Enabling Advanced Context-Based Multimedia Interpretation Using Linked Data

Ruben Verborgh, Davy Van Deursen, Erik Mannens, and Rik Van de Walle

Ghent University – IBBT, ELIS – Multimedia Lab
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent, Belgium
{ruben.verborgh,davy.vandeursen,rik.vandewalle}@ugent.be
http://multimedialab.elis.ugent.be/

Abstract. Current search technologies can only harness the ever increasing amount of multimedia data when sufficient metadata exists. Several annotations are already available, yet they seldom cover all aspects. The generation of additional metadata proves costly; therefore efficient multimedia retrieval requires automated annotation methods. Current feature extraction algorithms are limited because they do not take context into account. In this article, we indicate how Linked Data can provide information that is vital to create an interpretation context. As a result, advanced interactions between algorithms, information and context will enable more advanced interpretation of multimedia data. Eventually, this will reflect in better search possibilities for the end user.

Keywords: Linked Data, multimedia interpretation, Semantic Web

1 Introduction

A tremendous increase in multimedia content production and consumption characterized the last decade and will continue to shape the Web for generations to come. The biggest challenge is to serve consumers the content they need in a convenient format and in a seemingly instantaneous way. The availability of different search and browsing options (e.g., keyword search, faceted search, and content-based search) enables efficient retrieval, yet these techniques require rich metadata [8] in order to offer advanced retrieval operations.

Metadata generation remains a tedious and often manual task that fortunately can be assisted by automatic algorithms. However, these algorithms produce inexact results with an often unknown reliability. In this paper, we argue that Linked Data can play a crucial role in the annotation and interpretation of multimedia data by assisting algorithms. Additionally, the results of this annotation task can be published back to the Semantic Web, contributing to the growth of the Linked Data Cloud. This feedback loop proves important for annotating future data, as depicted in Fig. 1.

The success rate of current search techniques strongly depends on the availability of correct annotations, as illustrated below.

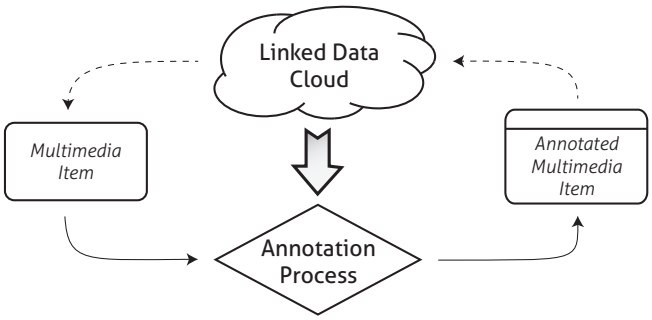

Fig. 1. Multimedia data annotation feedback loop

Keyword search Current search engines determine the content of an image by that of the surrounding text, which is inadequate for two reasons:

1. Although mostly a strong correlation between image contents and surrounding text exists, this is not always the case. Trivial examples account for this (e.g., try searching for *caption* or *alt*), as well as more realistic examples (e.g., a search for two people names often also returns other people).
2. Even a strong correlation does not imply understanding of the image contents. [6] lists some of the categories the ambiguous query *Washington* yields: persons, buildings, scenery, maps... In 2008, *Google* included the option to narrow the search to faces [4] – in fact a form of faceted search – but this does not solve the correlation problem nor help for the other categories.

Faceted search Faceted search or faceted browsing [16] enables searching through images in a faceted classification, e.g. simultaneous classifications in multiple categories. Evidently, this classification in categories requires annotations for each image, indicating its degree of membership to each of the categories.

Content-based search Content-based search techniques [13] measure the similarity of indexed items to sample items. However, images with very high visual similarity are for example unlikely to be of interest to the user, as he already disposes of the sample image. Instead, images with similar content can be relevant and could be explored with techniques such as faceted browsing, again requiring annotations. Conversely, content-based search can also be used to extend the other techniques, such as Google's *"find similar images"*.

Another major benefit of rich annotations is that they enable us to link imagery of subjects (such as George Washington and the White House) to their corresponding entities in the Linked Data Cloud, for example by using their *DBpedia* identifiers. Therefore, related images can show up on any search that results in entities from the Linked Data Cloud. That way, it would be possible to automatically provide keyword and faceted search on images, using the combined available knowledge about the entity.

2 Available annotations

Not all annotations have to be generated: some are already present upon acquisition of the multimedia item, others are added by consumers or people in a social network situation.

