Web and TV Seamlessly Interlinked: LinkedTV*

Lyndon Nixon**

STI International GmbH
Neubaugasse 10/15, 1070 Vienna, Austria
lyndon.nixon@sti2.org

Abstract. This paper reports on the vision of LinkedTV driven by the EU project of the same name[1], and the work done in its first year. LinkedTV is a new type of television (or audio-visual) experience where Web and TV content can be seamlessly interlinked based on the concepts present within that content. The project addresses how the Web and TV is converging in end devices, and particularly this paper focuses on how we intend to answer the research challenges that the LinkedTV vision raises.

Keywords: Smart TV, Networked Media, semantic multimedia, media annotation, Connected TV, future TV.

1 Introduction

Networked Media will be a central element of the Next Generation Internet. Online multimedia content is rapidly increasing in scale and ubiquity, yet today it remains largely still unstructured and unconnected from related media of other forms or from other sources. This cannot be clearer than in the current state of the Digital "Smart" TV market. The full promise and potential of Web and TV convergence is not reflected in offerings which place the viewer into a closed garden, or expect PC-like browsing of the Web on a distant TV screen, or extend the television with new functionalities which however lack any relation to the currently viewed TV programming.

Our vision of future **Television Linked To The Web (LinkedTV)** is of a ubiquitously online cloud of Networked Audio-Visual Content decoupled from place, device or source. Accessing audio-visual programming will be "TV" regardless whether it is seen on a TV set, smartphone, tablet or personal computing device, regardless of whether it is coming from a traditional or new media broadcaster, a Web video portal or a user-sourced media platform. Television existing in the same

* The other LinkedTV consortium partners are: RBB Rundfunk Berlin Brandenburg (Germany), Sound and Vision (Netherlands), University of Mons (Belgium), CONDAT (Germany), Noterik (Netherlands), Fraunhofer IAIS (Germany), CERTH-ITI (Greece), EURECOM (France), University of Economics Prague (Czech Rep), CWI (Netherlands), University of St Gallen (Switzerland).

** The LinkedTV Consortium.

[1] www.linkedtv.eu, Twitter@linkedtv

M. Mancas et al. (Eds.): INTETAIN 2013, LNICST 124, pp. 32–42, 2013.

ecosystem as the Web means that television content and Web content should and can be seamlessly connected, and browsing TV and Web content should be so smooth and interrelated that in the end even "surfing the Web" or "watching TV" will become as meaningless a distinction as whether the film is coming live from your local broadcaster, as VOD from another broadcaster, or from an online video streaming service like Netflix. As a result, not only commercial opportunities but also opportunities for education, exploration and strengthening European society and cultural heritage arise. Imagine browsing from your local news to Open Government Data about the referenced location to see voting patterns or crime statistics, or learning more about animals and plants shown in the currently viewed nature documentary without leaving that show, or jumping from the fictional film to the painting the character just mentioned to virtually visiting the museum when it can be seen, or seamlessly accessing additional information that has been automatically aggregated from multiple sources in order to get better informed on an important event that was just mentioned in the news.

Technologically, this vision requires systems to be able to provide networked audio-video information usable in the same way as text based information is used today in the original Web: interlinked with each other at different granularities, with any other kind of information, searchable, and accessible everywhere and at every time. Ultimately, this means creating hypermedia at the level of the Web whose original success was the underlying hypertext paradigm built into HTML. Hypermedia has been pursued for quite a while as an extension of the hypertext approach towards video information. This requires suitable descriptive models of media that allow for its interlinking, as well as client applications able to process and play out hypermedia based on those descriptions, but to avoid a fully manual and hence not scalable approach for the scale of the Web, it needs complex media analysis algorithms and is still an open issue of research. The **Television Linked To The Web (LinkedTV)** project aims at a novel *practical* approach to Future Networked Media based on four phases: annotation, interlinking, search, and usage (including personalization, filtering, etc.).

The rest of the paper is structured as follows: Section 2 introduces the scenarios of LinkedTV, which motivate our vision and will act as the basis for prototypes. Section 3 outlines the LinkedTV architecture and player, while Section 4 references the research challenges within the project and how we aim to answer them. Finally Section 5 concludes with the outlook for the realisation of LinkedTV as part of every citizen's future experience of television.

