Augmenting Europeana Content with Linked Data Resources

Bernhard Haslhofer
University of Vienna
Dept. of Distributed and Multimedia Systems
Vienna, Austria
haslhofer@cs.univie.ac.at

Elaheh Momeni
University of Vienna
Dept. of Distributed and Multimedia Systems
Vienna, Austria
momeni@cs.univie.ac.at

Manuel Gay
Austrian Institute of Technology
Vienna, Austria
manuel.gay@ait.ac.at

Rainer Simon
Austrian Institute of Technology
Vienna, Austria
rainer.simon@ait.ac.at

ABSTRACT

Annotations allow end users to augment digital items with information, which can then be exploited for search and retrieval. We are currently extending Europeana, a platform which links to millions of digital items in European institutions, with an annotation mechanism that exposes annotations as linked data and enriches newly created annotations with links to contextually relevant resources on the Web. In two demos we showcase how we integrated that kind of content augmentation into two clients that allow users to annotate videos and historic maps.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous

General Terms

Design, Human Factors, Languages

Keywords

Linked Data, Annotations, Semantic Web

1. INTRODUCTION

Many cultural institutions are currently investigating how to integrate the end users' knowledge with existing digital curation processes. Annotations are an obvious and well-understood solution (cf. [3]) for that. Therefore, they will also play a major role in the future development of Europeana (`http://www.europeana.eu`) a platform that links to millions of digital items (texts, images, videos, sounds) in cultural organizations all over Europe. Web-based annotation clients will allow users to extend existing bibliographic information about digital items by means of textual notes and comments. The underlying indexing infrastructure can exploit that information for search and retrieval.

Our goal is to link Europeana digital items with other resources on the Web. In a first prototype, we already integrated our LEMO multimedia annotation framework [2] with the existing Europeana infrastructure. All annotation resources managed by this framework have URIs assigned, are exposed as dereferencable resources on the Web, and deliver RDF when they are resolved. For representing annotation data, we developed the LEMO vocabulary (`http://lemo.mminf.univie.ac.at/annotation-core#`). It reuses and refines terms from the Annotea [4] vocabulary and defines several media-specific extension (e.g., images, video, audio). Demos of the various clients developed on-top-of LEMO are available at: `http://dme.arcs.ac.at/annotation/annotation-showcase.html`

Now we are investigating how to integrate resource linkage with the process of annotating digital items in order to augment Europeana content with contextually relevant information. We expect to retrieve links with high accuracy because we have human end-users interacting with annotation systems. The users know the semantics of the annotated resources and understand the meaning of the created annotation content within the context of their annotations. Automated linkage frameworks can hardly do better.

We implemented two demos that showcase our ideas how we can augment Europeana content with Linked Data resources by means of annotations: the first demo allows users to annotate **videos**. The systems proposes relevant links to DBpedia [1] resources and prompts the user to adjudicate their validity. The information behind those linked resources enables contextual and multilingual search for videos. In the second demo, users can annotate **historic maps**. The system also suggests contextually relevant links, which the user can confirm or decline interactively. In addition, the tool provides functionality for map geo-referencing. The system can exploit the geo-referencing information to (1) enable geographical search and retrieval of annotations, and (2) provide "spatial context" to arbitrary searches.

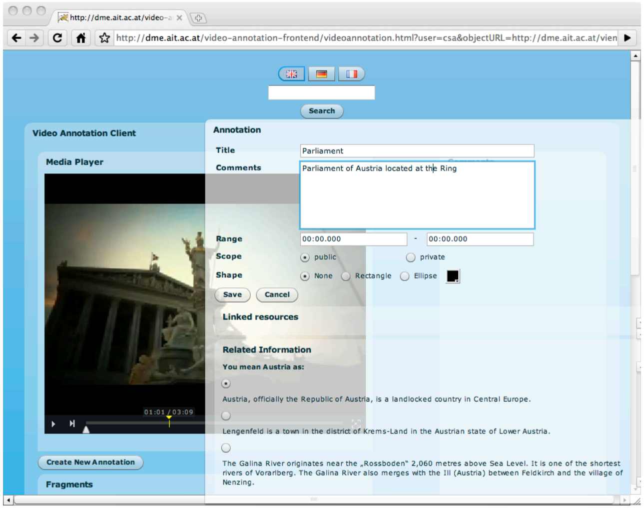

Figure 1: Video annotation client.