2.1 Metadata upon acquisition

The EXIF data format, describing metadata in digital photographs, is widely known [9]. It contains a lot of information that is mostly irrelevant for identifying photograph contents, such as camera type and aperture. Some recent cameras however, offer the option to include geographic coordinates in the EXIF data, thanks to a GPS receiver. The rise of versatile mobile devices that serve as phone, camera, and GPS navigator, have brought this into mainstream. These coordinates can be translated afterwards into a named location (country, city or even venue) and linked to the corresponding Linked Data entity. Furthermore, in combination with the EXIF timestamp, this information can be used to link the photograph to a particular event happening at that location and time [1, 7, 10].

2.2 Consumer and social metadata

Another common source of metadata, are annotations added by consumers upon publication of their material. Moreover, in some social networks, users are allowed to tag the content of others. Well-known examples are tags describing a variety of topics (place, time, scenery, people. . .). These tags however, suffer from the problem that they are informal. Therefore, several efforts to enhance their quality exist [3, 5].

Importantly, there is a growing tendency that stimulates users – perhaps unknowingly – to provide more formal tags. As examples, we cite the practices of person tagging, which creates a semantic link between a person and a photograph, and place tagging of photographs. Modern user interfaces have rendered this task intuitive for the majority of Internet users. Of course, the responsibility for the accuracy of these tags lies with these (sometimes anonymous) users, which is therefore arbitrary. It is tempting to assume that the formality of these tags is an indication of their reliability, which is a fallacy. In fact, their ease of creation can quickly lead to many formal but incorrect tags.

2.3 Production metadata

In production environments, there is an increasing tendency to generate metadata as part of the content creation process [2, 11, 12]. In each stage of the production process, relevant metadata is appended to the stage's end product and during the production process, metadata of previous stages can be reused. Ongoing research in this area reveals how these metadata can be turned into annotations that drive search applications.

3 Issues with contextless annotation

Automated multimedia analysis research has accomplished several milestones over the past decades in domains such as image feature extraction. While advanced techniques for various tasks exist, such as face recognition algorithms, most of them remain error-prone.

Another problem is that the output of these algorithms is often not formalized. For instance, a face recognition algorithm may recognize a certain face in an image, but does not output an entity that is linked to its corresponding Linked Data counterpart. A solution to this issue has been proposed in [14], where RDF is proposed as input and output for multimedia processing algorithms.

Still, the main issue remains the missing ability of feature extraction algorithms to collaborate on the annotation of a given multimedia item. Each algorithm is highly specialized, which is both its greatest strength and its greatest weakness. This degree of specialization comes indeed at the expense of losing an overview on the item under annotation. Humans, in contrast, possess the remarkable ability to shift rapidly between different levels of abstraction. This is why we can recognize faces in context, while we are not able to do so without.

Fig. 2 shows a clear example of a photograph of a face with and without context. It is evident that no human and no algorithm now and in the future will be able to recognize the person depicted in the photograph on the left with an acceptable certainty, given no additional information. Our human ability to unconsciously zoom in and out between different detail levels, provides us exactly with this information, given the photograph on the right. Upon recognizing one person as *Hillary Clinton*, we instantaneously realize the other person is *Bill Clinton* with a fairly high certainty.

This clearly the importance of context-awareness for feature extraction algorithms, a task that is hard because of the involved complexity. After all, it is impossible for algorithms to anticipate on all possible context parameters. Therefore, we should look into platforms that are able to integrate algorithms and context by means of knowledge and reasoning [15].

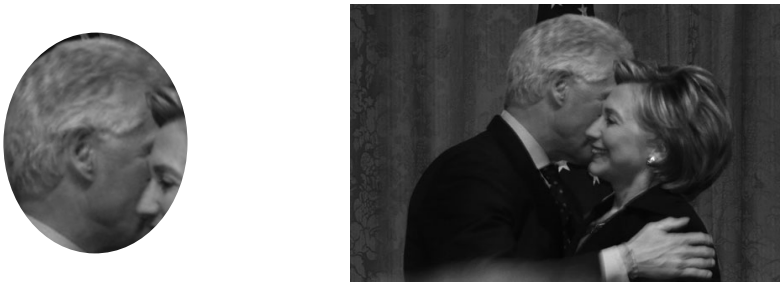

Fig. 2. A photograph of a face without context (left) and with context (right).

