

Seamlessly interlinking TV and Web content to enable Linked Television

Lyndon Nixon
MODUL University
Am Kahlenburg 1
1190 Vienna, Austria
lyndon.nixon@modul.ac.at

Vasileios Mezaris
CERTH-ITI
PO Box 60361
57001 Thermi-Thessaloniki,
Greece
bmezaris@iti.gr

Jan Thomsen
CONDAT AG
Alt Moabit 91d
10559 Berlin, Germany
jan.thomsen@condat.de

ABSTRACT

This paper explains Linked Television, a concept for interlinked TV and Web content within a shared digital ecosystem under the control of the content owner. An end to end platform is being developed to ingest, analyse and annotate TV programmes, and generate links to related Web content from the TV programme itself. We focus on a highly automated system available as a Web service to help lower the cost of enrichment of TV programming, and report on current demonstrators of the system's capabilities. While still the subject of research, we hope to provide wider access to this platform as a step towards making Linked Television a future reality.

Author Keywords

Linked Television; hypervideo; interactive TV; smart TV; second screen; annotation; enrichment

ACM Classification Keywords

H.3.1. Information Storage and Retrieval : Content Analysis and Indexing

1. INTRODUCTION

In the context of fundamental changes in television, viewers are increasingly engaging in parallel Internet usage to their TV programme consumption. There is also a trend towards looking up additional information related to the program being watched. Surveys in 2012 (summarized at <http://www.theguardian.com/technology/appsblog/2012/oct/29/social-tv-second-screen-research>) found that 75 to 85% of TV viewers are also using a second screen while viewing TV, while 37 to 52% are using that screen to follow what was going on in the TV program (A recent survey found 40% [1] with Google the first place to search for program-related information). Because viewers have limited means to conduct many of the information searches they desire, e.g. finding out which painting is hanging on the wall

behind a character is nearly impossible to formulate as a textual query to Google, such searches do not occur. On the other hand, providing such links as an additional service via connected devices can be a means to create new value for original television content.

Current approaches (Shazam, IntoNow) focus on costly, fully manual curation of related information for selected episodes of TV programming and that information is hardcoded and related mostly to the whole episode rather than specific topics or objects in it. Media owners perceive excessive costs in providing eased access to more granular related information for larger amounts of their content. Linked Television seeks to address this by providing integrated technology for ingesting, analysing and annotating TV programmes, and generate links to related Web content from the TV programme itself. We test this in two scenarios in association with European broadcasters in the domains of news and cultural heritage respectively. Both scenarios are the subject of operational demonstrators, running on top of LinkedTV technology provisioned as an online Platform with a Web service interface.

This paper introduces current results of Linked Television: the scenarios (Section 2), technologies for video analysis, annotation and enrichment (Section 3) and how the workflow between these technologies is realised in an online Platform (Section 4). To conclude, we consider how this solution can pave the way to a future integration of Web and TV to the benefit of content owners (Section 5).

2. LINKEDTV SCENARIOS

Linked Television currently pursues scenarios in two areas: *Interactive News*, together with the German broadcaster RBB and *Hyperlinked Documentary*, together with the Dutch broadcaster AVRO. Initial trials conducted with small, representative viewer groups validated the hypothesis that viewers would be interested in accessing additional information related to the TV program they were viewing [6]. In the Interactive News scenario, this is the RBB Aktuell evening news program, whereas in Hyperlinked Documentary we use episodes of the program Tussen Kunst & Kitsch (similar to the BBC's Antiques Roadshow). This content proves very relevant to Linked

Television as experts tend to mention various concepts while explaining antique objects without going into further depth, whereas viewers watch episodes specifically out of interest in those antique objects. Such "passing references" are typically the subject of further information wishes.

Videos have been made for both scenario demonstrators:

- The Linked News demonstrator
<http://www.linkedtv.eu/demos/linkednews/>
- The Hyperlinked Documentary demonstrator
<http://www.linkedtv.eu/demos/hyperlinkeddocu/>

3. CONTENT WORKFLOW

The first step towards establishing meaningful links between pieces of TV and Web content is to understand what each piece of content is about: this is a prerequisite for making informed linking decisions. Content understanding in the LinkedTV workflow is realized by automatically fragmenting and annotating in various ways the different forms of information (visual, audio, textual) that comprise a piece of TV content or other related content (e.g. a video clip with or without subtitles, or an audio recording).