2. VIDEO ANNOTATION

2.1 Goal and User Benefit

The video annotation tool we are currently building for Europeana showcases how users can augment existing online videos with related resources on the Web. This augmentation happens on-the-fly while the users are writing their annotations. It relies on user-feedback to verify the semantic validity of a link before it becomes part of the annotation. The goal is to exploit linked resource descriptions in the underlying search and retrieval infrastructure.

In Figure 1, we see an example annotation on a video sequence that shows the Austrian parliament in Vienna. A user has entered a simple note "The Parliament of Austria located at the Ring" and the system proposes a ranked list of potentially relevant DBpedia resources. Before an annotation is saved, the proposed resource can be marked as relevant or non-relevant.

In the context of Europeana, the augmentation of content with links to external Linked Data resources brings two main benefits for the end user. First, a large portion of Europeana contents, especially multimedia contents such as video and audio, have insufficient metadata assigned. Often this is because institutions do not have the capacities to manually catalogue the vast amount of digital resources they are hosting. With user-contributed annotations we widen that pool of metadata. By including links in annotations, we enable the underlying indexing infrastructure to answer queries based on contextually relevant information. From DBpedia, for instance, we can include the short descriptions (abstracts) of the resources that are related to the entities ("Austria", "Ring") in the annotation text. Second, by including translations of resource descriptions in the index, we enable a basic form of multilingual retrieval for multimedia resources. By linking to DBpedia we can resort to 91 different translations for all concepts that are currently described in the Wikipedia. If both concepts in the above annotation example are linked to the right DBpedia resources, users can find the video in their search results if they use a German query, such as "Österreich" or "Ringstraße".

Figure 2: Map annotation client.

2.2 Technical Details

For enriching textual annotations with links to relevant resources, the video client makes use of our *Semantic Enrichment Service*, which is a separate component that can be reused in any other annotation client. It provides a simple HTTP interface that accepts plain annotation text and returns a ranked list of potentially relevant resources.

The service consists of two main components:

- *NEREngine*: an interface to different, configurable implementations of Named Entity Recognition engines such as Gate[1] or OpenCalais[2]. It takes an annotation string as input and returns the recognized named entities sorted by their type (e.g., Place).
- *LODIndex*: a component that periodically indexes resources of a configured type (e.g., `dbpedia-owl:Place`) from the SPARQL interface of a given data source (e.g., `http://dbpedia.org/sparql`). By default it considers for each resource the values of all `rdfs:label` properties in all available languages. Thus, no matter what the input language of a passed named entity is, the service always returns matching resource URIs.

Internally, the service first sends the annotation string to the NEREngine and then retrieves a ranked list of relevant links from the LODIndex by passing the recognized named entities. First experiments show that organizing the indexed resources by their types and matching them with recognized entity types reduces semantic ambiguities and increases the quality of the proposed resources.

3. MAP ANNOTATION

3.1 Goal and User Benefit

Our second demonstrator shows how we integrated linkage with the annotation of digitized antique maps. The goal is to exploit semantic linking in conjunction with map georeferencing to provide "spatial context" to search and retrieval tasks.

Figure 2 shows an example screenshot. A user has zoomed to a region on the map and added an annotation to a specific

[1] `http://gate.ac.uk/sale/tao/index.html`
[2] `http://www.opencalais.com/`

area, using shape drawing tools provided by the front-end. As in the case of video annotation, the annotation tool starts to produce suggestions for potentially relevant linked data resources. In the user interface, these suggestions are superimposed on the map as a tag cloud. Each tag in the cloud represents one link suggestion. New tags are added to the tag cloud as additional suggestions become available, e.g., while the user is typing annotation text. At any time, the user can choose to accept a link suggestion simply by clicking on the corresponding tag.