2 LinkedTV Scenarios

LinkedTV will demonstrate its vision of weaving of television and the Web through three scenarios, each of which representing different aspects of the value and potential of the future Networked Media Web. These are a current affairs scenario, a documentary scenario, and a media artist scenario.

Current Affairs Scenario

In general, the envisaged service targets a broader audience. For the sake of a convincing scenario, however, we have sketched a few fictional users of the LinkedTV news service and their motivations to use it. For example, socially active retiree **Peter** watches the news show "rbb AKTUELL". One of the spots is about a fire at famous Café Keese in Berlin. Peter is shocked. He used to go there every once in a while, but that was years ago. As he hasn't been there for ages, he wonders how the place may have changed over the years. In the news spot, smoke and fire engines was almost all one could see, so he watches some older videos about the story of the famous location where men would call women on their table phones – hard to believe nowadays, now that everyone carries around mobile phones! Memories of these good old days make him happy and sad at the same time. After checking these very nice clips on the LinkedTV service, he returns to the main news show and watches the next spot on a new Internet portal about rehabilitation centres in Berlin and Brandenburg. He knows an increasing number of people who need such facilities. He follows a link to a map of Brandenburg showing the locations of these centres and bookmarks the linked information to check again later.

Documentary Scenario

In the documentary scenario, we have storyboarded with the persona Rita, an administrative assistant at the Art History department of the University of Amsterdam. She didn't study art herself, but spends a lot of her free time on museum visits, creative courses and reading about art. One of her favourite programmes is the Antiques Roadshow (Dutch title: Tussen Kunst & Kitsch), which she likes to watch because, on the one hand, she learns more about art history, and on the other hand because she thinks it's fun to guess how much the objects people bring in are worth. She's also interested in the locations where the programme is recorded, as this usually takes place in a historically interesting location, such as a museum or a cultural institute.

Rita is watching the latest episode of the Roadshow. The show's host, Nelleke van der Krogt, gives an introduction to the programme. Rita sees the show has been recorded in the Hermitage Museum in Amsterdam. She always wanted to visit the museum as well as finding out what the link is between the Amsterdam Hermitage and the Hermitage in St. Petersburg. She sees a shot of the outside of the museum and notices that it was originally a home for old women from the 17th century. Intriguing! Rita wants to know more about the Hermitage location's history and see images of how the building used to look. After expressing her need for more information, a bar appears on her screen with additional background material about the museum and the building in which it is located. While Rita is browsing, the programme continues in a smaller part of her screen. After the show introduced the Hermitage, a bit of its history and current and future exhibitions, the objects brought in by the participants are evaluated by the experts. One person has brought in a golden, filigree box from France in which people stored a sponge with vinegar they could sniff to stay awake

Fig. 1. Chi-Ro symbol in the TV program

during long church sermons. Inside the box, the Chi Ro symbol has been incorporated (Figure 1). Rita has heard of it, but doesn't really know much about its provenance and history.

Again, Rita uses the remote to access information about the Chi Ro symbol on Wikipedia and to explore a similar object, a golden pyx with the same symbol, found on the Europeana portal. Since she doesn't want to miss the expert's opinion, Rita pauses the programme only to resume it after exploring the Europeana content. The final person on the show (a woman in her 70s) has brought in a painting that has the signature 'Jan Sluijters'. This is in fact a famous Dutch painter, so she wants to make sure that it is indeed his. The expert - Willem de Winter - confirms that it is genuine. He states that the painting depicts a street scene in Paris, and that was made in 1906. Rita thinks the painting is beautiful, and wants to learn more about Sluijters and his work. She learns that he experimented with various styles that were typical for the era: including fauvism, cubism and expressionism. She'd like to see a general overview of the differences of these styles and the leaders of the respective movements.

During the show Rita could mark interesting fragments by pressing a button on her remote control. While tagging she continued watching the show but afterwards these marked fragments are used to generate a personalized extended information show based on the topics Rita has marked as interesting. She can watch this related / extended content directly after the show on her television or decide to have this playlist saved so she can view it later. This is not only limited to her television but could also be a desktop, second screen or smartphone, as long as these are linked together. She's able to share this information on social networks, allowing her friends to see highlights related to the episode.

