SemWebVid - Making Video a First Class Semantic Web Citizen and a First Class Web Bourgeois[*]

Thomas Steiner,

Google Germany GmbH, ABC-Straße 19, 20354 Hamburg, Germany
tsteiner@{google.com, lsi.upc.edu}[¶]

Abstract. SemWebVid[1] is an online Ajax application that allows for the automatic generation of Resource Description Framework (RDF) video descriptions. These descriptions are based on two pillars: first, on a combination of user-generated metadata such as title, summary, and tags; and second, on closed captions which can be user-generated, or be auto-generated via speech recognition. The plaintext contents of both pillars are being analyzed using multiple Natural Language Processing (NLP) Web services in parallel whose results are then merged and where possible matched back to concepts in the sense of Linking Open Data (LOD). The final result is a deep-linkable RDF description of the video, and a "scroll-along" view of the video as an example of video visualization formats.

Keywords: RDF, LOD, Linked Data, Semantic Web, NLP, Video

1 Introduction

Over recent years the use of Resource Description Framework (RDF) in documents has gained massive popularity with even mainstream media[2] picking up stories of big companies deploying RDF on their Web presence. However, these efforts have mainly concentrated on textual documents in order to annotate concepts like shop opening hours, prices, or contact data. Far fewer occurrences can be noted for RDF video description on the public Web. Related efforts are automatic video content extraction, or the W3C Ontology for Media Resource.

The development of SemWebVid was driven by the following objectives:

[*] This work is partly funded by the EU FP7 I-SEARCH project (project reference 248296)

[¶] The author is currently a PhD student at Universitat Politècnica de Catalunya, Department LSI, Campus Nord, Edifici Omega, Jordi Girona 1-3, 08034 Barcelona, Spain. We would like to thank Michael Hausenblas from DERI Galway for the review of this work and paper.

[1] Live demo at http://tomayac.com/semwebvid/, username: iswc2010, password: iswc2010

[2] http://www.nytimes.com/external/readwriteweb/2010/07/01/01readwriteweb-how-best-buy-is-using-the-semantic-web-23031.html

- Improve **searchability** of video content by extraction of contained entities and disambiguation of those entities (for queries like *videos of Barack Obama where he talks about Afghanistan while being abroad*).
- Enable **graphical representations** of video content through symbolization of entities (for e.g. video archives of keynote speeches where one could *graphically skim through long video sections at a glance*).

These goals can be reached through RDF video descriptions and we thus developed SemWebVid to create RDF video descriptions in a potentially automatable way based on live data found on YouTube.

2 SemWebVid Dataflow

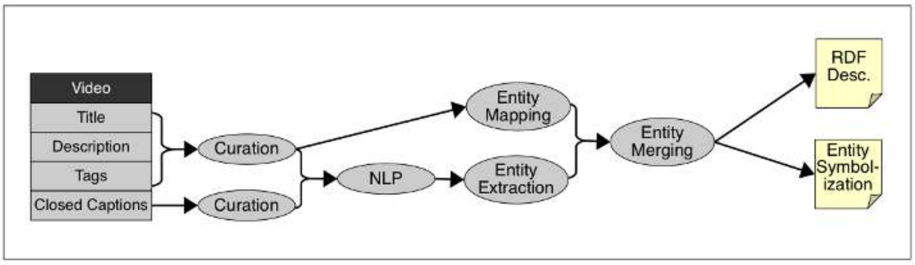


Fig. 1. SemWebVid Dataflow Diagram

The raw data for the RDF video descriptions consist of the beforementioned two pillars: the user-generated video title, video (plaintext) description, and tags on the one hand, and the user- or auto-generated[3] closed captions on the other.

