Survey of Semantic Media Annotation Tools for the Web: Towards new Media Applications with Linked Media

Lyndon Nixon[1] and Raphaël Troncy[2]

[1] MODUL University, Vienna, Austria
lyndon.nixon@modul.ac.at
[2] EURECOM, Sophia Antipolis, France
raphael.troncy@eurecom.fr

Abstract. Semantic annotation of media resources has been a focus in research since many years, the closing of the "semantic gap" being seen as key to significant improvements in media retrieval and browsing and enabling new media applications and services. However, current tools and services exhibit varied approaches which do not easily integrate and act as a barrier to wider uptake of semantic annotation of online multimedia. In this paper, we outline the Linked Media principles which can help form a consensus on media annotation approaches, survey current media annotation tools against these principles and present two emerging toolsets which can support Linked Media conformant annotation, closing with a call to future semantic media annotation tools and services to follow the same principles and ensure the growth of a Linked Media layer of semantic descriptions of online media which can be an enabler to richer future online media services.

1 Introduction

The amount of non-textual content being shared via the Internet is growing at a massive scale, yet the original Web technology around hypertext is proving too limited for the needs of online users to find and re-use online media easily. Metadata has become a vital component of media retrieval but the vast majority of online media assets have very limited metadata attached to them, e.g. Google Image search works primarily on the text surrounding media embedded into web pages. Online media collections may expose some basic properties following the Dublin Core model, i.e. a title, a description or some keywords, while Social Web media sharing sites tend towards open tagging of their content (folksonomies). Some additional metadata is typically included in the creation process, such as digital cameras capturing date/time and location of a photo. However, this still proves far from ideal for supporting rich and innovative media-centric applications where there may be complex queries as well as a need for meaningful organisation of relevant media items for browsing or re-use. With the current emergence of the so-called *Visual Web* - image-centric web sites (e.g. Pinterest,

Instagram, Tumblr) - and presumably the explosion very soon of an *Audiovisual Web* (Vine is a precursor of this but expect more professional content) the lack of actionable metadata is a major blocker to new innovation around retrieval and organisation of online media content.

On the other hand, the research community has been talking about semantic multimedia as a solution to richer media descriptions since over a decade - the basic idea being to move from basic metadata schemas to machine processable models with well defined properties and values, and provide sets of tools for manual or automatic creation of that metadata [3, 11, 8]. Research project showcases have repeatedly demonstrated improved media retrieval, faceted browsing or multimedia presentation as a result, e.g. [14]. However it has been typical of independent research activities to each choose different metadata models and vocabularies for the "semantic annotation" and build tools which were largely standalone and not connected to the rest of the media workflow. Furthermore, these models and tools for creating them have barely been applied to online media resources, and when they have been, separate approaches have not been well aligned meaning that semantic media demonstrators remain focused on individually annotated online collections instead of approaching online media retrieval and organisation at a truly Web scale. As a result, semantic media annotations remain at a small scale and in heterogenous formats today while the vast amounts of online media content remains unannotated or only attached to non-semantic metadata information.

In this paper, we want to address this situation. The authors believe we stand before a significant opportunity to make semantic annotations of online media a fundamental part of the Web fabric, enabling better Web media retrieval, reuse and re-mixing through new online services and applications able to leverage the semantic annotations. The timing is significant because the Visual Web is already here[3], and the Audiovisual Web (also driven by the shift of TV to digital and online) is only a matter of time. It needs to be driven by online tools and services which exhibit a consensus on interoperable input and output data formats, vocabularies and concept identifiers. Thus, in Section 2 we will propose a set of principles for Web-friendly semantic annotation of media which we call "Linked Media" and show how emerging Web specifications can address them. In Section 3, we look at currently available semantic multimedia annotation tools for the Web and compare these against those principles of "Linked Media", highlighting two recent developments for online annotation which conform to the Linked Media vision. Section 4 briefly outlines an example of new media applications potentially enabled by Linked Media. We conclude in Section 5 with a brief look into the future where significant amounts of online media are being semantically annotated and those annotations are being published on the Web, and call for future research work on semantic multimedia annotation to align itself with Linked Media.