The analysis of the visual modality starts by fragmenting the video into elementary temporal fragments, i.e. visual shots, using an extension of [12]. Shots are the basis of most of the subsequent visual analysis processes. Despite their significance for analysis, though, a shot does not necessarily represent the most suitable decomposition granularity for a human to consume the TV content, being typically just a few seconds long and not able to convey a meaningful event by itself. For this, fragmentation of the video into shots is followed by grouping the latter into larger, more self-contained temporal fragments, i.e. scenes, using the method described in [10] and extensions of it that also take into account the results of object re-detection. The latter results are produced by a technique for effecting spatio-temporal fragment creation, by finding all re-appearances of a known object throughout a video file [2].

For annotation of each content fragment, several different routes are followed: Optical character recognition, Visual concept detection [9], or Automatic speech recognition [8] as well as the analysis of textual transcripts (e.g. programme subtitles) are all possible sources of annotations (association of a concept to a content fragment). In the latter case, Named Entity Recognition [7] is performed across timed text material so that entities (concepts) can be associated to specific temporal segments of the TV programme (subtitle blocks), cf. [5], [4] for explanations of the approach. In order to combine all of these analysis and annotation results into a single metadata file for each TV programme, an unified metadata model has been created in LinkedTV (the LinkedTV Ontology, <http://linkedtv.eu/ontology>, extending [3] as well as a Web service which combines outputs of the different approaches

into a combined metadata record (Tv2RDF, <http://linkedtv.eurecom.fr/tv2rdf/>).

The aggregated annotation is used for the subsequent enrichment. Related content items on the Web, crawled according to a whitelist (so only content from trusted sites is considered), are also annotated with entities and entity matching techniques are used to determine, for an annotated fragment of the TV program, which set of Web links match best that annotation [11]. As a result, LinkedTV can automatically provide link sets to related Web content for the annotated TV program at no extra cost to the broadcaster - although, sensitive to the potential issues around offering automated links to the viewer, LinkedTV has also developed an Editor Tool where human editors can check and correct any annotations or suggested links easily¹.

4. LINKEDTV PLATFORM

The overall LinkedTV Architecture can be roughly divided into three conceptual layers (Figure 1):

- a) the LinkedTV Platform,
- b) the Linked Media Services (such as analysis, annotation and enrichment, bundling various individual sub-services), and
- c) a Service Layer which uses these services and supports client interaction, in particular for enriched TV playback.

This section gives an overview over the LinkedTV Platform and its integration approach with the other layers. The main task of the LinkedTV Platform is to support the Linked Media Workflow as defined and realized in LinkedTV. In itself it does not contain any components which generate data, but it rather provides the means to store, access, search, link, update or manage that data. Figure 2 shows the basic Linked Media Workflow from ingestion, analysis and annotation, to serialization and enrichment, generation and aggregation of the system metadata, and finally personalization and payout. The Service Layer provides API services for retrieval, accessing and manipulating data stored in the Platform. This includes the general LinkedTV services which are provided for the different components of the LinkedTV Platform for analysis, annotation, enrichment, or personalization, as well as external clients such as the Editor Tool or the LinkedTV Player. The Service Layer is completely realized through a set of REST services. The two main repositories, the Administration Database and the RDF Respository, are exposed via two different REST Base URLs <http://api.linkedtv.eu> and <http://data.linkedtv.eu>. The first one gives access only to complete media resources, whereas the second one gives

