VideoHypE: An Editor Tool for Supervised Automatic Video Hyperlinking

Lotte Belice Baltussen[1], Jaap Blom[1], and Roeland Ordelman[1,2]

[1] Netherlands Institute for Sound and Vision, Hilversum, The Netherlands
[2] University of Twente, Enschede, The Netherlands

Abstract. Video hyperlinking is regarded as a means to enrich interactive television experiences. Creating links manually however has limitations. In order to be able to automate video hyperlinking and increase its potential we need to have a better understanding of how both broadcasters that supply interactive television and the end-users approach and perceive hyperlinking. In this paper we report on the development of an editor tool for supervised automatic video hyperlinking that will allow us to investigate video hyperlinking in a real-life scenario.

Keywords: video hyperlinking, interactive television, video analysis, user studies, information extraction.

1 Introduction

Aiming towards richer interactive television experiences, broadcast companies are becoming increasingly interested in enriching television content with hyperlinks that connect the primary content to other data sources that could enhance the attractiveness of watching television in either a linear fashion or on-demand. By default, such links are currently created manually by broadcast companies' editorial departments. However, the identification of media-fragments [8] that can be used as *anchors* in the primary content, and the selection of appropriate link *targets* completely manually is labour intensive. Automating at least parts of the hyperlink generation process for television content is essential to be able to provide the users with a rich set of links that significantly enhance the user experience. A second argument for pulling technology into the link generation process is that manual hyperlink generation is inherently limited by a human editor's subjective view on anchor selection and limited view on possible target data sources. Note that the latter in return also confines the anchor selection process.

A baseline approach towards the automation of link generation is to deploy available time-labelled content descriptors such as subtitles, and automatically generated annotations (e.g., automatic speech recognition, visual concept detection). Entities such as names, people, places and objects can be extracted from the textual annotations to serve as anchor "candidates" that can be linked to other content sources. A curated white list of resources may define the domains that are considered for the link targets. The final step would then be to connect

the link anchors to the link targets. For example, consider a programme that is about *Rembrandt* and his painting *Night Watch*, which is displayed in the *Rijksmuseum* in *Amsterdam*. The entities "Rembrandt", "Night Watch", "Rijksmuseum" and "Amsterdam" can be extracted from the subtitles and could be identified unambiguously using Semantic Web URIs (Linked Data) [10] and linked to the Wikipedia articles about Rembrandt and the Night Watch, assuming that Wikipedia is part of the white list of link sources.

Although the use of a white list restricts the amount of anchors that can practically be linked, it is not the case that every anchor in a white list that *can* be linked is also a useful candidate. One can imagine that the usefulness of a candidate anchor depends among others on the content itself, user context, and the characteristics of the link target, such as its specificity and relevance. Ideally, the relevance of candidates for linking is determined automatically on the basis of context features. However, we need to have a better understanding of user behaviour in a video linking scenario to model the relevance of link candidates appropriately. We investigate user behaviour in a scenario in which an editorial department of a broadcast company creates links for an interactive television application, and evaluate how automatic link suggestion is controlled and perceived by the editors of the programme and how end-users that watch the programme in the interactive television application appreciate the links. In this paper we describe the first stage of the development of video hyperlink editor (*VideoHypE*) that uses rich, and partly automatically extracted content annotations for *supervised* hyperlink generation. The tool will be used and evaluated in real-life editorial link generation scenarios at broadcast companies.

The rest of this paper is structured as follows. In the next section (section 2) we briefly describe the underlying technology of the *VideoHypE* tool. In section 3 we report on a requirements elicitation session with editorial staff of a broadcaster interested in video hyperlinking and close with a future work section (section 4).

2   Link Suggestion Workflow

The envisaged *VideoHypE* tool will use the input of a processing chain incorporating audio and visual analysis and information extracting. The chain consists of shot detection for video segmentation into shots [6], spatial-temporal and visual concept detection [2], face detection [7], face clustering and face recognition for respectively detecting, grouping and tagging faces of persons in a video, and object re-detection for retrieving instances of pre-selected images within video frames. With respect to audio analysis technologies we will use speaker identification [4] for identifying certain pre-selected persons of interest related to the user scenarios, automatic speech recognition (ASR) [5], and audio fingerprinting for media synchronisation, e.g. for synchronising a second screen application with the main screen.

Based on ASR transcripts, existing programme metadata and subtitles, named entities are extracted, disambiguated and finally enriched with related content

(generally linked data resources). Combined with the previously compiled annotations the end-result of a video being processed is a number of annotated media fragments stored as RDF triples in a Virtuoso semantic repository[1]. The VideoHypE tool is built on top of this semantic repository and includes mechanisms to assess the relevance of the automatically generated links and push them to a user interface [1].

