node Abbreviation form of RDFS.

In turn, we define the *mTag* Ontology Mappings Description Language (*m*OMDL), which provides a tool for declaring semantic correspondences between different *mtags*. A *m*OMDL document declares a set of mapping rules. Each mapping rule consists on one or more `FromField` elements and one or more `ToField` elements.

The following example shows a one-to-one mapping rule which maps `v1:date` mtag (formatted as `dd/mm/yyyy`) into the `v2:date` mtag formated as `dd.mm.yyyy`.

```
<MappingRules>
    <FromField xsi:type="SourceFieldType">
      <XPathExpression>//Creator//GivenName
      </XPathExpression>
    </FromField>
    <FromField xsi:type="SourceFieldType">
      <XPathExpression>//Creator//FamilyName
      </XPathExpression>
    </FromField>
    <ToField xsi:type="FormattedTargetFieldType">
      <XPathExpression>v2:creator</XPathExpression>
      <ReplaceWithRegExp>$1 $2</ReplaceWithRegExp>
  </ToField>
    </MappingRule>
</MappingRules>
```

Table 4 Example one-to-many mapping

We envisage that the mappings will be internally solved by using Semantic Web tools, such as SPARQL queries and rules, but the architecture just specifies an independent way of declaring them.

6. Querying images

MPEG has recently finished the ISO/IEC 15938-12 a.k.a. MPEG Query Format (MPQF), which standardizes a query language for multimedia repositories. We have chosen MPQF as the main search interface for the proposed architecture.

It is possible to specify conditions regarding *mtags* within the MPEG Query Format. The same elements used to refer the XML metadata can be used to refer to a certain *mtag*. The example in Table 5 shows a query asking for images titled "Blade Runner".

```
<MpegQuery>
    <Query>
      <Input>
        <QueryCondition>
          <Condition xsi:type="Equal">
              <StringField typeName="mtag">dc:Title
              </StringField>
              <StringValue>Blade Runner</StringValue>
          </Condition>
        </QueryCondition>
      </Input>
    </Query>
</MpegQuery>
```

Table 5 Example MPQF query with a *mtag*-based condition

7. Conclusions and future work

This paper has presented a set of architectural patterns which aim overcoming the problems related to the multiplicity of metadata vocabularies in the design of interoperable image search&retrieval systems. The key points are 1) The usage of standard interfaces as ISO/IEC 15938-12 (MPEG Query Format) for querying and ISO/IEC 24800 (JPSearch) for metadata management, 2) The definition of the *mTags* statement-based metadata model, 3) The definition of a core metadata ontology, based on the work of the W3C Media Annotations Working Group and 4) The provision of practical mechanisms to define semantic relationships (e.g. mappings) between different metadata ontologies.

The paper introduces the notion of a "composable metadata model" as an alternative to traditional choice of a monolithic model. This is achieved managing the metadata annotations at the statement level. This fine granularity facilitates the achievement of the "LEGO Metaphor", allowing users the maximum freedom when using and defining metadata tags. This approach, in combination of semantic mappings and a core ontology, allows overcoming the limitations of the traditional approach for managing metadata using only XML schemas.

The authors contributed the approach to JPEG [8][9], and its being studied to become part of JPSearch Part 2 [5].

8. Acknowledgements

This work has been partly supported by the Spanish government (TEC2008-06692-C02-01) and the European Network of Excellence VISNET-II (IST-2005-2.41.6), co-funded under the European Commission IST 6th Framework Program.

9. References

[1] ISO/IEC 15938-12:2008 Information Technology - Multimedia Content Description Interface - Part 12: Query Format.

[2] Ruben Tous and Jaime Delgado. L7, an MPEG-7 Query Framework. 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS 2007), Barcelona, November 2007. IEEE Computer Society Press. ISBN 0-7695-3030-3. p. 256-263

[3] Mario Döller, Ruben Tous, Matthias Gruhne, Kyoungro Yoon, Masanori Sano, and Ian S. Burnett. The MPEG Query Format: On the Way to Unify the Access to Multimedia Retrieval Systems. IEEE Multimedia, ISSN: 1070-986X, volume 15, issue 4, 2008.

[4] ISO/IEC 24800-3:2008 CD Information technology - JPSearch - Part 3: JPSearch Query format. Output document from the 45th ISO/IEC JTC 1/SC 29/WG 1 Poitiers meeting, July 7th to 11th, 2008

[5] ISO/IEC WD 24800-2:2008 WD Information technology - JPSearch - Part 2: Schema and Ontology Registration and Identification. Output document from the 45th ISO/IEC JTC 1/SC 29/WG 1 Poitiers meeting, July 7th to 11th, 2008.

[6] The Dublin Core Metadata Initiative (DCMI). See http://dublincore.org/.

[7] Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 February 2004. See http://www.w3.org/TR/rdf-concepts/.

[8] Ruben Tous and Jaime Delgado. WD version 4 of JPSearch Part 2 (24800-2) Registration, Identification, and Management of Metadata Schema and Ontology. Input document number wg1n4869 to the 47th ISO/IEC JTC 1/SC 29/WG 1 San Francisco meeting, January, 2009.

[9] Ruben Tous, Jaime Delgado, Anna Carreras, and Eva Rodriguez. Proposal of a New Part for JPSearch. ISO/IEC 24800 Part 6 - Metadata Ontology. Input document number wg1n4786 to the 46th ISO/IEC JTC 1/SC 29/WG 1 Busan meeting, October, 2008.

[10] Ontology for Media Object 1.0. W3C Working Draft 17 March 2009. http://www.w3.org/TR/mediaont-1.0.

[11] Flickr's Machine Tags. http://www.flickr.com/groups/api/discuss/72157594497877875/.