

NexGenTV: Providing Real-Time Insight during Political Debates in a Second Screen Application

Olfa Ben Ahmed¹, Gabriel Sargent², Florian Garnier³, Benoit Huet¹ Vincent Claveau²,
Laurence Couturier³, Raphaël Troncy¹, Guillaume Gravier², Philémon Bouzy³, Fabrice Leménorel⁴
¹ EURECOM, ² CNRS/IRISA, ³ AVISTO, ⁴ WildMoka

benoit.huet@eurecom.fr, vincent.claveau@irisa.fr, laurence.couturier@avisto.com, fabrice@wildmoka.com

ABSTRACT

Second screen applications are becoming key for broadcasters exploiting the convergence of TV and Internet. Authoring such applications however remains costly. In this paper, we present a second screen authoring application that leverages multimedia content analytics and social media monitoring. A back-office is dedicated to easy and fast content ingestion, segmentation, description and enrichment with links to entities and related content. From the back-end, broadcasters can push enriched content to front-end applications providing customers with highlights, entity and content links, overviews of social network, etc. The demonstration operates on political debates ingested during the 2017 French presidential election, enabling insights on the debates.

CCS CONCEPTS

• **Information systems** → **Multimedia content creation; Video search; • Computing methodologies** → *Information extraction; Visual content-based indexing and retrieval;*

1 INTRODUCTION

Television is undergoing a revolution with the convergence of broadcast and Internet diffusion. People watch TV on their computers, tablets or smartphones. At the same time, they use those devices to search for information related to the program (second screen usage) and share opinions on social networks (social TV). Social networks are also widely used to share short snippets: funny sequence, controversy in a debate, key point in a sport's program, etc. Facing this situation, the NexGenTV project¹ aims at developing authoring tools for broadcasters to easily develop second screen applications and facilitate social TV. To do so, there is a need for near real-time automated tools to easily identify clips of interest, describe their content, and facilitate their enrichment and sharing.

We present a system for the repurposing of political debates on second screen and social TV. The system implements a complete workflow from broadcast stream to front-end applications on mobile devices, taking advantage of the synergy between content-based multimedia analysis and social network analytics. Political debates

constitute a prototypical use-case for second screen and social TV features: accessing information on politicians on second screen during the debate is now common practice; there are a lot of social reactions to political debates; many media actors seek to reuse short video excerpts, e.g., to illustrate an online newspaper article; people also easily share excerpts on which they comment; etc.

The NexGenTV platform consists of two main components. The back-office ingests TV streams and provides advanced content segmentation, description and enrichment features so as to rapidly select clips or information to share. These features rely on a combination of social network analytics, content-based multimedia analysis and ontological description. The back-office is used to feed in almost real-time a front-end application where one can see the clips selected in the back-office along with additional information automatically added. The front-end can be sought of as a mobile app devoted to one political debate, where one receives push notifications of enriched interesting sequences and can explore and share these sequences as they build.

2 CONTENT ANALYSIS TECHNOLOGY

The NexGenTV application builds on core functionalities to segment, describe and enrich TV stream based on analysis of the stream content and social network analytics.

Face detection and recognition. Persons appearing on screen are central elements of debates, e.g. to quantify the appearance time of politicians, to link their interventions to entity references, to identify their circle of friends [10], etc. A person recognition pipeline, including face detection, alignment and recognition, enables to find and identify key personalities in a program. A real-time face tracking system is used to detect and track multiple faces. Each of the detected face track is embedded in a low-dimension feature space with Facenet [9], on top of which support vector classifiers were built to recognize 35 politicians and journalists.

Speaker diarization. Speaker turns are helpful cues for content segmentation. This is specially true in political debates where a speech turn frequently corresponds to a politician developing his ideas on a topic. The NexGenTV platform implements a standard bottom-up clustering approach to speaker diarization. Hidden Markov models are used to segment speech portions into pseudo-sentences. Two stage clustering using Gaussian mixture models [2] is used to group pseudo-sentences into speaker turns. Though not state-of-the-art, this system is however rather fast and exhibit sufficient performance.

Term extraction. Keyword (term) extraction is crucial to characterize a segment for description and categorization. We use a combination of symbolic and numerical approaches thus able to extract multi-word expressions. All nominal compounds (e.g., noun,

¹NexGenTV is supported by the FUI Program of Banque Publique d'Investissement.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MM '17, October 23–27, 2017, Mountain View, CA, USA

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4906-2/17/10.

<https://doi.org/10.1145/3123266.3127929>

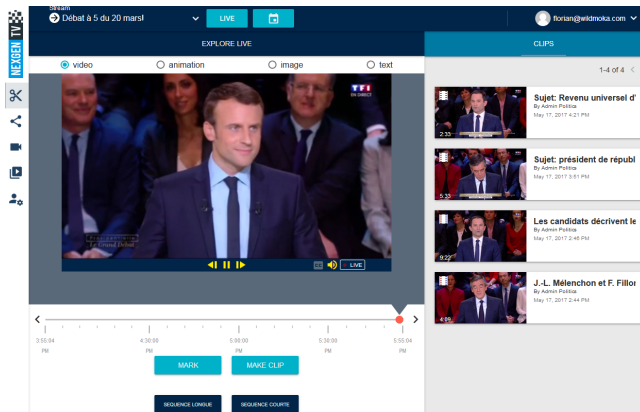


Figure 1: Screenshot of the back-office interface.

noun adj) are extracted and normalized to account for small variations. Each of the resulting candidates is evaluated based on its frequency and discriminativeness, as measured with the Okapi-BM25 weighting scheme [8]. Normalized compound sequences are finally ranked from the most relevant to the least.