---

[3] *The Triumph of the Visual Web*, `http://readwrite.com/2013/12/18/visual-web-2013-trends`

2 Principles for Linked Media

Linked Media is a manifesto for a flexible, interoperable set of specifications for semantic media annotation, linking and presentation. It addresses metadata models, vocabularies for concepts, syntax for media fragments and Web based publication for subsequent retrieval and further processing. It has become possible now thanks to a consensual specification of how to refer to media fragments on the Web, annotate media and unambiguously refer to concepts and has become necessary with the scale of online media, making retrieval and re-use a pressing challenge.

Non-textual media such as audio-visual streams are not well integrated into the Web, where hypertext has traditionally been the baseline and thus links are made via textual anchors between Web pages which may embed non-textual media. The core problem is that there is not yet a Web-wide shared approach to annotating Web media such that the media metadata could consequently be found and used for linking across media collections, linking into and out of online documents, or also to generate links through the growing Web of Data. As such, we have identified core research issues around media on the Web today which are vital to be taken into account if, we believe, Web media is to be fully integrated into the Web of linked content and services in the future:

1. **Web media needs to be annotated in terms of its online parts, both along spatial and temporal dimensions**, since it is too imprecise to say that an atomic media item is about a concept X where that concept may only relate to a (small) part of the media. For a long time, we have lacked a standard means to refer to spatial or temporal segments of media on the Web.

2. **Web media needs to be annotated with terms which represent a shared understanding of a domain or identification of a thing.** When these terms are provided in a machine-understandable manner, we can say that they are drawn from an ontology. Providing ontologies and means to describe things using an ontology has been the activity of the Semantic Web community for many years.

3. **Web media needs to be annotated using a media ontology which supports the above two issues.** There is no agreed annotation schema for media on the Web, with the best known example MPEG-7 proving both to be too complex and not formally structured enough to be usable in this context. The W3C has proposed a Media Ontology, which seeks to capture common properties of different media annotation schemas and provide as a result a means to map between them [5]. However, we add an additional requirement for that ontology, which is that of capturing the type of the link between the media fragment and the represented thing. Current media annotations barely consider this aspect.

4. **The expressed representation of different concepts by different media fragments in different ways shall be the basis to interlink media across the Web.** By annotating media with concepts connected to larger,

shared domain models, we allow machines to choose and rank media resources on the Web by conceptual relevance and organize and present sets of media resources in a meaningful way based not just on the concepts but how those concepts relate to one another.

Linked Media is a reaction to the current heterogeneity of media descriptions on the Web [9]. Just as Web browsers could never have been successful without a core consensus on the Web page mark up (HTML), Web media can not fulfil its potential without a core consensus on the available descriptive metadata. While a W3C effort to define a common vocabulary for media description is laudable [5] the result is too generic for meaningfully annotating media to concepts, being restricted to a single, underspecified 'keyword' property. Both subject and object of the annotation also need clearer guidelines to ensure interoperability. The desire to open up multimedia annotations to the Web of Data, where they are understandable and sharable across systems, is shared by the creators of the Open Annotation Model (OAM) [13]. They also find that the subject of annotation is often a segment of a media item, and promote the re-use of fragment identification mechanisms already defined in the Web architecture [15].

Regarding the object of annotations, the Web of documents has long had the same problem in annotating its documents for better search and retrieval, since a shared understanding of the concepts the documents are annotated with was needed - neither free text annotation nor use of keywords or tags proved to be a solution immediately unless the differing usage of terms across systems and users could be aligned. Linked Data [4] is an emergent answer to this issue of Web-wide annotation by using the same principles as the Web itself with the concepts used in annotation: identify each concept by an URI, resolve these URIs to descriptions about those concepts, and create links between those URIs so that machines can browse concepts like humans browse Web pages. As noted in a paper outlining the design rationale for the OAM [2] it is the first Web media annotation model to embrace both Media Fragments as subject and Linked Data as object of Web media annotations, thus serving as a suitable starting point for Linked Media annotation.

In the next section, we will look at the currently available tools for semantic annotation of multimedia on the Web and compare their functionality to these Linked Media principles which find their realisation in the specifications mentioned above. We will show how work being done in two recent projects - ConnectME[4] and LinkedTV[5] - most closely follows these goals and outlines how this contributes to the wider vision of semantically annotated Web media.