3   Hyperlink Editor Requirements

In order to develop a *VideoHypE* tool that is consistent with the needs of broadcasters engaging into video hyperlinking, we need user requirements from the editors of the interactive television applications that are going to use the tool in a real-life scenario. We are currently looking at two use scenarios for hyperlink suggestion: an interactive news scenario based on news content from a German broadcaster[2], and interactive documentary scenario on a Dutch programme called "Tussen Kunst en Kitch" (TKK), an antiques roadshow from Dutch public broadcaster AVRO. In this paper we focus on the scenario for the TKK programme.

To elicit requirements for the tool, we organised an interview session with the editorial members and technical staff of the TKK programme (further referred to as TKK staff) in which we explained the concept of extracting media fragments and suggesting hyperlinks and asked them about their view on addressing hyperlinking in such a manner. Point of departure was a mutual agreement on the fact that automatic video hyperlinking could be a helpful tool in an interactive television application scenario, either fully automated or in an editorial setting. With an eye on the current experimental state of technology, the TKK staff expects that some kind of supervision on the links that are created will always be required. One important reason for keeping control over the links is that broadcast companies fear that providing links that are not appreciated by end-users may harm the appreciation of the television programme.

The corrections of the hyperlinks can be done on several levels. First of all, the editor could merely accept and reject the hyperlinks. However, an extra level of control could be added that allows editors to also *add* hyperlinks not provided by the system. Another important aspect of the editor tool will be the granularity on which the work will take place. The system can provide hyperlinks on the level of shots, scenes or even spatial elements within shots and scenes (e.g. a painting within a shot). The TKK editors indicated that they would most likely prefer to work on chapter level, i.e. a scene of a TKK episode in which a specific art object is discussed. The reason for this is that it is easier for them to work in increments, instead of having to describe an episode as a whole. Furthermore, the TKK staff suggests to allow selection of entities for linking on a global level to increase efficiency. For example, they would like to select a specific entity

---

[1] `http://virtuoso.openlinksw.com/`
[2] We use a news show programme from the German broadcaster Rundfunk Berlin Brandenburg (RBB).

type, such as *Person: Rembrandt van Rijn* or *Location: Amsterdam* and accept or reject this type, which automatically results in the hyperlinks connected to these entities either being accepted or rejected for the entire chapter.

Obviously, the TKK staff are very interested in the expected precision of the link suggestion process, and whether this would help to provide better links using less human resources. They suggest that it could be useful to be able to control the cut-off point of a ranked list of links as a function of the amount of supervision that can be provided. In general, editorial staff would have about 20-30 minutes to work on this editor tool, but this can differ per episode. For example, when there is only limited time for supervision, the amount of suggested links should be small (low recall is fine) and have a high probability of being relevant (precision is important). When there is more time to supervise the suggested links it would be interesting to explore the link suggestions more abundantly for a bit more icing on the cake.

3.1   On the Definition of Relevance

The definition of precision and relevance is in this context a crucial issue. Although the concept of hyperlinking in general –based on people's experience with linking in webpages on the World Wide Web– is well-known, the unfamiliarity with the concept of video hyperlinking (see also [9]) complicates the discussion. During the interview session with the TKK staff, different perspectives on link relevance could be distinguished. For example, relevance can be regarded from a television production point-of-view where links should add to the intended "message" of the television programme. Clearly, producers have a detailed view on how they should reach specific audiences and providing video hyperlinks can be regarded as just an extension of the original medium. This perspective can have many gradations, from really having the creator in the loop to a broadcast company that demands control over their perceived identity.

Heading more towards the user playing a role in the definition of relevance, it was suggested that links could be typified as being relevant for a more general audience and relevant for a specific (type of) user. Interestingly, this typology may align well with the amount of supervision that is required. Obviously, it is less difficult to define guidelines on what an audience in general may be interested in with respect to links which would allow to focus supervision on a limited set of highly relevant "general purpose" links. Moving towards more specific (types of) users, supervision could be reduced to global levels (e.g., by defining acceptable links on the basis of user/audience types) or even omitted and left to application-side personalisation approaches.

Next to general and user-specific link relevance, also programme specific relevance was mentioned. For an antiques roadshow type of programme for example, context information on the time-period in which an art object was made is regarded as being very relevant. Note that looking at the programme level for assessing the relevance of a link is different from a content-based assessment as the former typically applies to all the programmes of this type whereas the latter applies to a specific programme item. On another level related to link relevance is

the difference between how links are perceived while watching a live programme or when watching it using an on-demand service. However, one can argue that the boundaries between live broadcast and on-demand are fading.