The Media Art Scenario

This scenario has focused on the other hand on "personalized remixing of TV" using several feature dimensions. Features can be extracted from two main sources. The first one is the automatic analysis of the video itself. We use state-of the art frameworks to segment the videos into scenes. For each scene, three kinds of features can be extracted concerning object detection, event detection and emotion detection. On the other side, features can be extracted from behavioural observation, especially

by using RGBD cameras (like the Microsoft Kinect). Those features can bring information about viewer's attention or change in behaviour (was not moving but now he is moving, was talking but now he stopped and looks towards the TV, etc.). During playout of video content with a structure it is not possible to adapt the video itself to the viewer as parts of the video cannot be avoided or moved without breaking the scenario continuity. In this context automatically segmented scenes can be stored during playout to be used after the content visualisation or during a pause (advertisement or user defined program pause). The scenes to be stored can be either automatically stored (they contain objects of interest for the user as described in his profile, the user had a sudden behaviour change, …) or be manually stored (by using a specific gesture recognized by an RGBD camera).

A mock-up of the interface could be the one in Figure 2 with the main content in the middle and the band of scenes with their characteristics (activity detection, objects, viewer reaction). This menu could also be located on a second screen to avoid distracting too much the viewer. During pauses or after the content is finished, the user can access to all the scenes which were saved or the ones he manually saved during playout. The viewer has access to the enrichment of those videos (hyperlinks, links to other videos…) based on the three kinds of features which were extracted (object-based, event-based and emotion-based). The viewer can than keep a subset of scenes which summarizes in a personal way the video he saw with additional content he added from the enrichment links. This can be sent towards social networks. The information coming from these summaries can be used to augment the personal profile for more efficient enrichment personalisation.

The scenes coming from already seen content and additional content from automatic enrichment can finally be used as a scene database for semi-automatic remixes and mash-ups. An initial video scene is selected and placed in the centre of the screen while 3 clusters are formed around using object-based, event-based and emotion-based features to compute the similarity to the seed video scene. The viewer can than mix the current seed with another video from one of the three clusters as in the figure below.

Fig. 2. Mock interface for the media art scenario

3 LinkedTV Architecture

To realize all these scenarios an architecture for a LinkedTV infrastructure, integrating different media technologies into an agreed workflow, has been developed based on analysis of the use case requirements, technical feasibility and identification of the project research goals.

The LinkedTV platform analyses and annotates external videos, generates media fragments and enriches them semantically with external information from the Web and Linked Open Data sources. The annotated tags and links can be adjusted and enhanced by an editor through tools for annotation and hyperlinking. Based on the enriched and interlinked media fragments various user applications with personalized user interfaces for clickable video allow the user to access the provided videos. Within LinkedTV three different user scenarios demonstrate the possibilities enabled by the LinkedTV approach. This includes a Web client variant using a Browser with the full potential of HTML5 for clickable video based on a two-way HTTP communication and a second client variant, using HbbTV, HTTP and a TV-set with reduced functionality.

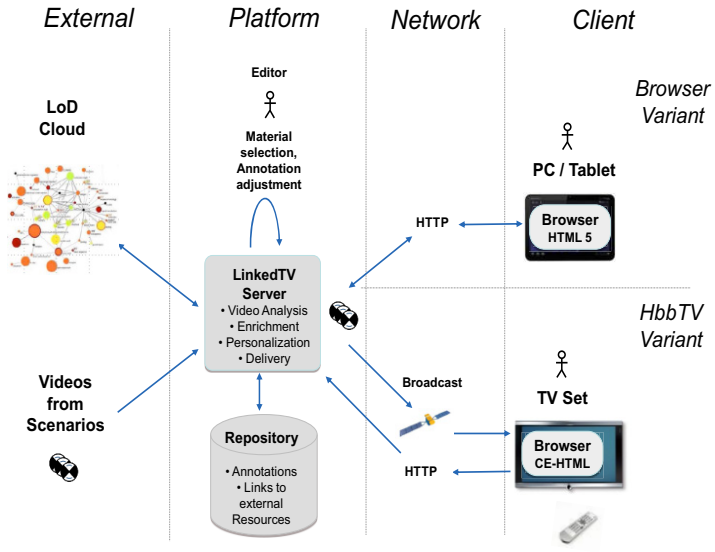

Fig. 3. LinkedTV System Overview

The LinkedTV platform employs a Service Oriented Architecture (SOA) with a division in three layers each consisting of several components. The use of REST Services ensures an efficient, flexible and fault tolerant communication between all components. The SOA architecture allows together with the adoption of standards for formats, interfaces and protocols the exchange of platform components by third party software, distributed development of components, scalability and multi-lingualism.

