A Survey of Multimedia Annotation Localisation on the Web of Linked Data

Gary Lefman[1], Dave Lewis[1], Felix Sasaki[2]
[1]CNGL the Centre for Global Intelligent Content, Trinity College Dublin, Ireland
[2]Language Technology Lab, German Research Centre for Artificial Intelligence (DFKI), Germany
E-mail: lefmang@tcd.ie, dave.lewis@cs.tcd.ie, felix.sasaki@dfki.de

Abstract

Multimedia annotation generates a vast amount of monolingual data that helps to describe audio, video, and still images. These annotations are, however, unlikely to be useful to people that do not communicate in the same language. Annotations may also have insufficient context for people from different cultures. There is a demand for localised annotations as well as localised multimedia. If annotated resources are to be shared effectively on the Web of Linked Data, they need to be connected to similar resources that have already been adapted for other languages and cultures. In the absence of Linked Data, monolingual annotations remain trapped in silos and cannot, therefore, be shared. We have identified a gap in the localisation continuity between multimedia annotations and the Web of Linked Data. Flickr was examined as an example, and also as a representative candidate, of open social media platforms. Localisation was also taken into consideration when looking at current multimedia ontologies and Linked Data frameworks.

Keywords: Localisation, Multimedia Annotation, Linguistic Linked Open Data, Multimedia Ontology, RDF, OWL, Flickr

1. Introduction

When people annotate images, videos, and graphics on the Web, they are describing their observations and experiences in a manner that can be shared and retrieved. Multimedia annotations may contain observable and clearly recognisable characteristics of a scene, such as a tree on a grassy hill, and metaphysical properties, like emotions, that may be derived from the same scene. Multimedia annotations on an open social media platform may be written in different languages, and by people who might represent completely different cultures. A user from England might look at our scene and add an annotation with the term "tree", a user from Wales sees the tree and annotates it with "coeden" (Welsh:tree), and another user from Wales adds "bren" (Welsh:wood). The three different annotations all represent the same object on this image. Searching for any of these terms would show this image. However, if each annotation was added to a different image, each one on a different platform, we would want to ensure a resource for all three images was extracted in a search for "coeden".

Multimedia is usually annotated in one language; the user's primary language, or the language commonly used by a community of users. Regardless of the language used, an open social media platform may index these annotations in a monolingual manner. There will be no simple way to index or search for an annotation by language if, for example, there is no language tag in the resource URI to explicitly identify it. Nor will it be easy to link it to other monolingual resources. These multilingual annotations are effectively mixed within a single container. Any attempt to link this data would be pot luck because the target language of an annotation cannot be guaranteed without intervening translation. This applies equally to intra- and inter-platform environments.

Gracia et al. (2012) suggest that users expect access to the Web of Data to be *intuitive, appealing, and effortless*. We believe this is also true for Linguistic Linked Open Data. Linking semantically related resources across the Web of Data *can improve the situation* (Sasaki, 2013) of, what would otherwise be, disparate linguistic platforms. The alternative is to replicate and directly adapt every annotation. This can be expensive and time-consuming. Whereas linking related media annotation utilises existing multilingual resources.

We carried out this survey to determine if there are gaps in the continuity between multilingual annotated multimedia and open social media platforms across the Web of Linked Data. We selected Flickr[1] , an open social media platform for sharing images and videos, on which to focus the survey because it provides users with an annotation feature. Since the lexical and semantic value of annotations has already been well-defined (Cimiano et al., 2010), our approach was to examine the *practicality* of multimedia annotation localisation.

Whilst writing about multimedia localisation, we are broadly referring to the adaptation of audio and visual media to meet the specific needs of natural languages and cultures. In this paper we concentrate only on images, but in the context of Flickr this may also apply to video as well.

2. Related Work

There has been a considerable amount of research into the extraction of multimedia annotations, and linking them semantically with other resources. However, none of the studies appear to have examined the relevance of localised annotations and their impact on the Web of Data.

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

The most recent work (Im & Park, 2014) exploited the semantic relationship between the tags of image pairs using RDF (W3C, 2004) and OWL (W3C, 2012c), whilst utilising the link analysis algorithm HITS to rank comparable annotations. They also used the Linked Open Data provider DBpedia[2] for its heuristics and ontology classes in order to avoid disambiguation between tags. Their "Linked Tag" approach appears to have operated in isolation of other open social media platforms. This would have decreased the opportunity for access to resource annotations in languages other than the language used in their experiments.

On 1 November 2013, a project called LIDER[3] was set up to study Linguistic Linked Open Data (LLOD) and its applicability to multilingual and multimedia content. Among other items, its two-year mission is to provide guidelines and best practices for building and exploiting resources in this area on the Web. The front-end of LIDER is the W3C Linked Data for Language Technology[4] (LD4LT) community group, whose remit spans across the wider linguistic linked data arena.