3 Curation of Raw Data

Before the entity mapping and extraction steps, the raw data need to be curated. While the video titles are typically trouble-free as they are usually very descriptive, the main problem with the plaintext video descriptions is that sometimes they get abused for non-related spam-like messages or comments rather than providing a proper summary of the video content. Unfortunately this is hard to detect, so in the end we decided to simply use them as is. With regards to tags the main issue are different tagging styles. As an example see potential tags for the concept of the person Barack Obama:

- `"barack", "obama"` (all words separated, 2 tags)
- `"barack obama"` (space-separated, 1 tag)
- `"barackobama"` (separate words concatenated, 1 tag)

The first split style is especially critical if complete phrase segments are expressed in tag form:

[3] YouTube allows for auto-generation of closed captions through speech recognition: http://youtube-global.blogspot.com/2010/03/future-will-be-captioned-improving.html

- "one", "small", "step", "for", "a", "man"

For our demonstration we use an API from Bueda[4] in order to split combined tags into their components and try to make sense of split tags. We use Common Tag to represent tags in the application.

The curation step for closed captions mainly consists of removing speaker and hearable events syntax noise from the plaintext contents, and obviously the cues (time markers for each caption). This can be easily done using regular expressions, the syntax being a variation of ">>Speaker:" and "[Hearable Event]".

4 Entity Extraction and Mapping

We try to map the list of curated tags back to entities using plaintext entity mapping Web services[5] from DBpedia [1], Sindice [2], Uberblic, and Freebase. This works quite well for very popular tags (samples below from the DBpedia URI Lookup Web service, all results are prefixed with http://dbpedia.org/resource/"):

- "barack obama" => Barack_Obama

It somewhat succeeds for very generic tags (though with obvious ambiguity issues):

- "obama" => Obama,_Fukui

It fails for specific tags ("ggj09" was a tag for the event "Global Game Jam 2009"):

- "ggj09" => N/A

It is thus very important to preserve **provenance** data in order to judge and estimate the quality of the mapped entities. With regards to the curated closed captions, description, and title we work with NLP Web services[6], namely OpenCalais, Zemanta, and AlchemyAPI. For the test cases we used (famous speeches, keynotes) results were relatively accurate from our judgments. In a final step the detected entities are merged, and a symbolization for each entity gets retrieved by means of a heuristic approach, including Google image search.

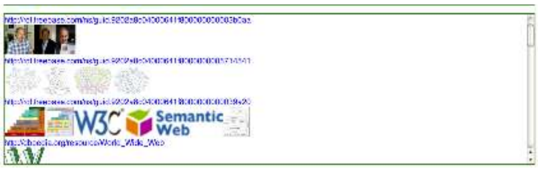


Fig. 2. Graphical symbolizations of several entities (TimBL, Semantic Web, etc.)

5 Description of the SemWebVid Demonstration

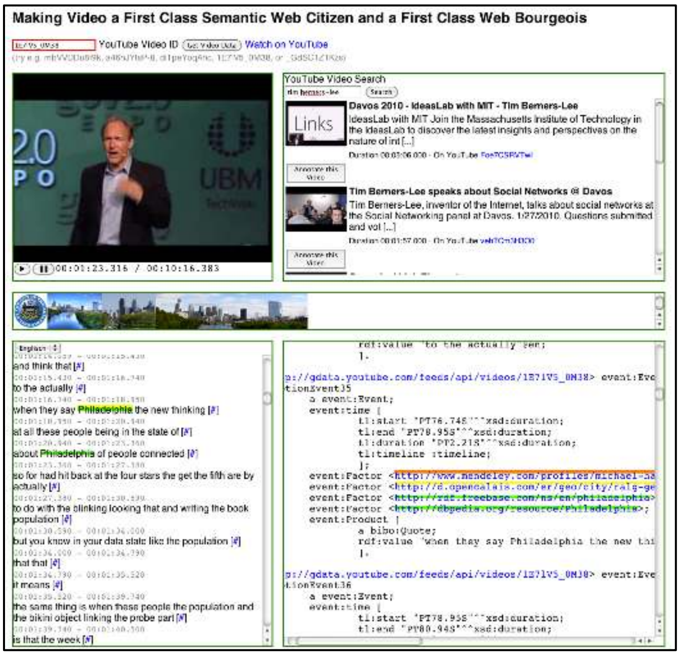
SemWebVid is designed to be an online Ajax application for interactive use. Unfortunately the terms and conditions of some of the NLP Web services involved do

[4] http://www.bueda.com/developers

[5] http://platform.uberblic.org, http://www.freebase.com, http://sindice.com, http://dbpedia.org

[6] http://opencalais.com, http://zemanta.com, http://alchemyapi.com

not allow for a SemWebVid API, however, due to its design both on-the-fly RDF description generation and permanent linking to previous descriptions are possible.