3.2   Link Targets

The choice for a set of link targets or link target domains, defined beforehand as being useful for linking in the context of a certain programming, plays a significant role in the process. When the set is small and confined in terms of entities that serve as anchors for linking (e.g., person names like "Rembrandt van Rijn" with biographic information available in a white-listed source like Wikipedia), the precision of automatic link generation can be expected to be higher than would be the case when there are several link target domains that allow different entity types for establishing links, as the limitations can serve as a restriction on the variability in the processes of anchor selection (e.g., only focus on automatic person detection) and link target suggestion (e.g., disambiguation needs to be done only at person level).

Hence, both from a technical point of view and from the viewpoint of a broadcaster, the definition of a "white-list" of one or more link target domains is an important step. In the discussions with the TKK staff the possibility to use self-curated content sets, such as a photo database of art objects taken on the location where the antiques roadshow programme was recorded, was also addressed. Another suggestion was to use the frequency of a single link suggested within a certain time frame as a way to collect evidence on the probability that a certain link is relevant. Finally, from the interview session some ideas on addressing the user were made such as the possibility to bookmark links and/or media fragments, and to have the user in the feedback loop for assessing relevance and/or rank links (e.g., by collecting thumbs-up/thumbs-down information, or use implicit feedback via the clicks). Note however, that the latter would be less feasible for video hyperlinking in live programmes. Using different colours for different link types was another suggestion.

4   Future Work

Using the link extraction workflow as described in section 2 and the requirements provided by the broadcaster's editorial team, the *VideoHypE* tool will be built in close collaboration with the TKK staff. A working version of the tool will be presented to them, and the tool will be formally evaluated with TKK editorial members. The output of the tool –the links suggested by the link extraction workflow and validated by the TKK editors using the tool– will be evaluated with end-users watching the antiques roadshow programme in an interactive application featured with video hyperlinks. The evaluation with editors will provide us with more insight into a supervised approach with respect to video hyperlink generation and, in its slipstream, with the performance levels of automatic link extraction. The evaluation with end-users will give us the opportunity to develop

a better understanding about user preferences, personalisation and appreciation of video hyperlinking.

References

1. Baltussen, L.B., Leyssen, M.H.R., Ossenbruggen, J., van, O.J., Blom, J., Leeuwen, P., van, H.L.: Antiques Interactive. In: EuroITV 2012 – Adjunct Proceedings, pp. 23–24. Fraunhofer Institute for Open Communication Systems, FOKUS, Berlin (2012)
2. Stein, D., Apostolidis, E., Mezaris, V., Patz, N., Müller, J.: Semi-Automatic Video Analysis for Linking Television to the Web. In: EuroITV 2012 – Adjunct Proceedings: FutureTV Workshop, pp. 154–161. Fraunhofer Institute for Open Communication Systems, FOKUS, Berlin (2012)
3. Romero, L.P., Traub, M.C., Leyssen, H.R., Hardman, L.: Second Screen interactions for Automatically Web-enriched Broadcast video. In: Submitted to: ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013) "Exploring And Enhancing the User Experience for Television" Workshop, Paris, France (2013)
4. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing 10, 19–41 (2000)
5. Schneider, D., Schon, J., Eickeler, S.: Towards Large Scale Vocabulary Independent Spoken Term Detection: Advances in the Fraunhofer IAIS Audiomining System. In: Proc. SIGIR, Singapore, pp. 34–41. CTIT, Enschede (2008)
6. Tsamoura, E., Mezaris, V., Kompatsiaris, I.: Gradual transition detection using color coherence and other criteria in a video shot meta-segmentation framework. In: 15th IEEE International Conference on Image Processing, ICIP 2008, pp. 45–48 (2008)
7. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: International Conference on Image Processing Proceedings, vol. 1, pp. I-900–I-903 (2002)
8. Troncy, R., Leeuwen, P., van, G.J.: Specification of the Media Fragment URI scheme, D2.1, LinkedTV (2012),
   `http://www.linkedtv.eu/wp/wp-content/uploads/2012/11/`
   `LinkedTV_D2.1_Specification_of_the_Media_Fragment_URI_scheme.pdf`
9. Aly, R., McGuinness, K., Kleppe, M., Ordelman, R., O'Connor, N.E., de Jong, F.: Link Anchors in Images: Is there Truth? In: Proceedings of the 12th Dutch Belgian Information Retrieval Workshop (DIR 2012), pp. 1–4 (2012)
10. Lašek, I., Kliegr, T., Dojchinovski, M., Sahuguet, M., Rizzo, G., Huet, B., Troncy, R.: Specification of Web Mining Process for Hypervideo Concept Identification (2012)