3. Multimedia Annotations and Metadata

Annotation of images involves the application of an extra layer of information that is related to it, whether this is applied manually by a user or automatically by a computer-based agent. In addition to annotations, metadata also plays a role in describing multimedia for the purpose of cataloguing, searching, and retrieving etc. Metadata, on the other hand, tends to be more definitive, providing technical attributes of the media format and how it was created.

Multimedia annotation data is typically stored in an external resource, such as databases like the Microsoft Research Annotation System (Grudin & Bargeron, 2005), and freeDB[5]; although a limited amount of relevant meta-data may be wrapped-up inside media file containers. Scalable Vector Graphics (SVG) (W3C, 2011) is one example. Self-contained annotations are immensely portable, but they are, however, silos of information that must be extracted and stored externally if they are to be useful to the Web of Linked Data. This is because self-contained annotations are complex, making it harder to index and search them. The use of an open media annotation format makes it easier to extract and index their metadata. Externalising metadata simplifies the process of mapping its resources into a common vocabulary, and linking it to other data. Therefore indexing, searching, and integration with other content become easier, too.

Current multimedia metadata formats like MPEG-7 or ID3 ((W3C, 2012b) provides an overview) mostly do not rely on Semantic Web technologies and were developed before the Semantic Web itself was conceived (Sjekavica et al., 2014). This leads us to believe that the first hurdle to cross is the standardisation of multimedia annotation techniques. This challenge might be exaggerated when we also consider that multimedia annotation techniques in use today are essentially monolingual in nature. This isn't necessarily a concern from a linguistics standpoint, because effective internationalisation allows for the adaptation for other languages. But it does present a potential problem when examining the rest of the localisation process. During localisation, images can also be adapted. Thus objects may move or change shape, and the spoken language might change, too. The Media Fragments URI specification (W3C, 2012a) provides the technologies to tackle this problem. It is also important to provide a means for the same (or similar) resource to be linked, so that they may all be included in the wide scope of queries.

4. Ontologies and Linked Open Data Frameworks for Multimedia

Multimedia ontologies provide a formal vocabulary for the identification and arrangement of audio and visual resources. They support the semantics of images in a manner that is consistent, permitting successful storage and retrieval of multimedia properties. General ontologies like schema.org provide domain-specific areas for media. Schema.org relies on a simple taxonomic model. This can be serialised as microdata or in the linked data RDFa Lite 1.1 (W3C, 2012d) form, but lacks the expressive power of OWL ontologies, that does not go beyond subClassOf relations. It is possible, though, to generate Linked Data from microdata. Nonetheless, there has been a movement towards mapping it to the Web of Data (Nogales et al., 2013). Dublin Core (DCMI, 2012) is another generic ontology that describes documents. It can be applied to several models; the Resource Description Framework (RDF), and the more simplified RDF in attributes (RDFa). Dublin Core supports the multimedia types (classes) `Image`, `MovingImage`, `Sound`, and `StillImage`. Some of the properties for these classes that can be localised, such as `dc:description`, `dc:title`, and `dc:subject`.

Dedicated multimedia ontologies, on the other hand, are more conducive to describing image resources. The W3C Ontology for Media Resources (MediaONT) (W3C, 2012b) was purposefully designed with the Web of Data in mind. It provides a level of abstraction to interrelate the aforementioned, not Semantic Web based multimedia metadata formats like MPEG-7 or ID3.

MediaONT is essentially monolingual in nature, pertaining to a single semantic concept per element or property. Some ontologies provide a way of identifying the natural language of the data that is applied to it. But they do not necessarily cater for data provided in multiple

[2] http://dbpedia.org/
[3] http://www.lider-project.eu/
[4] http://www.w3.org/community/ld4lt/
[5] http://www.freedb.org/

languages within the same property. Dublin Core and MediaONT, for example, use `Language` and `language`, respectively. The value of the language code applied to them is invariably BCP47 (IETF, 2009) or the less granular RFC-3066 (IETF, 2001) format. Since the ontologies do not directly support multiple variations within the same property, they rely upon a wrapper to contain the linguistic variations.

5. An Example: Multimedia Annotation in Flickr

Flickr is an open social media platform for sharing images and video. Users can apply spatial annotations to selected regions of images shared by others, as well as their own. Flickr annotations are called tags, which are heterogeneous folksonomies that mostly contain unstructured data (Concas et al., 2014). This kind of tagging requires users to interpret image contents (Konkova et al., 2014). Interpretations may be personal or biased, depending upon the users' social context and language. In addition to tags, structured Exif (CIPA, 2012) metadata in the form of embedded camera, lens, and exposure properties may be recorded. Devices fitted with a GPS receiver may also record geospatial data in the form of longitude and latitude coordinates.