The LinkedTV Platform is divided in three main layers: 1) the Analysis and Annotation Layer containing the components and interfaces for the analysis, annotation and enrichment components, 2) the Presentation Layer containing the components for developing personalized LinkedTV applications for end users and 3) the Linked Media Layer containing all the metadata generated including the services to access them, as well as management tools.

Connected to the Analysis and Annotation Layer there are editorial tools for 1) the Media Selection and Analysis Tool to select new videos for analysis and include them into the platform 2) the Annotation Tool to adjust automatically generated annotations and 3) the Hyperlinking Tool for the manual insertion of links associated to certain objects in the video.

The Presentation Layer provides the basis for the development of specific end user components and applications: 1) a Hypervideo Player for HTML5 to retrieve and view hyperlinked video, 2) a HbbTV compliant player for TV applications and 3) Specific applications to perform the LinkedTV Scenarios with individual user interfaces and features such as the media use case.

The first step has been to develop the HTML5 based Hypervideo Player. It combines media fragments together with annotations from the Annotation Layer and displays these on the video canvas. Video hotspots are used to allow editors and end users to find objects using layered technology. The visual hotspots also provide access to related content and other available media on the Web. The HTML5 technology allows the Hypervideo Player to be used on a broad range of systems, including PCs, Smart TVs and mobile devices.

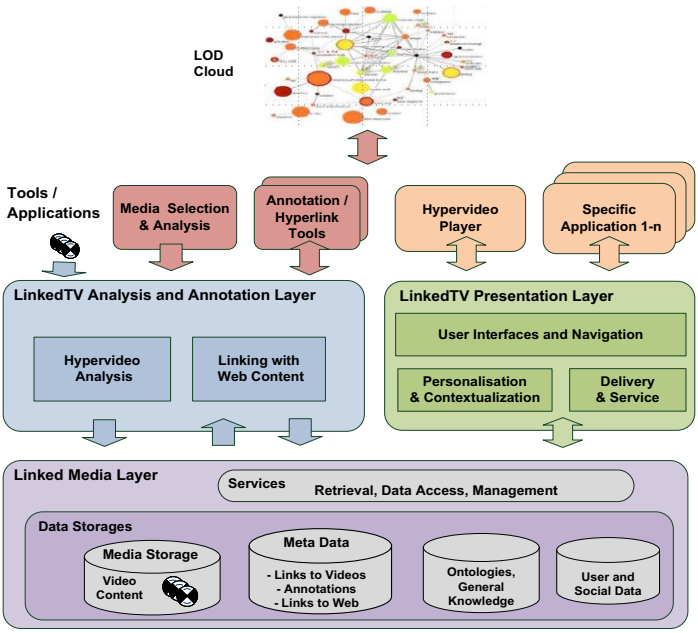

Fig. 4. Architecture overview

Fig. 5. Hypervideo player with hotspot

4 Research Challenges

The described architecture incorporates a number of components that must provide a certain level of quality and efficiency around specific functionalities that become the subject of project research challenges. These challenges can be categorized into the domains of (i) media analysis, (ii) cross-media interlinking, (iii) LinkedTV user interfaces and navigation, and (iv) LinkedTV personalisation and contextualisation.

Media Analysis [1]. Based on the scenarios envisioned within LinkedTV, we need to be (semi-)automatically deriving from the A/V content information about the concepts being present in that content. For example, persons could be identified via face and/or voice, and objects of interest should be detected and tracked. Due to the broad target domains it cannot be guaranteed that established (domain specific) databases of annotated media (for helping classifiers identify objects in new video material) contain enough instances of concepts present in new video material to analyse. This is why we need strong clustering and re-detection techniques so that an editor only needs to label a concept once and can automatically find other instances within the video itself, or within a larger set of surrounding videos. For keywords needed to tag the videos, both more abstract concepts and events should be recognized. Also needed for keywords, as well as for the named entity recognition, is an automatic speech recognizer whenever there are no subtitles available, or, if subtitles are given, forced alignment techniques to match the timestamps to the video on a word level rather than on an utterance level which might be to coarse-granular for our needs. Finally, larger videos should be temporally segmented based on their content to provide reasonable time-stamp limits for hyperlinks.