4 Creating annotation context with Linked Open Data

4.1 Knowledge-driven annotation

Information that connects various concepts in different ways is readily available on several semantic data sources, including DBpedia. We can imagine an annotation platform consulting these data sources to retrieve information to either complete or validate found results. The annotation process could in fact entirely take place on the Semantic Web and consist of:

- general and specific knowledge from the Linked Data Cloud;
- intelligent services such as feature extraction algorithms;
- a Semantic Web reasoner applying rule-based knowledge to entities.

The general knowledge could drive the process, deriving concrete knowledge about a specific multimedia item.

Linked Data knowledge On the one hand, automated interpretation requires general knowledge about annotations, consisting of both ontologic and rule-based knowledge. In the case of images, an example of the former is *"images can contain regions that depict a face"*, and an example of the latter *"if an image contains a region that depicts a person's face, then the image depicts the person"*. This general knowledge helps the platform decompose the task of interpreting an image into smaller subtasks. On the other hand, specific knowledge about concrete topics is essential to provide a context for the item under interpretation. Their presence proves vital for complex reasoning schemes. Examples include ontological and instance-based knowledge about people and interpersonal relations.

Intelligent services Automated annotation algorithms could interact with Semantic Web knowledge if we approach algorithms as Semantic Web services [14]. We can describe algorithms as regular Web services and invoke them similarly. Each input and output parameter thus gains semantic value, which enables us to form complex service compositions. Together with general and specific Linked Data knowledge, they enable diverse intelligent interactions with the context.

As an example, we consider a face recognition algorithm that operates on Fig. 2. The visible features of Bill Clinton alone will be insufficient to achieve successful recognition. The nearby presence of Hillary Clinton could substantiate the assumption that the other person relates to her in some way. A search on the Linked Data Cloud for relatives, friends and co-workers would yield a list of possibilities which we can pass in a semantic way to the recognition service. As a result, the algorithm could take these suggestions into consideration and limit its search space, significantly increasing the odds of arriving at the correct conclusion. We could then further cross-check the solution by checking with available metadata, such as GPS location and time, and maybe retrieve the event at which the photograph was taken.

Reasoning Semantic Web reasoning plays a crucial role in this application domain. In the above example, we intuitively assumed that the simultaneous depiction of two persons implies some kind of connection between them. This knowledge, which can be derived for instance by statistic methods, needs to be available formally. A reasoner is then able to instantiate this knowledge on concrete entities, e.g., to retrieve people connected to Hillary Clinton.

An important topic here is dealing with imperfections. Most data in the Linked Data Cloud are represented as absolute facts without association about possible vagueness or uncertainty. While this is sometimes applicable and even desired, automated interpretation inherently needs to deal with predictions and assumptions. As a consequence, these imperfections need to propagate when reasoning on intermediate data.

4.2 Feedback to the Linked Data Cloud

Furthermore, results from an annotation process can also be pushed back to the Linked Data Cloud. As indicated, multimedia searches are an interesting and important application. Feedback enables other useful applications as well, e.g.:

- New data could be inferenced based on the generated annotations. For example, the fact that several people were at a given place at a given time could indicate their attendance of an event.
- Feedback data could enhance future annotation tasks, for example by means of statistics of which people appear together frequently.

In general, feedback will potentially link many different concepts, contributing to the coherence of the Linked Data Cloud while at the same time complementing it with multimedia data.

4.3 Platform model

Fig. 3 displays a model for our platform, clearly indicating the interaction between various building blocks and marking the importance of Linked Data for the interpretation process.

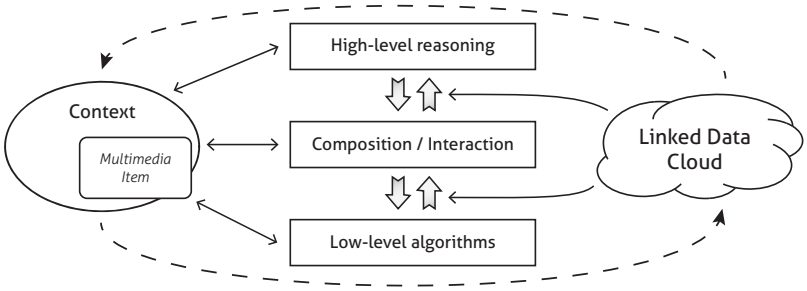

Fig. 3. Use of linked data in a context-aware interpretation process

5 Concluding remarks and future work

In this article, we outlined how Linked Data forms an cornerstone of a context-based multimedia interpretation process. The incorporation of Linked Data creates a holistic view that integrates the different aspects of annotating. Several important topics for future research emerge, including:

- the interaction of feature extraction algorithms with Linked Data;
- the representation of uncertainties associated with multimedia interpretation;
- the feedback of the interpretation process towards the Linked Data Cloud.