Making Video a First Class Semantic Web Citizen and a First Class Web Bourgeois

1E71V5_0M38 YouTube Video ID Get Video Data Watch on YouTube

YouTube Video Search

Davos 2010 - IdeasLab with MIT - Tim Berners-Lee

IdeasLab with MIT Join the Massachusetts Institute of Technology in the IdeasLab to discover the latest insights and perspectives on the nature of int [...]

Tim Berners-Lee speaks about Social Networks @ Davos

Tim Berners-Lee, inventor of the Internet, talks about social networks at the Social Networking panel at Davos. 1/27/2010. Questions submitted and vot [...]

00:01:23.316 / 00:10:16.383

```
and think that [#]
00:01:15.430 - 00:01:16.340
to the actually [#]
00:01:16.340 - 00:01:18.550
when they say Philadelphia the new thinking [#]
00:01:18.550 - 00:01:20.540
at all these people being in the state of [#]
00:01:20.540 - 00:01:23.360
about Philadelphia of people connected [#]
00:01:23.360 - 00:01:27.380
so for had hit back at the four stars the get the fifth are by actually [#]
00:01:27.380 - 00:01:30.590
to do with the blinking looking that and writing the book population [#]
00:01:30.590 - 00:01:34.000
but you know in your data state like the population [#]
00:01:34.000 - 00:01:34.790
that that [#]
00:01:34.790 - 00:01:35.520
it means [#]
00:01:35.520 - 00:01:39.740
the same thing is when these people the population and the bikini object linking the probe part [#]
00:01:39.740 - 00:01:40.300
is that the week [#]
```

```
        rdf:value "to the actually"@en;
        ].

p://gdata.youtube.com/feeds/api/videos/1E71V5_0M38> event:Eve
tionEvent35
    a event:Event;
    event:time [
        tl:start "PT76.74S"^^xsd:duration;
        tl:end "PT78.95S"^^xsd:duration;
        tl:duration "PT2.21S"^^xsd:duration;
        tl:timeline :timeline;
        ];
    event:Factor <http://www.mendeley.com/profiles/michael-ha
    event:Factor <http://d.opencalais.com/er/geo/city/ralg-ge
    event:Factor <http://rdf.freebase.com/ns/en/philadelphia>
    event:Factor <http://dbpedia.org/resource/Philadelphia>;
    event:Product [
        a bibo:Quote;
        rdf:value "when they say Philadelphia the new thi
        ].

p://gdata.youtube.com/feeds/api/videos/1E71V5_0M38> event:Eve
tionEvent36
    a event:Event;
    event:time [
        tl:start "PT78.95S"^^xsd:duration;
        tl:end "PT80.94S"^^xsd:duration;
```

Fig. 3. SemWebVid screenshot showing Tim Berners-Lee's infamous potato chips speech at gov2.0 Expo 2010. Below the video box the concept of the city of Philadelphia is symbolized. The left lower box shows the closed captions directly, the right box the RDF description.

6 Conclusion and Future Work

While we are not the first[7] to connect RDF (and thus Linked Data) with video, SemWebVid's contribution is to present an automatic text-based way to generate RDF video descriptions. Future work is among other things to determine whether the expected searchability improvements pay off the high processing efforts.

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: Proc. of 6th Int. Semantic Web Conf., 2nd Asian Semantic Web Conf., November 2008, pp. 722–735.
2. Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., Tummarello, G.: Sindice.com – A Document-oriented Lookup Index for Open Linked Data. International Journal of Metadata, Semantics and Ontologies, 3 (1), 2008.

[7] Sack, H: http://www.hpi.uni-potsdam.de/fileadmin/hpi/FG_ITS/papers/Harald/DSMSA09.pdf