The lack of tag structure in Flickr presents a problem, in the sense that there is no ontology to which users can apply their tags. All of the terms are essentially collected in a single container or set (Marlow et al., 2006), without any form of classification. Therefore, it is unlikely that the semantic value of these tags can be determined from the relationships between images, alone. To compound this issue, many users can tag the same image with a variety of terms that have the same meaning. This becomes apparent when considering users may apply tags in different natural languages, as observed by (Allam, 2013). For example, user A tags an image of a cat with the English word "cat" and user B tags a different image of a cat in Norwegian as "katt". Since no ontology is employed, a search for "katt" will show only the images of cats that were tagged with the Norwegian term. Images that were tagged with the English term will not be shown due to this *disagreement in vocabulary* (Marlow, Naaman et al., 2006). Furthermore, images of people with the personal name "Katt" will be returned, emphasising the ambiguity that is introduced with the lack of ontological structure. Therefore localised tags are *meaningless to a global audience* (Konkova, Göker et al., 2014) if there is no facility to link heterogeneous tags.

Managing localised multimedia annotations on the Web of Data appears to be a challenge that may stem from the source of the annotations. In Flickr's case, ambiguity is introduced through the absence of ontology and language identification. There have, however, been several successful attempts to extract relationships between Flickr tags across the Web of Data. One example is LinkedTV. This was a project that presented a URI-based RESTful Linked Services Infrastructure (Nixon, 2013), which used a model for aggregating tags from Flickr, and leveraged MediaONT to classify them. They also used RDF to bridge the gap between tagged images in Flickr and the Web of Data. This allowed for the extraction of related content from other online services like YouTube[6] and Instagram[7] . However, they explicitly ignored RDF labels that were not in English. So there was a missed opportunity to utilise a Linguistic Linked Open Data source (Chiarcos et al., 2012), such as DBpedia, to extract resource URIs from relationships with other languages.

Flickr also recognises machine tags (Flickr, 2014), which are annotations with text that conforms to the syntax `namespace:predicate=value`. To carry out a search, the machine tag is appended to a Flickr URL and submitted. The `namespace` and `predicate` properties can be any term with the only restriction being that they must match the regular expression `^[a-zA-Z0-9_].*`. The value may consist of a double-quote encapsulated string containing any percent-encoded character. Both `namespace` and `predicate` are uncontrolled, so the user is free to enter whatever they like (Yee, 2008), although Flickr does offer a few suggestions. Interestingly, one of these suggestions refers to Dublin Core, using the form `dc:title=value`. This namespace, however, appeared to be rarely used. A quick experiment applied to a Web browser demonstrated this through the use of the wildcard URI `http://www.flickr.com/photos/tags/dc:title=*`. This resulted in only 78 images tagged with the title property, which is a considerably small number considering that over 580 million photos were added to the service in 2013 alone (Michael, 2014). The wildcard URI `http://www.flickr.com/photos/tags/dc:*` for the entire Dublin Core name space resulted in 132,789 images spanning 10 properties. Those properties included `dc:identifier` and `dc:author`, and `dc:subject`. It's worth noting that `dc:author` is not an authoritative Dublin Core metadata term, which highlights the lack of control over the use of namespace and predicates.

The ability to annotate Flickr multimedia with machine tags, albeit unstructured and loosely controlled, does provide an open channel to resources that would be beneficial to the Web of Data. The challenge is a lack suggestion when users annotate resources. Better management of machine tags could be gained through the recognition of annotations starting with `dc:`. Users could then be presented with a choice of authoritative Dublin Core properties from which to choose. This would result in a hybrid of the "set" and "suggestive" classifications proposed by Marlow, Naaman et al. (2006). Of particular interest is `dc:language`, which would offer greater flexibility in matching related resources in Linguistic Linked Open Data. This feature could also be extended to

[6] http://www.youtube.com/
[7] http://instagram.com/

the MediaONT namespace `ma:` to support several additional properties that are absent from Dublin Core, although, there is no reason why users cannot use it now. It was observed that Flickr documentation of machine tags was sparse, which may have contributed to poor adoption of the Dublin Core namespace.

6. Conclusion

We have examined the role of localisation in multimedia annotation and how annotation data relates to multimedia ontologies and Linked Open Data. The focus of our survey has been on the open social media platform called Flickr. The goal was to identify gaps in the continuity between multilingual annotated images and the Web of Linked Data. To the best of our knowledge, there has been no consideration of localisation in the multimedia annotation technologies examined in this paper. Where multimedia ontologies are present, they are not inherently multilingual. This provides an opportunity for Linguistic Linked Open Data to bridge the gap between multimedia annotation in social media and the Web of Linked Data. Linguistic Linked Open Data would not only provide a way to semantically link annotations between languages, but also link annotations across other open social media platforms.