3   Survey of online semantic media tools

Past and present research activities have also contributed to the implementation of tools for enabling the semantic annotation of media on the Web. An earlier

---

[4] http://www.connectme.at/

[5] http://www.linkedtv.eu/

survey made by Dasiopoulou *et al.* [1] is remarkable in that it is only a few years earlier and yet exhibits a lack of tools which work online (in the browser), not all can annotate media which is available at an URL rather than on the local machine, nor do any show any usage of Linked Data URIs or the set of W3C specifications referred in 2. The goal of semantic multimedia research on the Web must be to enable the publication online of interoperable and re-usable semantic descriptions which can support new online tools and services for media retrieval, re-use and re-mixing. Initial approaches to allow for images and video to be part of Linked Data has focused on natural language processing and clustering techniques across the free text tags attached to media on Web 2.0 sites like Flickr [6] and YouTube [7], but without re-publication of the extracted descriptions for subsequent re-use. We look at recent annotation tools which have emerged since the previous survey, are available online and create semantic descriptions of Web based media which could be published and re-used. We examine if they now more adequately address the below requirements, in line with our Linked Media principles, of:

- tool or service is online and allows to annotate media of different types via URL
- media can be annotated in terms of its parts and the media fragmentation follows a non-proprietary approach supported by other tools
- media can be annotated by concepts where a shared understanding is possible across systems
- media ontology used for the generated descriptions is interoperable across systems

3.1 Review of existing semantic multimedia annotation tools

We consider the following web-based[7] semantic multimedia annotation tools:

- Annomation[8]: this is a browser based tool for video annotation. It is currently restricted to educational material available within the Open University. Tags can be added at any point in the video timeline and given a duration. A number of vocabularies are supported for the tags, including DBPedia and GeoNames. The resulting video annotations re-use several ontologies, but seem to be saved back into the tool's own repository, i.e. they are only available again to the same tool.
- Annotorius[9]: this is an image annotation tool which is browser-based, implemented in JavaScript. It allows the attachment of free text descriptions to a spatial region. A Semantic Tagging plugin suggests named entities for the inserted text, which map to DBPedia resources. Annotations use their own JavaScript data objects for persistence and sharing.

---

[6] http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/
[7] We omit, on purpose, standalone softwares and plugins outside the context of a browser for performing the annotations
[8] http://annomation.open.ac.uk/annomation
[9] http://annotorious.github.io/

- YUMA[10]: developed in the EuropeanaConnect project, it supports image, audio and video. Both DBPedia and Geonames resources can be annotation targets, and are suggested from free text or location references respectively. Annotations can be exported as RDF using a tool-specific vocabulary.
- SMAT[11] - Semantic Multimedia Annotation Tool - promises to allow the annotation using domain ontologies of fragments of content items within a rich internet application. Video can be accessed from any streaming server and annotated with spatial or temporal fragments connected to a term from a preloaded domain ontology. It is targeted to pedagogical usage and seems to be focused on demonstrating the act of media annotation rather than any wider re-use.
- SemTube[12] is a prototype for semantically annotating YouTube videos developed within the SemLib EU project. It allows attaching annotations to both spatial and temporal fragments, with annotations being free text, Freebase terms or full RDF triples. A faceted browser then allows users to explore their annotated videos. It appears functional but seems to be only enabled to save and retrieve annotations within a host server.
- Pundit[13] is an open source Web document annotation tool that has developed out of the SemLib EU project. It incorporates however only image annotation at the moment, allowing regions of the image to be annotated with LOD terms or freely chosen ontology URIs. A client can be downloaded and installed for local annotation of online Web pages which are saved to and retrieved from an instance of a Pundit server.
- IMAS[14] is a Web-based annotation tool developed within the SALERO EU project. Structured descriptions can be produced for media assets retrieved from a repository [16]. SALERO developed its own ontologies for annotating media and describing relationships between media according to the needs of the media production domain. The tool only allows global annotation of media resources and not annotating parts of them, and the output is specifically intended for the needs of producers (e.g. subsequent rediscovery of media) rather than for publication to the Web.
- ImageSnippets[15] enables to tag images using Linked Data resources. Interestingly, tagged images can then be published to the Web, both with descriptions embedded in the image data and included in the HTML file as RDFa metadata, on the fly. However, the tool does not yet support fragment-based annotation and it is restricted to the image medium. It is currently in beta but looks promising, except that its current open annotation approach could suffer from shared public image annotations not being interoperable due to a lack of a common annotation vocabulary among authors.

