Multilingual Semantic Linking for Video Streams: Making "Ideas Worth Sharing" More Accessible

Daan Odijk
ISLA, University of Amsterdam
Amsterdam, The Netherlands
d.odijk@uva.nl

Edgar Meij
Yahoo! Research
Barcelona, Spain
emeij@yahoo-inc.com

David Graus
ISLA, University of Amsterdam
Amsterdam, The Netherlands
d.p.graus@uva.nl

Tom Kenter
ISLA, University of Amsterdam
Amsterdam, The Netherlands
tom.kenter@uva.nl

ABSTRACT

This paper describes our submission to the Developers Challenge at WoLE2013, "Doing Good by Linking Entities." We present a fully automatic system which provides intelligent suggestions in the form of links to Wikipedia articles for video streams in multiple languages, based on the subtitles that accompany the visual content. The system is applied to online conference talks. In particular, we adapt a recently proposed semantic linking approach for streams of television broadcasts to facilitate generating contextual links while a TED talk is being viewed. TED is a highly popular global conference series covering many research domains; the publicly available talks have accumulated a total view count of over one billion at the time of writing. We exploit the multilinguality of Wikipedia and the TED subtitles to provide contextual suggestions in the language of the user watching a video. In this way, a vast source of educational and intellectual content is disclosed to a broad audience that might otherwise experience difficulties interpreting it.

1. INTRODUCTION

The way we watch television is changing. Broadcasts are increasingly consumed on or with an interactive device in the vicinity. A survey by Nielsen, consisting of 12k mobile users, found that 70% of tablet owners and 68% of smartphone owners use their devices whilst watching television [5]. These observations indicate that the television audience interacts more and more with the content they are consuming, thus extending the viewing experience. The interaction includes not only producing and consuming broadcast-specific social media, but may also provide additional content created exclusively for the interactive device, such as additional background information.

Due to the popularity of broadcasts, not only on television but also on mobile devices and on the Internet, people are increasingly gaining easy access to an amount of information that was previously inconceivable. All of these technologies are readily available and extensively used, but understanding everything that is being broadcast can be a difficult task. Therefore, an automatic method for providing background information is called for, preferably in the users' native language. We propose to use a state-of-the-art approach for semantic linking on television broadcast data to improve the user experience based on our recent work [4, 6]. In particular, our aim is to provide contextual suggestions to a user viewing a broadcast in the form of links to Wikipedia articles in her own language in a fully automatic manner.

We provide a demo implementation of our system for TED talks. TED is a global series of conferences that started out in 1984, with the aim of bringing together people from technology, entertainment, and design. Since its inception, TED has expanded rapidly and it is currently one of the world's most famous conference series with a significant number of local and global initiatives, including TED Conference, TEDGlobal, TEDWomen, and TEDx. The most well known instance, however, is the online repository of so-called TED talks.[1] TED talks aim to share what happens at TED with the world or, using their own slogan, share "ideas worth spreading." They have attracted a global audience in the millions and a total view count of over one billion. As such, they offer a rich source of information from all kinds of research areas including technology, entertainment, design, business, science, and global issues.

There are two caveats however. First, TED talks are typically presented in English—although subtitles are usually provided in a large number of languages. Given the pace in which some talks are presented, having to keep up with reading the closed captions and simultaneously watching the presenter and her/his slides can be a daunting task. Second, although the talks should be accessible to the general audience, some of the talks may contain jargon or technical terms and can thus be quite inaccessible. To make these talks more understandable, providing relevant background information is essential. In this demo, we use a state-of-the-art approach for semantic linking on television broadcast data to improve

[1] See `http://www.ted.com/`.

the user experience [6]. Our approach is based on stream-based semantic linking, which takes as input a stream of text from closed captions and links phrases in this stream to Wikipedia articles. Given the large number of languages in which Wikipedia is available, our aim is to provide semantic links in the language of choice. A particularly interesting case is linking to articles in Simple English,[2] which lowers the barrier for native English speakers.

The remainder of this paper is organized as follows. In Section 2 we present the literature on which our application is based. Section 3 details the application and its functionalities, and we end with a concluding section.