The partners of the LinkedTV consortium have access to state-of-the-art techniques that can pursue these goals. The main challenge will therefore be to interweave the individual analysis results into refined high-level information. For example, person detection can gain information from automatic speech recognition, speaker recognition and face recognition. Also, in order to find reasonable story segments in a larger video, one can draw knowledge both from speech segments, topic classification, and video shot segments. As a final example, video similarity can be

estimated with feature vectors carrying information from the concept detection, the keywords, the topic classification and the entities detected within the video.

In a first step, we already collected all semi-automatically produced analysis results into a single annotation file that can be viewed with a dedicated tool, and will use this in order to establish ground truth material on the scenario data. We already validated and extended various methods – object re-detection, shot segmentation, face detection and keyword extraction - and increased their accuracy in LinkedTV scenarios[2]. A project deliverable summarizes all first year achievements [7].

Cross-Media Interlinking. LinkedTV will use an ontology-based data model in order to represent all the information about television programs that needs to be managed by the system. This data model will be based on various well-established vocabularies from the multimedia and television domain.

First, LinkedTV receives legacy data either coming from the content itself or produced by the broadcaster. This data generally includes properties such as title, short description, format, duration, etc. that will be represented using the Ontology for Media Resources [2]. It includes as well broadcasting information that can be represented according to the BBC ontology [3] and the SchemaDotOrgTV vocabulary [4]. Second, the results of the analysis performed over the video in previous stages (shot detection, face recognition, ASR, etc.), which are currently available in XML like formats will be RDF-ized using appropriate additional ontology properties. The decomposition of media content into pieces will be directly addressed by the use of Media Fragment URIs [5]. Finally, the Open Annotation Core Data Model [6] will enable to associate real world concepts as annotations to media fragments and to state more information about this annotation such as its provenance or level of trust. In most of the cases, this information is scarce, incomplete and sometimes inaccurate. It is then necessary to retrieve additional contextual information from different external sources. These sources will mainly be datasets from the LOD cloud such as the BBC Programmes, LinkedMDB or Geonames, as well as various Social Networks where fresh media, reactions and opinions are available.

We have established the following workflow for managing the metadata in the LinkedTV system: first, the legacy metadata and the analysis results are serialized into RDF by performing certain translation processes between the original files and the corresponding vocabularies[3]. During this phase, some NER techniques are also applied over the ASR files in order to extract named entities from the video[4]. Then, once this data is already available in the knowledge base, LinkedTV iteratively executes background operations for retrieving missing information, enrich existing data, and interlink local entities with similar ones in other external datasets. At the end of the complete process, a complete RDF graph can be accessed by the LinkedTV player in order to show the viewer the information he needs. The LinkedTV Ontology was defined in a project deliverable [8].

---

[2] Several video demos are posted at `http://www.linkedtv.eu/demos-materials/online-demos/#core-media`

[3] See the online RDF Metadata Generator `http://linkedtv.eurecom.fr/tv2rdf`

[4] Using NERD `http://nerd.eurecom.fr`

User Interfaces and Navigation. Users are familiar with interaction with video in terms of using text to search for video fragments on the internet and of navigating the timeline of video using concepts such as "fast-forward". Users are also used to navigating the plethora of web pages on the internet, using search engines and bookmarks to find information they are looking for. The goal of LinkedTV from the user perspective is that s/he should be able to manipulate video material combining these two interaction paradigms. The challenge for the project is to ensure that the information, and entertainment, experience is enhanced, rather than frustrated by being presented with a bewildering assembly of vaguely related videos, where they lose both the advantages of passively watching edited video or being in control of finding specific clips of their own choice. Our research questions are to what extent we can provide links to usefully related video content without disrupting the viewing experience. At the same time we need to be able to provide unobtrusive interaction tools that allow the material to be explored freely. This led to a particular focus on second screen solutions. The Antiques Roadshow and news scenarios were used to explore the commonalities of the tools needed and specific ways of presenting the combination of functionalities to users in an appropriate LinkedTV UI [9].