---

[10] http://dme.ait.ac.at/annotation/
[11] http://www.kp-lab.org/tools/semantic-multimedia-annotation-tool-smat
[12] http://www.semedia.dibet.univpm.it/semtube/
[13] http://thepund.it
[14] http://salero.joanneum.at/imas/
[15] http://www.imagesnippets.com/

– OpenVideoAnnotation[16] plans to offer a web-based tool to collaboratively annotate video on the web, at the fragment level and using the Open Annotation ontology. Annotations are free text comment and tag based, but it is not yet clear if Linked Data will be supported nor if spatial fragments will be included. This tool is clearly promising but this is still a work in progress with a soon to be launched beta program.

Having reviewed the most recent tools known to the authors, we highlight work of two projects the authors have been involved in which are continuing the task of supporting online semantic media annotation.

3.2 The ConnectME toolset

In the ConnectME project, media fragment descriptions were generated out of industry partners existing media systems: in one case, a proprietary CMS and in the other case a Drupal installation. The Figure 1a shows the extended proprietary CMS, where the legacy metadata fields filled in by the media channel owner (title, description, keywords for example) are complemented by a "Start Video Annotation Tool" button. When the button is pressed, the metadata in the CMS for this media asset is published to an API on the metadata repository using the mediaRSS format. An internal script maps the mediaRSS information into a new media asset description using a lightweight metadata model [17], reusing a subset of the W3C Media Ontology and the Open Annotation Model. The fragments are associated to concepts with specific properties (which can be extended) such as `explicitlyMentions` or `implicitlySeen`, allowing media systems processing these descriptions to make distinctions based on the concept being "represented" by the media fragment.

In the case of Drupal, a dedicated RDF Module is used to write Drupal node data into a RDF model and another dedicated module[18] is able to publish this RDF to the Linked Media Framework[19] (LMF) used as media metadata repository. Initial entity extraction for the media metadata is performed in the LMF using a trained Apache Stanbol instance over the media title, description and keywords, generating DBPedia resources. A Web based annotation tool has been developed [20] in which a media asset may be selected and its annotation inspected, using a timeline view (Figure 1b) under the video frame to clearly show descriptions along the media's temporal fragments and allow editors intuitive editing of temporal fragment start/end times by drag & drop. Spatial fragments are also displayed for a selected annotation, if present, and can be changed by drag & drop of a spatial overlay over the video frame. The annotations are shown with their concept labels, with the addition/editing of annotations taking place in an easy-to-use wizard which allows plain text entry and suggests concepts to

---

[16] `http://www.openvideoannotation.org/`

[17] `https://connect.me/ontology`

[18] `https://code.google.com/p/lmf/wiki/DrupalModule`

[19] `https://code.google.com/p/lmf/`

[20] `https://annotator.connectme.at`

the annotator, providing a preview text to allow checking the correct concept is selected. Finished annotations are saved back to the repository, where they are used to enable automatic enrichment of video in a HTML5 based player [21].
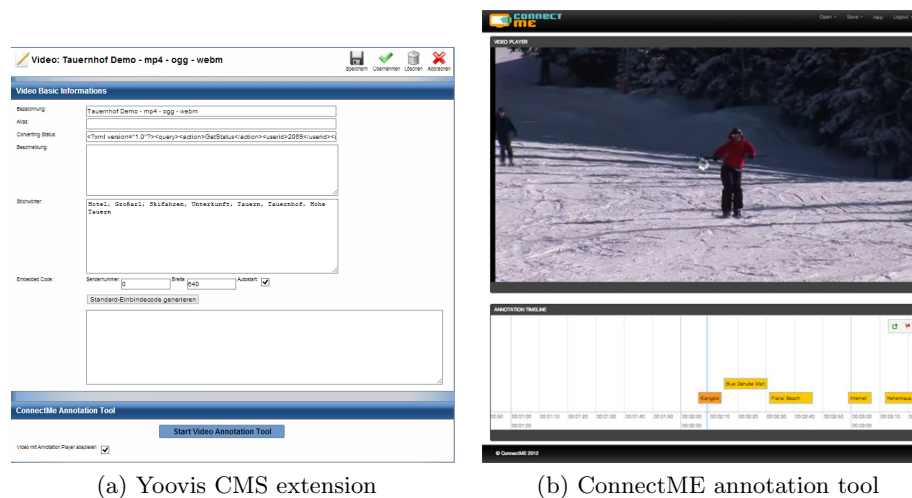