2. BACKGROUND

A significant task in building and maintaining the Semantic Web is link generation. Links allow a person or machine to explore and understand the web of data more easily: when you have linked data, you can find related data. The Linking Open Data (LOD) initiative extends the web by publishing various open data sets and by setting links between items (or concepts) from different data sources in a (semi-)automated fashion. The resulting data commons is termed the Linking Open Data cloud and provides a key ingredient for realizing the Semantic Web. Unstructured data resources—such as textual items [1–4] or multimedia items [6, 7]—can be enriched by mapping their content to structured knowledge repositories like the LOD cloud. This type of enrichment may serve multiple goals, such as explicit anchoring of the data sources in background knowledge, enabling new forms of intelligent search and browsing; authors or readers of a piece of text may find mappings to the LOD cloud to supply useful pointers. Examples include linking to concepts capturing or relating to the contents of the document.

For live television, edited broadcast-specific content is hard to prepare in advance. Preparing such content can be very time-consuming, resulting in a very limited number of television shows that actually provide such facilities. We present an approach for automatically generating links to background information in real-time. Our approach automatically generates links to Wikipedia. This process is commonly known as semantic linking and has received much attention in recent years [1–4, 6, 7]. Such links are typically explanatory, enriching the link source with definitions or background information. In the setting of television broadcasts, the links can provide fully automatic go-read-here functionality for streaming video data.

We base our semantic linking approach for TED talks on transcripts used for subtitles, thereby effectively casting the task as one of identifying and generating links for elements in the stream of subtitles. Note that the subtitles do not actually have to be shown; they are merely used as a textual stream to generate links that may then be shown through a visual representation. Traditional document-based approaches to semantic linking are not suited for this task, as links need to be generated continuously from this stream. On the other hand, using semantic linking approaches for short text, that completely ignore the streaming nature of the material, would mean missing out on important contextual signals. Hence, in order for our semantic linking approach to be effective in the context TED talks, it needs to be fast, able to disambiguate between candidate links in real-time, and leverage streaming data so as to capture context. Furthermore, since the viewer is dividing her attention between the actual talk and the suggested links, the information that is being offered needs to be of high quality, i.e., have high precision. We build upon the approach presented in [6] and use the so-called "sense probability" to generate the links. The next section provides a more detailed description of this method and the application.

[2]See `http://simple.wikipedia.org/`.

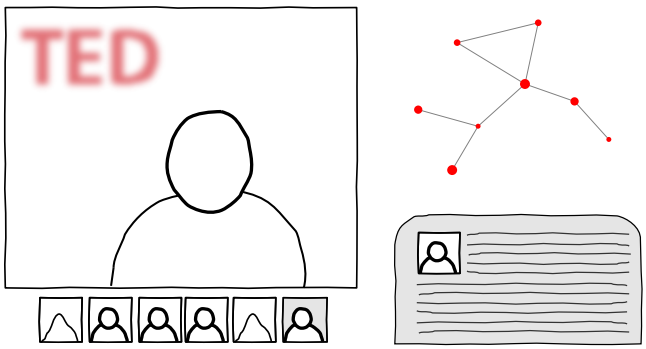

Figure 1: Sketch of the interface of the Semantic TED demo.

3. APPLICATION DESCRIPTION

A sketch of the interface of our Semantic TED demo is shown in Figure 1; the demo application itself is online. Our demo performs real-time semantic linking of textual streams; we link the transcripts of TED talks to Wikipedia articles. While the user watches a TED talk, links are presented to Wikipedia pages relevant to the topics being discussed in the talk. While the actual linking is done based on transcripts that are used as subtitles, the subtitles themselves do not have to be visible while watching the talk.

To enable us to present the TED talks in our envisioned way, we use a proxy server to the official TED website. The TED starting page is replaced by one of our own making that embeds the official starting page. This allows us to modify the webpage where a talk is presented to also present the related links we provide. No preprocessing is needed for our demo application. When a user watches a TED talk in our demo, it is processed and presented on-the-fly.

In the transcripts, phrases are identified as potential links to Wikipedia articles (candidate articles), based on a Wikipedia dump of April 12, 2012. We use lexical matching on all constituent n-grams of each sentence to produce these links. To select what links are relevant we use a heuristic function called sense probability (SP) that has been shown to perform well on streams of television broadcasts [6]. It captures both whether a link is correctly disambiguated and the likelihood of a phrase being a link. Sense probability is computed as follows:

$$SP(a,w) = \frac{|L_{a,w}|}{\sum_{w' \in W} n(a,w')}, \quad (1)$$

where $L_{a,w}$ denotes the set of all links with anchor text a and target Wikipedia article w and $n(a,w)$ is the number of occurrences of the text a in w. We use a threshold of 0.2 for links to be presented to a viewer.