Content hyperlinking. One distinguishing feature of second screen applications is the ability to provide links from a clip to other related clips, a task that we refer to as hyperlinking [1, 4, 5]. In NexGenTV, hyperlinking exploits subtitles. State-of-the-art document embedding [7], trained on French newspapers, is used to represent each clip by a vector. Using a database of such clip vectors, approximate nearest neighbor search techniques are used to efficiently retrieve on the fly a small number of related entries in the database for a given clip, the latter acting as a query.

Tweet collection and analysis. Social media is an important source of insight for broadcasted political events. A large number of persons share their views on the debate in quasi real-time, making it possible to analyze the TV broadcast in light of social reactions. Messages related to a program are collected on Twitter based on predefined hashtags and keywords. Relevant named entities are then extracted combining regular expressions and conditional random fields. Sentiment analysis [6] is also performed on each tweet, allowing to detect the polarity (neutral, positive, negative or mixed) relying on a recurrent neural network. This approach has been shown to obtain state-of-the-art results while being robust enough to handle grammatical variability.

3 INTERFACES

The NexGenTV platform has two interfaces: the back-office desktop interface allows TV channels to easily select and enrich content, partially automating the selection and enrichment of key clips to push; the front-end consists of a mobile application receiving selected content and enrichment from the back-office.

TV streams are ingested within the back-office where the main screen provides single-click clip selection, exploiting the result from speech and speaker turn detection: options are provided to select long or short clips and to easily adjust the boundaries of the clip. A list of clips already selected also appears in the main screen, as illustrated in Fig. 1. After selecting a clip, the edition interface enables adding information to the clip before pushing the enriched clip to the front-end application. An initial description

of the clip is automatically generated thanks to term extraction from subtitles, entity linking leveraging from face recognition and topic characterization [3] exploiting an ontological description of political life during the 2017 presidential election. Combining these elements, links can also be made to a description of the political program of the candidate for the clip's topic. Content linking also provides potential links to related clips or to additional sources previously ingested, e.g. previous debates or news shows. The back-office editing screen enables to re-arrange all these elements and the selection of relevant links before publication to the front-end.

In the front-end interface, new clips appear on the timeline as they are published from the back-office. They can be viewed, along with their description and with links to entries in our knowledge base (e.g., the bio of a politician, description of events of interest, program of a candidate on a given topic) and to related content. The front-end application also enables to view real-time statistics on the debate: time of speech or visual presence per candidate, amount of reactions on Twitter with clips at key instants, popularity for each candidate from opinion mining.

The demonstration operates on a collection of about 192 h of videos broadcasted by major French TV channels, totaling around 100 h of political debates and 92 h of TV news for enrichment. The average length of a debate is approx. 3 h, varying from less than 30 m to almost 4 h. The database for enrichment of a clip with hyperlinks consists of clips that were extracted from the debates and from the news shows. The ontology of the 2017 presidential election was created from scratch and describes the bio of all politicians involved, a list of topics discussed in the campaign (relation with EU, wealth tax, etc.) and synthetic records of the political program of each candidate on each topic.

REFERENCES

- [1] Evlampios Apostolidis, Vasileios Mezaris, Mathilde Sahuguet, Benoit Huet, Barbora Červenková, Daniel Stein, Stefan Eickeler, José Luis Redondo Garcia, Raphaël Troncy, and Lukáš Pikora. 2014. Automatic fine-grained hyperlinking of videos within a closed collection using scene segmentation. In *ACM Multimedia 2014*. ACM, 1033–1036.
- [2] Mathieu Ben, Frédéric Bimbot, and Guillaume Gravier. 2005. A model space framework for efficient speaker detection. In *European Conference on Speech Communication and Technology*.
- [3] Vincent Claveau and Sébastien Lefèvre. 2015. Topic segmentation of TV-streams by fast transform and vectorization. *Computer Speech and Language* 29, 1 (2015), 63–80. <https://doi.org/10.1016/j.csl.2014.04.006>
- [4] Maria Eskevich, Gareth J.F. Jones, Shu Chen, Robin Aly, Roeland Ordelman, Danish Nadeem, Camille Guinaudeau, Guillaume Gravier, Pascale Sébillot, Tom De Nies, Pedro Debevere, Rik Van de Walle, Petra Galušćáková, Pavel Pecina, and Martha Larson. 2013. Multimedia Information Seeking through Search And Hyperlinking. In *Intl. Conf. on Multimedia Retrieval*.
- [5] Maria Eskevich, Huynh Nguyen, Mathilde Sahuguet, and Benoit Huet. 2015. Hyper Video Browser: Search and Hyperlinking in Broadcast Media.. In *ACM Multimedia 2015*. ACM, 817–818.
- [6] Grégoire Jadi, Vincent Claveau, Béatrice Daille, and Laura Monceaux-Cachard. 2016. Evaluating Lexical Similarity to build Sentiment Similarity. In *Language and Resource Conference, LREC, portoroz, Slovenia*.
- [7] Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *International Conference on Machine Learning*.
- [8] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Found. Trends Inf. Retr.* 3, 4 (April 2009), 333–389. <https://doi.org/10.1561/15000000019>
- [9] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 815–823.
- [10] Shujun Yang, Lei Pang, Chong-Wah Ngo, and Benoit Huet. 2016. Serendipity-driven Celebrity Video Hyperlinking. In *ACM ICMR'16*. ACM, New York, NY, USA, 413–416. <https://doi.org/10.1145/2911996.2912029>