(a) Yoovis CMS extension

(b) ConnectME annotation tool

Fig. 1: Export to semantic annotation integrated into CMS (courtesy Yoovis GmbH, http://www.yoovis.tv) (left) and online video annotation tool (right)

3.3 The LinkedTV toolset

In the LinkedTV project, a Web service is provided for ingesting various types of related information for a media asset and generating RDF descriptions according to the LinkedTV ontology [22] as a result. This combines several ontologies with a specific extension for its use case of modelling the (automated) annotation of media fragments and their association to related content (hyperlinking). It uses the idea of Media Fragments annotated (via the Open Annotation Model) with semantic concepts but can also model a wider range of information, including initial outputs from media analysis processes (which can subsequently be used for the semantic annotation), provenance information and different levels of granularity (fragments can correspond to `Shots`, `Scenes` or `Chapters`). The effect is that of enabling a fuller media fragment description which fits the full media lifecycle within a media management process, e.g. able to preserve information about where an annotation came from, which analysis results (e.g. entity extraction from subtitles) were used to create an annotation, or a series of edits

---

[21] https://player.connectme.at
[22] http://linkedtv.eu/ontology

to a fragment description via an annotation tool. The service aggregates the results of all the different input processing steps and a Web interface is available at `http://linkedtv.eurecom.fr/tv2rdf/`:

1. EXMARaLDA[23], a format for aggregating media analysis results obtained after the execution of different low level feature analysis processes over media content. They include shot segmentation, scene segmentation, concept detection, object detection, automatic speech recognition, face detection, keyword extraction, etc.
2. TV Anytime[24], a metadata format for legacy information from broadcasters such as title, description and keywords.
3. SRT subtitles file, using entity extraction via NERD[25], a REST service which aggregates results from many different online entity extraction services [12, 6], and associating the entities to a temporal fragment of the media.

A Web based editor tool has also been developed [26]. It loads the RDF descriptions from the LinkedTV Platform [27], making a distinction between the media "entities" (the concepts each fragment is annotated with) and "enrichments" (in LinkedTV, on the basis of the media annotations, hyperlinks to related content are also suggested). The LinkedTV metadata generator is already integrated within the LinkedTV platform so that generated descriptions of media assets are available to the Editor Tool. The tool allows an editor to select a specific chapter of the video, browse the existing annotations on the platform, and select them for the "media fragment description" or add new annotations. The manually selected annotations are saved back to the platform while attaching relevant provenance information (using the PROV ontology) which has the effect that - while the existing annotations are preserved and can be returned to - the manually selected annotations can also be selected out from the repository easily. The idea of the Editor Tool is to support content editors at TV broadcasters who will want to proof all automatically generated annotations and use, in the subsequent media workflow, only annotations they manually selected. Figure 2 shows the main interface of the Editor Tool:

Summarizing, two online media annotation UIs are available which can take results from prior analysis steps and store RDF data in dedicated platforms for which some front-end applications have already been developed for media enrichment. Furthermore, a REST service is available for producing semantic media descriptions for any online media making use of acquired analysis results, (TV Anytime) descriptive metadata and/or subtitling. If we consider the entire list of surveyed tools and services with respect to the requirements outlined at the beginning of the section, (Table 1) we observe that these ConnectME and LinkedTV implementations are closest to the vision and goal of publishing Linked Media for new media applications.

