Linking Entities for Enriching and Structuring Social Media Content

Raphaël Troncy
EURECOM
Biot, France
raphael.troncy@eurecom.fr

ABSTRACT

Social media platforms such as Twitter, Facebook or LinkedIn become a reliable source of news and play a key role for being aware of events around the world. Using social media to recognize, enrich or summarize events is however very challenging. In the first part of this talk, we will present ADEL, a novel hybrid architecture for an adaptive entity linking system, that combines methods from the natural language processing, information retrieval and semantic fields. The framework enables to link all the mentions occurring in a text to their entity counterparts in a knowledge base. It is modular and adaptive since it enables to process text written in different languages and of different kind (newswire, tweets, blog posts, etc.) while entities can be of common types (PERSON, LOCATION and ORGANIZATION) or specific ones (dates, numbers) and be disambiguated in generic or specialized knowledge bases. We will show how ADEL can outperform the state-of-the-art systems on the reference NEEL challenges that happens in the yearly #Micropost workshop (2014-2016).

In the second part of this talk, we will present a framework that can collect microposts from more than 12 social platforms and that contain media items, as a result of a query – for example a trending event. We will then show how we can automatically create different visual storyboards that reflect what users have shared about this particular event. The visualization emphasizes the different aspects of storyboards. A graph view shows the relationships between microposts and topics that we automatically extract, while the timeline view emphasizes the time dimension. The user can watch and interact with the summarized view of all the topics or select a particular one with the additional details. In addition, the states of different views are persistent through the URLs which makes easy sharing possible.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2892109.

BIO

Dr. Raphaël Troncy[1] is an Associate Professor in the Data Science Department of EURECOM leading the Multimedia Semantics group. He is also co-chair of the W3C Media Fragments Working Group and W3C Incubator Group on Multimedia Semantics, contributes to the W3C Web Annotations Working Group and numerous W3C Community Groups such as Schema.org or Open Linked Education. He is working on information extraction (project ASRAEL[2] and in particular named entity linking (framework NERD[3]), ontology modeling in the multimedia domain (project DOREMUS[4]) and large scale data integration. Applications range from smart cities (project 3cixty[5]) to social TV and second screen (projects LinkedTV[6] and NexGen-TV[7]) where semantic web, information extraction and multimedia analysis technologies are used together.

Acknowledgments

This work has been partially supported by Bpifrance within the NexGen-TV Project, under grant number F1504054U.

1. REFERENCES

[1] J. Plu, G. Rizzo, and R. Troncy. Revealing Entities from Textual Documents Using a Hybrid Approach. In *3rd International Workshop on NLP & DBpedia*, 2015.

[2] G. Rizzo, M. van Erp, and R. Troncy. Benchmarking the Extraction and Disambiguation of Named Entities on the Semantic Web. In *9th International Conference on Language Resources and Evaluation (LREC'14)*, 2014.

[1] http://www.eurecom.fr/~troncy
[2] http://asrael.limsi.fr/
[3] http://nerd.eurecom.fr/
[4] http://doremus.org/
[5] http://3cixty.com/
[6] http://www.linkedtv.eu/
[7] http://nexgentv.fr/