The links produced in this manner are displayed in two ways. First, we represent the stream of links under the video player by a thumbnail and title of each link, taken from the corresponding Wikipedia article. Second, we display the links in a network structure to the right of the video player. Related links are connected by leveraging the Wikipedia graph structure, defined by the hyperlinks between Wikipedia articles. In the Wikipedia graph, each Wikipedia article is considered a node—edges are drawn when articles refer to one another by hyperlinks in the article's body. In our network visualization, edges between two links are drawn in a force-directed graph when either of the links' articles link to the other. In this way, related links form natural clusters, providing an intuitive overview of 'core' topics of a talk.

Furthermore, to provide visual cues as to the "importance" of links in the network, we reflect the sense probability by the font-size of its title in the network visualization; larger titles reflect links that are correctly disambiguated and likely to be relevant.

Finally, to ensure the graph isn't cluttered by an overly large number of nodes, we prune concepts that fall under a threshold of a weight that is calculated by combining a links' sense probability with that of its related links in the context, as defined below. Singleton nodes are thus more likely to be pruned than highly connected nodes, leaving a coherent set of concepts. The weight of each node is computed as follows:

$$W(a,w) = SP(a,w) + \lambda \sum_{\{a',w'\} \in N(W)} SP(a',w') \qquad (2)$$

where $N(w)$ denotes the set of all incoming and outgoing links for target Wikipedia article that occur in the context of this talk. We use $\lambda = 0.6$ and also a threshold of 0.6 for links to be presented in the graph.

Clicking on either nodes in the network or thumbnails in the stream results in its Wikipedia article being opened in a new browser window, effectively enabling the "read more" functionality.

In order to enable multi-linguality, we use the inter-language Wikipedia links to "translate" the English Wikipedia articles into the language of choice. Currently, we support Spanish, French, and Dutch although more languages can be added easily.

4. CONCLUSION AND FUTURE WORK

In this paper we have described our submission to the "Doing Good by Linking Entities" Developers Challenge at WoLE2013. We have applied our semantic linking system for live broadcast streams to enrich publicly available online symposia. In particular, we utilize the transcripts of TED talks to generate links to relevant Wikipedia pages to make these talks more easily understandable. Furthermore, we exploit the inherent multi-linguality of Wikipedia to provide semantic links in the language of choice, thus further lowering the barrier to understand these symposia.

We can further improve the quality of linking by analyzing the user interactions with the demo, e.g. by interpreting the clicks on provided links as positive feedback for the quality of the link. In [6], we have shown that we are able to significantly improve the quality of links to related background information using a supervised machine learning approach in a similar setting in realtime.

As the system is applicable to any kind of (live) stream data, it can easily be employed in different settings such as live political speeches, online educational programs, etc.

ACKNOWLEDGEMENTS

This research was partially supported by the European Union's ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme, CIP ICT-PSP under grant agreement nr 250430, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements nr 258191 (PROMISE Network of Excellence) and 288024 (LiMoSINe project), the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.061.814, 612.061.815, 640.004.802, 727.-011.005, 612.001.116, HOR-11-10, the Center for Creation, Content and Technology (CCCT), the Hyperlocal Service Platform project funded by the Service Innovation & ICT program, the BILAND project funded by the CLARIN-nl program, the Dutch national program COMMIT, the ESF Research Network Program ELIAS, the Elite Network Shifts project funded by the Royal Dutch Academy of Sciences (KNAW), and the Netherlands eScience Center under project number 027.012.105.

REFERENCES

[1] D. Graus, T. Kenter, M. Bron, E. Meij, and M. de Rijke. Context-based entity linking - university of amsterdam at tac 2012. In *TAC 2012*, 2013.

[2] E. Meij, M. Bron, L. Hollink, B. Huurnink, and M. de Rijke. Mapping queries to the Linking Open Data cloud: A case study using DBpedia. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):418 – 433, 2011.

[3] E. Meij, M. Bron, B. Huurnink, L. Hollink, and M. de Rijke. Learning semantic query suggestions. In *Proceedings of the 8th International Conference on The Semantic Web*, ISWC 2009, 2009.

[4] E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM 2012, 2012.

[5] Nielsen. In the U.S., tablets are TV buddies while ereaders make great bedfellows, May 2012. `http://bit.ly/L4lf9E` [Online; accessed May 2012].

[6] D. Odijk, E. Meij, and M. de Rijke. Feeding the second screen: Semantic linking based on subtitles. In *OAIR '13*, To appear.

[7] C. Snoek, B. Huurnink, L. Hollink, M. de Rijke, G. Schreiber, and M. Worring. Adding semantics to detectors for video retrieval. *IEEE Transactions on Multimedia*, 9(5):975–986, 2007.