---

[23] `http://www.exmaralda.org/`

[24] `http://www.tv-anytime.org/`

[25] `http://nerd.eurecom.fr/`

[26] `http://editortool.linkedtv.eu/`

[27] `http://api.linkedtv.eu/`

Tool	Input Format	Fragment Support	Conceptual Annotation	Ontology
Annomation	Video only. URL loaded from local repository	Temporal fragments, proprietary syntax	DBPedia and GeoNames resources	Reuse of several vocabularies (FOAF, Dublin Core, Timeline, SKOS, Geo)
Annotorius	Image only. URL via JavaScript code	Spatial fragments, proprietary syntax	DBPedia resources	No ontology, proprietary JavaScript data model
YUMA	Audio, image or video. Reference by URL	Spatial and temporal fragments, using SVG	DBPedia and GeoNames resources	Own RDF-based vocabulary
SMAT	Flash video (FLV). Refer to a streaming server by URL	Spatial and temporal fragments, proprietary syntax	Terms selected from preloaded domain ontologies	Unclear, seems to not have any export option
SemTube	YouTube video by URL	Spatial and temporal fragments, presumed proprietary syntax	Freebase resources. Terms can be imported from other ontologies	Own RDF-based vocabulary
Pundit	Image only as part of a Web document	Spatial regions, presumed proprietary syntax	Freebase resources. Terms can be imported from other ontologies	Own RDF-based vocabulary
IMAS	Image and video from a media repository	No fragmentation of media	Assume ability to re-use any ontology term	SALERO ontologies, extensible
ImageSnippets	Image only, can be opened from URL	No fragmentation of media	Users are free to import terms from LOD or create their own	LIO ontology
Open Video Annotation tool	Video, unclear if access will be possible via URL	Only temporal fragments in the demo. Format unknown	Free text or tags, unclear if LOD will be supported	Open Annotation Model
ConnectME Annotation tool	Video, access via URL or from repository	Spatial and temporal fragments according to Media Fragment URI syntax	DBPedia and Geoname resource suggestion, extensible	ConnectME ontology, W3C Media Ontology and Open Annotation Model
TV2RDF service	Video, access via URL	Spatial and temporal fragments according to Media Fragment URI syntax	Depending on entity extraction service. Focus on using DBPedia	LinkedTV Ontology, W3C Media Ontology and Open Annotation Model
LinkedTV Editor Tool	Video, access via a media resource registered on LinkedTV Platform	Spatial and temporal fragments according to Media Fragment URI syntax	Any URI can be provided. Autocomplete available for suggesting DBPedia resources	LinkedTV Ontology, W3C Media Ontology and Open Annotation Model

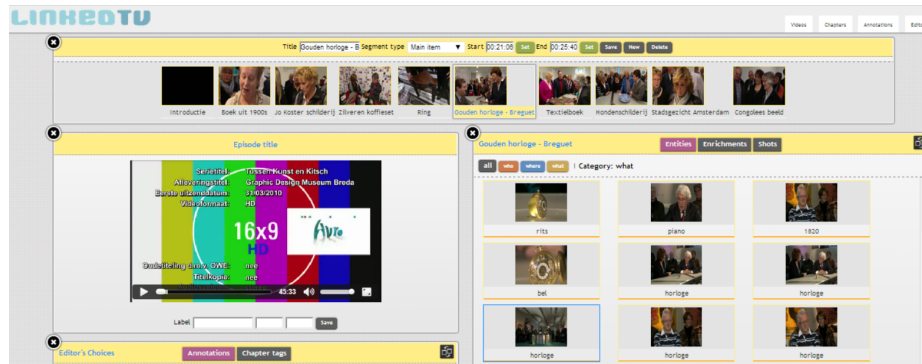Table 1: Survey of online semantic media annotation tools

Fig. 2: LinkedTV Editor Tool

4 New Media applications with Linked Media systems

Future semantic media systems will incorporate analysis technology (for media fragment creation and classification) prior to (semi-automatic) semantic annotation of those fragments and subsequently copyright tools can support the attachment of licenses to the media fragments before the metadata is stored for use in new media applications. A number of such applications are being prototyped in the MediaMixer project [28] to highlight the value of semantic multimedia technology in the context of specific industry use cases, including dynamic video clip selection for newsrooms, license modeling for user generated content, and creating topical channels of online learning video content.

We briefly describe in this section the demonstrator prepared together with the VideoLectures.NET platform. VideoLectures.NET hosts more than 16 000 video lectures and it is currently challenging to find specific topics within those lectures as the current implementation relies on the indexing of basic top level metadata about each lecture (title and description). Media fragment and semantic technology has been introduced in a proof of concept with a selected subset of site materials. We focus here on the annotation, a fuller description of the implementation can be found at [10]. VideoLectures manages natively XML descriptions of lectures (title, description, category, keywords). We map this metadata into a TV Anytime like description using a XML template and simple script to perform the conversion. Analysis results are acquired from both shot segmentation algorithms run on the video itself and the slide transition XML documents, providing two different granularities of video temporal fragmentation. Using the TV2RDF service described in the section 3.3, we can process these results as well as video transcripts created automatically by an ASR software[29] and acquire an aggregated semantic description of the lecture in terms of its salient fragments and Linked Data concepts attached to each fragment. In

---

[28] http://community.mediamixer.eu/demonstrators
[29] http://www.translectures.eu/tlk/

the resulting demonstrator[30], a search on specific terms can now return video fragments from different lectures where the term has been extracted. For example DNA returns no results in the current Videolectures.NET site, while the MediaMixer demonstrator provides many matches across different video lectures (Figure 3).