7. Future Work

Exposing multimedia annotations to the Web of Linked Data will increase accessibility to multilingual information, for machines and people alike. With this in mind, we would like to continue research into linking social media folksonomies across languages and across social media platforms, with a view to integrating information with Linked Open Data resources. We will consider MediaONT to formalise multimedia annotations in social media, using RDF/OWL, and investigate whether Media Fragments URI can play a role or not.

8. Acknowledgements

This research is partially supported by the European Commission as part of the LIDER project (contract number 610782) and by the Science Foundation Ireland (Grant 12/CE/I2267) as part of the CNGL Centre for Global Intelligent content (www.cngl.ie) at Trinity College Dublin.

9. References

Allam, H. (2013). Social, Technical, and Organizational Determinants of Employees' Participation in Enterprise Social Tagging Tools: A Conceptual Model and an Empirical Investigation. PhD Thesis, Dalhousie University.

Chiarcos, C., Hellmann, S., et al. (2012). Linking linguistic resources: Examples from the open linguistics working group. Linked Data in Linguistics. Springer.

Cimiano, P., Montiel-Ponsoda, E., et al. (2010). A note on ontology localization. *Applied Ontology*, 5**,** 127-137.

CIPA (2012). Exchangeable image file format for digital still cameras: Exif Version 2.3, December 2012. Available: http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf

Concas, G., Pani, F. E., et al. (2014). Using an Ontology for Multimedia Content Semantics. Distributed Systems and Applications of Information Filtering and Retrieval. Springer.

DCMI (2012). DCMI Metadata Terms, 14 June 2012. Available: http://dublincore.org/documents/dcmi-terms/

Flickr. (2014). Flickr Tags FAQ [Online]. Available: http://www.flickr.com/help/tags/ [Accessed 02/02/2014].

Gracia, J., Montiel-Ponsoda, E., et al. (2012). Challenges for the multilingual web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 11**,** 63-71.

Grudin, J. & Bargeron, D. (2005). Multimedia annotation: an unsuccessful tool becomes a successful framework. *Communication and Collaboration Support Systems. TH a. TIEK Okada. Ohmsha*.

IETF (2001). Tags for the Identification of Languages, RFC 3066, January 2001. Available: http://www.ietf.org/rfc/rfc3066.txt

IETF (2009). Tags for Identifying Languages, BCP47, September 2009. Available: http://tools.ietf.org/search/bcp47

Im, D.-H. & Park, G.-D. (2014). Linked tag: image annotation using semantic relationships between image tags. *Multimedia Tools and Applications***,** 1-15.

Konkova, E., Göker, A., et al. (2014). Social Tagging: Exploring the Image, the Tags, and the Game. *Knowledge Organization*, 41.

Marlow, C., Naaman, M., et al. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. In Proceedings of the seventeenth conference on Hypertext and hypermedia, 2006. ACM, 31-40.

Michael, F. (2014). How many photos are uploaded to Flickr every day, month, year? [Online]. Available: http://www.flickr.com/photos/franckmichel/6855169886/ [Accessed 05 February 2014].

Nixon, L. (2013). Linked services infrastructure: a single entry point for online media related to any linked data concept. In Proceedings of the 22nd international conference on World Wide Web companion, 2013. International World Wide Web Conferences Steering Committee, 7-10.

Nogales, A., Sicilia, M.-A., et al. (2013). Exploring the Potential for Mapping Schema. org Microdata and the Web of Linked Data. Metadata and Semantics Research. Springer.

Sasaki, F. (2013). Metadata for the Multilingual Web. *Translation: Computation, Corpora, Cognition*, 3.

Sjekavica, T., Gledec, G., et al. (2014). Advantages of Semantic Web Technologies Usage in the Multimedia Annotation and Retrieval. *International Journal of Computers and Communications*, 8**,** 41-48.

W3C (2004). Resource Description Framework (RDF), W3C Recommendation, 10 February 2004. Available: http://www.w3.org/TR/rdf-schema/

W3C (2011). Scalable Vector Graphics (SVG) 1.1 (Second Edition), W3C Recommendation, 16 August 2011. Available: http://www.w3.org/TR/SVG11/

W3C (2012a). Media Fragments URI 1.0 (basic), 25/09/2012. Available: http://www.w3.org/TR/media-frags/

W3C (2012b). Ontology for Media Resources 1.0, W3C Recommendation, 09 February 2012. Available: http://www.w3.org/TR/mediaont-10/

W3C (2012c). OWL 2 Web Ontology Language Document Overview (Second Edition), W3C Recommendation, 11 December 2012. Available: http://www.w3.org/TR/owl2-overview/

W3C (2012d). RDFa Lite 1.1, W3C Recommendation, 07 June 2012. Available: http://www.w3.org/TR/rdfa-lite/

Yee, R. (2008). Understanding Tagging and Folksonomies. *Pro Web 2.0 Mashups: Remixing Data and Web Services***,** 61-75.