As TV becomes more interactive, it moves closer to a gaming environment. While the goals of LinkedTV are specifically related to providing information to users, interaction devices more familiar in the gaming environment can be used to enhance the exploration experience. We also explore where these interaction devices, such as Kinect, can help the user retain a sense of control in a complex linked and time-based information environment. An example is a dance exploration scenario where we will explore the potential of interacting with the video space by dancing.

Personalisation and Contextualisation. The provision of recommendations for Media Fragments to end users requires the enhancement of current recommender technology. Innovative methods to derive semantic fingerprints from the fragment properties, surrounding objects, current scene and spoken text are needed to provide users with personalized related information while watching TV. LinkedTV has published a first user schema for capturing viewer's interests and a set of approaches to implicitly gather those interests via user interaction with the LinkedTV content as well as tracking attention and emotions during watching via a Kinect installation [10].

LinkedTV plans to show hyperlinked video on tablets and TV Sets. After a first demonstrator for TVs and tablets using HTML5[5] we want to show clickable video for TV devices based on hbbTV. This will require innovative solutions for features not initially provided by HTML5 (e.g. access to resources such as a broadcast video stream and its metadata) or hbbTV 2.0 (e.g. hyperlinks placed over dynamically moving objects or alternative interaction models such as pointing devices based on gesture control or second screens). This summary of research challenges reflects the need for the cross-European collaborative work of expert organisations in the respective domains, which is being enabled by the LinkedTV EU project.

5 See a screencast of the 1st year demo at `http://www.linkedtv.eu/demos-materials/online-demos/#scenarios`

5 Outlook for LinkedTV

Television is changing. However, uptake of the new Smart TVs will increase at a gradual rate and currently the Web-TV offering is not capturing the interest of the critical mass of viewers. Key growth is seen presently in the younger adult demographic with use of social TV apps in a second screen device. Finally, industry lock-in is largely seeking to limit third party OTT (over the top) services that would take revenue from existing customers (e.g. cable subscribers). In the next years connected TV platforms will not only become more present but the quality and intuitiveness of their UI experience will improve. The legacy media – TV and video content - industry will either co-opt the new technology within their own controlled services or be pushed out by innovative OTT services (cf. the history of the music industry and the Web). The LinkedTV vision – seamlessly interlinking Web and TV content in an unified interactive, audio-visual experience on the end device – has both technological and business barriers to overcome. The LinkedTV project is working on overcoming the technological barriers, and will also monitor the shifting TV business landscape. As the project comes to an end in April 2015, TV will already be very different from it has been traditionally, and LinkedTV will be ready to promote its vision for television, connected to the Web, its metadata and content, and to innovative services for analysis, annotation, linking, personalisation and presentation of such LinkedTV content.

Acknowledgements. This work is supported by the Integrated Project LinkedTV (www.linkedtv.eu) funded by the European Commission through the 7th Framework Programme (FP7-287911).

References

1. Stein, D., Apostolidis, E., Mezaris, V., de Abreu Pereira, N., Müller, J.: Semi-Automatic Video Analysis for Linking Television to the Web. In: Proc. FutureTV Workshop, at EuroITV Conference 2012, Berlin, Germany (July 2012)
2. http://www.w3.org/TR/mediaont-10/
3. http://www.bbc.co.uk/ontologies/programmes/2009-09-07.shtml
4. http://www.w3.org/wiki/SchemaDotOrgTV
5. http://www.w3.org/TR/media-frags/
6. http://www.openannotation.org/spec/core/
7. Mezaris, V., et al.: State of the Art and Requirements Analysis for Hypervideo. LinkedTV project deliverable D1.1 published at http://linked.tv/deliverables
8. García, J.L., Tröncy, R., Vacura, M.: Specification of lightweight metadata models for multimedia annotation. LinkedTV project deliverable D2.2 at http://linked.tv/deliverables
9. Leyssen, M., Traub, M., Hardman, L., van Ossenbruggen, J.: LinkedTV user interfaces sketch. LinkedTV project deliverable D3.3 at http://linkedtv.eu/deliverables
10. Tsatsou, D., Mezaris, V., Kliegr, T., Kuchar, J., Mancas, M., Nixon, L., Klein, R., Kober, M.: User profile schema and profile capturing. LinkedTV project deliverable D4.2 at http://linkedtv.eu/deliverables