Fig. 3: VideoLecturesMashup: finding lecture snippets for a topic

As a next step, since the fragment annotations use Linked Data resources (DBPedia), semantic search will be implemented so that search can also take into account synonyms, multilingualism, and topical relevance. We have taken technologies originally developed in the LinkedTV project developed to enrich TV programming with links to related content, and applied the same tools to enable fragment level search over lecture videos. It is an important contribution to have separate "building blocks" of what we call overall a Linked Media system (as it respects the principles of Linked Media outlined in Section 2) that can support different domains and end-user applications with the same data formats and approaches to semantic media annotation.

The experiences gained with demonstrators in different domains help validate the Linked Media vision and the technology, with annotation results being usable in a number of front end applications such as LinkedTV's interactive TV second screen application (Figure 4, cf. http://linkedtv.eu/demos/linkednews). A full end-to-end workflow is being implemented within the server-side LinkedTV Platform with the goal to provide users with a simple means to ingest their video content and get access to the semantic annotation results via a SaaS model. As reflected in the survey of the previous section, it is these sort of tools which are preventing a wider uptake by media owners and enterprises of semantic annotating their media and re-use in new contexts.
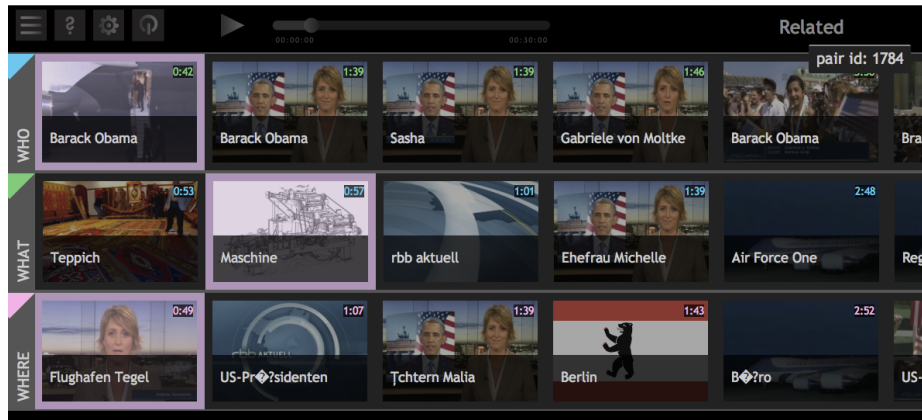
---

[30] http://mediamixer.videolectures.net

Fig. 4: LinkedTV second screen app displaying concepts annotated in a TV program

5 Conclusion

This paper began by acknowledging that, despite the benefits that semantic multimedia technology is perceived to offer - improved media management, retrieval and re-use - the amount of online media today having semantic annotations is desperately low. We have identified barriers to greater uptake of this technology in the unfortunate heterogeneity of approaches coming out of research and lack of integration with the Web itself as technology platform. We refer to principles for Linked Media which can connect online media descriptions more usefully into the Web of Data. A survey of current annotation tools reflected on a lack of consideration of these requirements in their implementation and has highlighted online tools and services from the ConnectME and LinkedTV projects which are closer to the vision and goal of eased, online publication of re-usable semantic descriptions of media. Demonstrators such as the VideoLecturesMashup point to the feasibility of following this Linked Media approach to ensure semantic media annotation can be integrated into media workflows and the resulting annotations re-used across different media applications.

Just as the Linked Data movement has worked hard to encourage data owners to embrace the value of online publication of their data in a structured, interlinkable manner, we must also address the challenge of encouraging media owners to publish semantic descriptions of their media online in a similarly structured and interlinkable manner. A growing body of semantic media annotations which follow interoperable data models, structures and concept vocabularies so that computer systems can search and link across them could support the creation of new media applications for search, re-use or re-mixing of online media. It should be noted that this is not insisting on the open publication of the media assets themselves: they may still be found behind paywalls or restricted by licenses,

but through the publication of semantic descriptions they become more findable to applications which may need them for a re-use or re-mixing task.