All these topics illustrate the enormous potential of Linked Data for advanced automated multimedia interpretation.

Acknowledgments

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research Flanders (FWO-Flanders), and the European Union.

References

1. Chen, C., Oakes, M., Tait, J.: A location data annotation system for personal photograph collections: Evaluation of a searching and browsing tool. In: International Workshop on Content-Based Multimedia Indexing, 2008. CBMI 2008. (Jun 2008)
2. Debevere, P., Van Deursen, D., Van Rijsselbergen, D., Mannens, E., Matton, M., De Sutter, R., Van de Walle, R.: Enabling Semantic Search in a News Production Environment. Proceedings of the 5th International Conference on Semantic and Digital Media Technologies (SAMT 2010) (Dec 2010)
3. Lee, S., De Neve, W., N Plataniotis, K., Man Ro, Y.: MAP-based image tag recommendation using a visual folksonomy. Pattern Recognition Letters 31(9), 976–982 (Jan 2010), http://dx.doi.org/10.1016/j.patrec.2009.12.024
4. O'Malley, S.: New search-by-style options for Google Image Search. Official Google Blog (Dec 2008), http://googleblog.blogspot.com/2008/12/new-search-by-style-options-for-google.html
5. Overell, S., Sigurbjörnsson, B., Van Zwol, R.: Classifying tags using open content resources. Proceedings of the Second ACM International Conference on Web Search and Data Mining pp. 64–73 (2009)
6. Rahurkar, M., Tsai, S., Dagli, C., Huang, T.: Image Interpretation Using Large Corpus: Wikipedia. Proceedings of the IEEE 98(8), 1509 – 1525 (2010), http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5484723
7. Sarin, S., Nagahashi, T., Miyosawa, T., Kameyama, W.: On automatic contextual metadata generation for personal digital photographs. In: The 9th International Conference on Advanced Communication Technology (Feb 2007)
8. Smith, J., Schirling, P.: Metadata standards roundup. IEEE MultiMedia (Jan 2006), http://www.computer.org/portal/web/csdl/doi/10.1109/MMUL.2006.34

9. Tešić, J.: Metadata practices for consumer photos. IEEE MultiMedia (Jan 2005), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1490501
10. Troncy, R., Malocha, B., Fialho, A.: Linking events with media. Proceedings of the 6th International Conference on Semantic Systems (Jan 2010), http://portal.acm.org/citation.cfm?id=1839759
11. Van Rijsselbergen, D., Van De Keer, B.: Movie script markup language. Proceedings of the 9th ACM symposium on Document engineerin (Jan 2009), http://portal.acm.org/citation.cfm?id=1600193.1600231
12. Van Rijsselbergen, D., Verwaest, M., Van De Keer, B., Van de Walle, R.: Introducing the Data Model for a Centralized Drama Production System. IEEE International Conference on Multimedia and Expo, 2007 (Jan 2007), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4284725
13. Veltkamp, R., Tanase, M.: A survey of content-based image retrieval systems. Content-based image and video retrieval (Jan 2002), http://www.cs.uu.nl/groups/MG/multimedia/publications/art/socbirs02.pdf
14. Verborgh, R., Van Deursen, D., De Roo, J., Mannens, E., Van de Walle, R.: SPARQL Endpoints as Front-end for Multimedia Processing Algorithms. In: Proceedings of the Fourth International Workshop on Service Matchmaking and Resource Retrieval in the Semantic Web at the 9th International Semantic Web Conference (ISWC 2010) (Nov 2010)
15. Verborgh, R., Van Deursen, D., Mannens, E., Poppe, C., Van de Walle, R.: Enabling Context-aware Multimedia Annotation by a Novel Generic Semantic Problem-Solving Platform. Multimedia Tools and Applications special issue on Multimedia and Semantic Technologies for Future Computing Environments (2011)
16. Yee, K., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. Proceedings of the Special Interest Group on Computer–Human Interaction (Jan 2003), http://portal.acm.org/citation.cfm?id=642611.642681