¹ <http://editortool.linkedtv.eu>

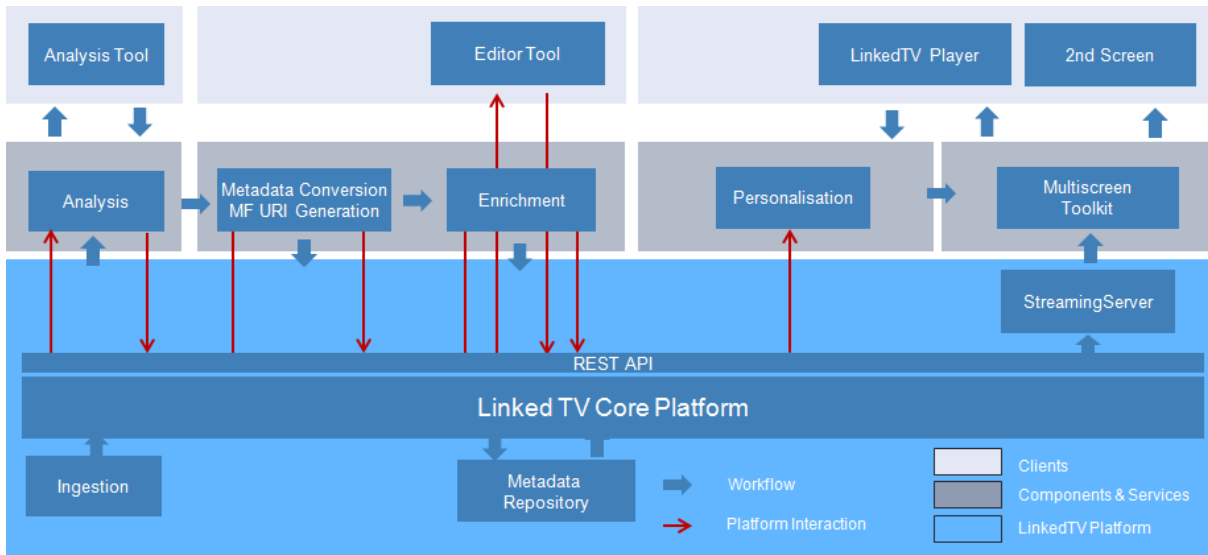


Figure 1. LinkedTV Architecture

access to the RDF data of complete media resources as well as media fragments, annotations, and more. Both REST APIs ensure that the LinkedTV Platform is an open platform which can be used by other clients as well as integrated with other IPTV Environments.

This data is requested by the LinkedTV Player application (which may run on a companion device, e.g. in a tablet) which visualises the concepts in the video annotation to the viewer. The current Player UI has taken the approach to organize concepts along the axes of "who" (persons), "where" (locations) and "what" (other) (see Figure 3). Concepts are shown as thumbnails which are highlighted when the concept is currently 'active' in the TV programme. When a concept is selected (touch event) available enrichments for that concept are shown, e.g. a description (Wikipedia) or related media.