An obvious difference between maps and other types of media, such as video, is that maps are inherently concerned with the depiction of geography. Therefore, when a user draws a shape on the map and adds an annotation to it, it is likely that there is a strong relation between the annotation, and the geographical area covered by the shape. This, for example, is the case in Figure 2, where the user has placed an annotation on a place name inscription near the Strait of Gibraltar, and made an explanatory note on the origin of the place name ("The Pillars of Hercules"). In order to allow the system to compute the geographical coordinates for the shape, the map annotation tool includes functionality for geo-referencing: Geo-referencing establishes an explicit correspondence between the map-image's pixel coordinates and a defined geographical coordinate system such as the well-known WGS84 (or "GPS" latitude/longitude) coordinates. The tool supports this task by allowing users to create so-called "control points". Control points are identifiable points (e.g. landmarks, cities, natural formations, etc.) on the map to which the geographical coordinates are known. These control points can then be used as a basis to translate between both coordinate spaces.

As soon as the map has been geo-referenced with control points, the tool can provide two main benefits to the end user. First, annotations become geographically searchable: By including the annotations' spatial footprints into the index, the user can find annotations which pertain to a certain place, country or region. Furthermore, the underlying indexing infrastructure can use the footprint information to include links to geographical linked data sources such as Geonames.org. This provides further contextual metadata, in addition to the contextual links derived from the annotation's text using NER. Second, the system can potentially provide "spatial context" to any (keyword) search result: For example, when a user searches for the term "Pillars of Hercules", the system will be able to indicate the region around the Strait of Gibraltar as relevant to this query.

3.2 Technical Details

The browser-based user interface is built on the open source OpenLayers JavaScript Web mapping library. Similar to popular Web map services like Google Maps or Yahoo! Maps, the user interface provides a full-screen drag- and zoom-able representation of the map.

To produce contextually relevant link suggestions based on the annotation text, the Map Annotation Showcase relies on the same semantic enrichment service as the the video annotation tool.

Geo-referencing is based on computation of an affine transformation from the closest neighbor control points. Advantages of this approach are its computational simplicity, as well as the fact that it can be applied to any type of map, irrespective of scale or projective properties. On the downside, geo-referencing through control points is an approach that is inherently based on interpolation. Therefore, results will always represent approximations, and accuracy may vary from map to map. Nonetheless, trials with scans of antique maps from the 16th century have so far yielded promising results [6]. A more in-depth discussion of the technical details of the map annotation tool and the tag-cloud user interface is also presented in [7].

4. SUMMARY AND FUTURE WORK

Annotations are a well-understood mechanism to augment digital items with additional information. With our LEMO framework we can already expose that information as linked data on the Web. By integrating linkage with annotation processes and interfaces we can augment digital items with relevant resources. The linked resource descriptions can then be exploited for search and retrieval.

In our future work, we will concentrate on the improvement of our semantic enrichment service and will further investigate the possibilities this brings for multi-lingual retrieval. Another goal is to enable interoperability with other annotation systems by supporting the model currently developed by the Open Annotation Collaboration (OAC) [5].

Acknowledgments

The work has partly been supported by the European Commission as part of the eContentplus program (EuropeanaConnect).

5. REFERENCES

[1] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. Dbpedia - a crystallization point for the web of data. *J. Web Sem.*, 7(3):154–165, 2009.

[2] B. Haslhofer, W. Jochum, R. King, C. Sadilek, and K. Schellner. The lemo annotation framework: weaving multimedia annotations with the web. *Int. J. on Digital Libraries*, 10(1):15–32, 2009.

[3] J. Hunter. *Collaborative Semantic Tagging and Annotation Systems*, volume 43, chapter 2. Information Today Inc., 2009.

[4] J. Kahan and M.-R. Koivunen. Annotea: an open RDF infrastructure for shared Web annotations. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 623–632, New York, NY, USA, 2001. ACM Press.

[5] R. Sanderson and H. V. de Sompel. Making web annotations persistent over time. In *Joint JCDL/ICADL International Digital Libraries Conference (JCDL 2010)*, 2010. forthcoming.

[6] R. Simon, J. Korb, C. Sadilek, and M. Baldauf. Explorative User Interfaces for Browsing Historical Maps on the Web. In *5th International Workshop on Digital Approaches in Cartographic Heritage*, Vienna, Austria, 2010.

[7] R. Simon, C. Sadilek, J. Korb, M. Baldauf, and B. Haslhofer. Tag Clouds and Old Maps: Annotations as Linked Spatiotemporal Data in the Cultural Heritage Domain. In *Workshop On Linked Spatiotemporal Data 2010*, Zurich, Switzerland, 2010.