If this is to become a true possibility in the soon to emerge Audiovisual Web, it is critical that current semantic multimedia research agrees to a consensus on how annotation is to be done: how it fits with outputs of prior media analysis, how it handles the different granularities of annotation on a single media asset, what metadata models, data formats and conceptual vocabularies it uses, how annotation output can be easily re-used by media management systems, in media retrieval, or for (multi)media presentation or re-purposing. This paper proposes to the semantic multimedia research community to embrace a consensus around the so-called Linked Media principles and has highlighted specifications which can provide for the necessary cross-system interoperability as well as validated this approach with first demonstrators building applications on top of Linked Media-conforming semantic media descriptions.

Acknowledgments

References

1. Stamatia Dasiopoulou, Eirini Giannakidou, Georgios Litos, Polyxeni Malasioti, and Yiannis Kompatsiaris. *A Survey of Semantic Image and Video Annotation Tools*, volume 6050 of *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution: Bridging the Semantic Gap*, chapter 8, pages 196–239. Springer, 2011.
2. Bernhard Haslhofer, Robert Sanderson, Rainer Simon, and Herbert Sompel. Open annotations on multimedia Web resources. *Multimedia Tools and Applications*, pages 1–21, 2012.
3. Jane Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In 1st *International Semantic Web Working Symposium (ISWC)*, pages 261–281, 2001.
4. Tim Berners Lee. Linked Data Principles. Design Issues, 2006.
5. WonSuk Lee, Werner Bailer, Tobias Bürger, Pierre-Antoine Champin, Jean-Pierre Evain, Véronique Malaisé, Thierry Michel, Felix Sasaki, Joakim Söderberg, Florian Stegmaier, and John Strassner. Ontology for Media Resources 1.0. W3C Recommendation, 2012.
6. Yunjia Li, Giuseppe Rizzo, José Luis Redondo García, Raphaël Troncy, Mike Wald, and Gary Wills. Enriching media fragments with named entities for video classification. In 1st *Worldwide Web Workshop on Linked Media (LiME'13)*, pages 469–476, Rio de Janeiro, Brazil, 2013.
7. Yunjia Li, Giuseppe Rizzo, Raphaël Troncy, Mike Wald, and Gary Wills. Creating enriched YouTube media fragments with NERD using timed-text. In 11th *International Semantic Web Conference (ISWC), Demo Session*, Boston, USA, 2012.

8. Frank Nack, Jacco van Ossenbruggen, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). *IEEE Multimedia*, 12(1), 2005.

9. Lyndon Nixon. The importance of Linked Media to the Future Web. In *1st International Workshop on Linked Media (LiME)*, Rio de Janeiro, Brazil, 2013.

10. Lyndon Nixon, Tanja Zdolsek, Ana Fabjan, and Peter Kese. Video Lectures Mashup: remixing learning materials for topic-centred learning across collections. In *OCW Global Conference*, Ljubljana, Slovenia, 2014.

11. Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). *IEEE Multimedia*, 11(4), 2004.

12. Giuseppe Rizzo and Raphaël Troncy. NERD: A Framework for Unifying Named Entity Recognition and Disambiguation Extraction Tools. In *13th Conference of the European Chapter of the Association for Computational Linguistics (EACL'12)*, Avignon, France, 2012.

13. Robert Sanderson, Paolo Ciccarese, and Herbert Van de Sompel. Open Annotation Data Model. W3C Community Draft, 2013.

14. Guus Schreiber, Alia Amin, Mark van Assem, Victor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Zhisheng Huang, Janneke van Kersen, Marco de Niet, Borys Omelayenko, Jacco van Ossenbruggen, Ronny Siebes, Jos Taekema, Jan Wielemaker, and Bob Wielinga. MultimediaN E-Culture Demonstrator. In *5th International Semantic Web Conference (ISWC)*, 2006.

15. Raphaël Troncy, Erik Mannens, Silvia Pfeiffer, and Davy Van Deursen. Media Fragments URI 1.0 (basic). W3C Recommendation, 2012.

16. Wolfgang Weiss, Tobias Bürger, Robert Villa, P. Punitha, and Wolfgang Halb. Statement-Based Semantic Annotation of Media Resources. In *4th International Conference on Semantic and Digital Media Technologies (SAMT)*, 2009.