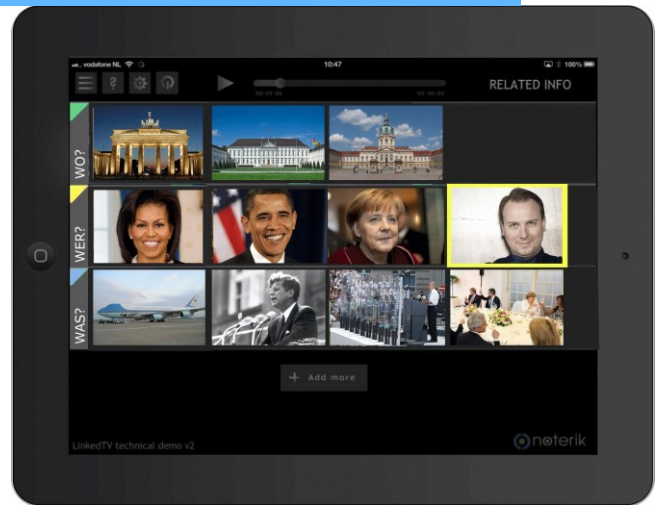


Figure 3. LinkedTV Player User Interface

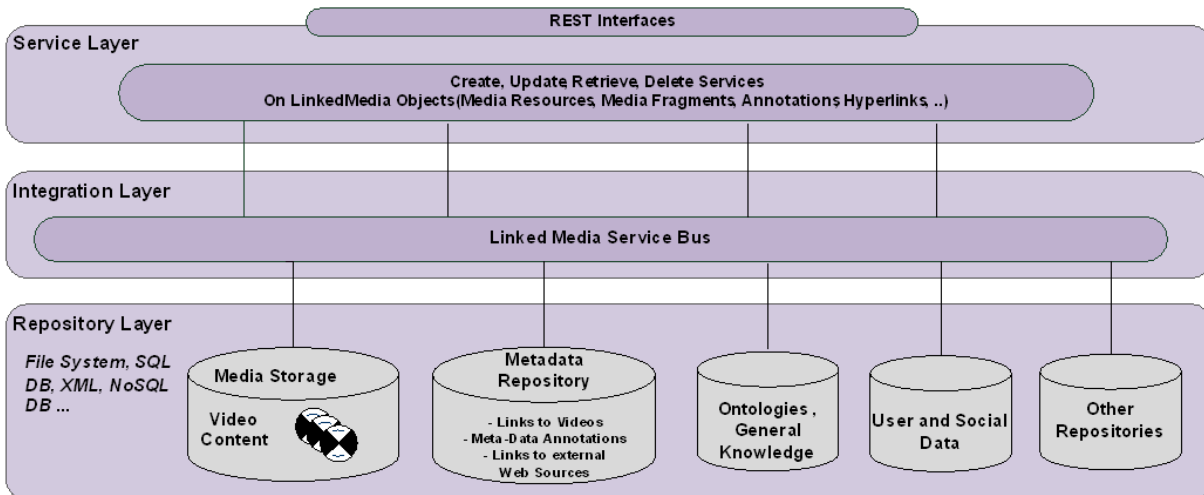


Figure 2. Linked Media Workflow

CONCLUSION

LinkedTV believes that facilitating the richer annotation of TV content and the re-use of annotation to dynamically enrich TV content with links to related information and Web content is a worthwhile contribution to the future of television. An implementation has been made of a Platform which supports the workflow to ingesting video material, analysing it, annotation and generating enrichments. First scenario demonstrators are available which show the possibilities of using LinkedTV-sourced annotation and enrichment in a second screen application for seamlessly accessing related information while watching a TV program.

As work in progress, LinkedTV is focused in its final year on achieving greater accuracy in its automated analysis, annotation and enrichment workflow. Current outputs still require some effort in curation of the annotations and enrichments by TV editors, which is being facilitated by an online Editor Tool we are developing within the project. A dashboard will be implemented on the Platform to allow media owners to more easily track the status of the processing of their content, including integration of the Editor Tool to allow for visualisation and correction of the extracted annotations and enrichments. Aside from technical challenges to produce more accurate automatic results, LinkedTV faces also a business challenge in its goal to bring Linked Television to the masses. Ideas like opening up more metadata about valuable TV content to other applications, or giving viewers the chance to browse away from that content to other information, are challenging to the TV industry yet become necessities to survive - in our opinion - in a world where TV is no longer the only source of content nor consumed independently from other sources, and added value services and interactivity will be game changers: the questions remains whether the game will change for the traditional broadcasters or the emerging OTT providers first.

ACKNOWLEDGMENTS

The research leading to this paper was supported by the European Union's 7th Framework Programme via the project LinkedTV (GA 287911) <http://www.linkedtv.eu>

REFERENCES

1. J. Abreu, P. Almeida, B. Teles, and M. Reis. Viewer behaviors and practices in the (new) television environment. In Proceedings of the 11th European Conference on Interactive TV and Video, EuroITV

'13, pages 5-12, New York, NY, USA, 2013. ACM.

2. E. Apostolidis, V. Mezaris, and I. Kompatsiaris. Fast object re-detection and localization in video for spatio-temporal fragment creation. In MMIX'13 at IEEE ICME, San Jose, CA, USA, 2013.

3. B. Haslhofer, R. Sanderson, R. Simon, and H. Sompel. Open annotations on multimedia Web resources. *Multimedia Tools and Applications*, pages 1-21, 2012.

4. Y. Li, G. Rizzo, J. L. R. Garcia, R. Troncy, M. Wald, and G. 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.

5. Y. Li, G. Rizzo, R. Troncy, M. Wald, and G. Wills. Creating enriched YouTube media fragments with NERD using timed-text. In 11th International Semantic Web Conference (ISWC), Demo Session, Boston, USA, 2012.

6. L. Nixon. Scenario Demonstrators. LinkedTV Deliverable 6.2, 2013.

<http://www.slideshare.net/linkedtv/scenario-demonstrators>

7. G. Rizzo and R. 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.

8. J. Schwenninger, D. Stein, and M. Stadtschnitzer. Automatic Parameter Tuning and Extended Training Material: Recent Advances in the Fraunhofer Speech Recognition System. In Workshop Audiosignal- und Sprachverarbeitung, Koblenz, Germany, 2013.

9. P. Sidiropoulos, V. Mezaris, and I. Kompatsiaris. Enhancing video concept detection with the use of tomographs. In IEEE ICIP, Melbourne, Australia, 2013.

10. P. Sidiropoulos, V. Mezaris, I. Kompatsiaris, H. Meinedo, M. Bugalho, and I. Trancoso. Temporal video segmentation to scenes using high-level audiovisual features. *IEEE Trans. on Circuits and Systems for Video Technology*, pages 1163-1177, 2011.

11. D. Stein, A. Oktem, E. Apostolidis, V. Mezaris, J. R. Garcia, R. Troncy, M. Sahuguet, and B. Huet. From Raw Data to Semantically Enriched Hyperlinking: Recent Advances in the LinkedTV Analysis Workow. In NEM Summit, Nantes, France, 2013.

12. E. Tsamoura, V. Mezaris, and I. Kompatsiaris. Gradual transition detection using color coherence and other criteria in a video shot meta-segmentation framework. In IEEE ICIP-MIR, pages 45-48, San Diego, CA, USA, 2008.