ELEVATE-Live: Assessment and Visualization of Online News Virality via Entity-Level Analytics

Govind[(✉)], Céline Alec, and Marc Spaniol

Department of Computer Science, Université de Caen Normandie,
Campus Côte de Nacre, 14032 Caen Cedex, France
{govind,celine.alec,marc.spaniol}@unicaen.fr

Abstract. Recent research has shown significant progress in forecasting the impact and spread of societal relevant events into online communities of different languages. Here, raising contents to the entity-level has been the driving force in "understanding" Web contents. In this demonstration paper, we present a novel Web-based tool that exploits entity information from online news in order to assess and visualize their virality.

Keywords: Entity-level web analytics · Web semantics
Analytics interface

1 Introduction

"Viral News" is a standing term that describes news that receives perception beyond average and, thus, spread at high speed and/or extremely wide. In particular, the Web allows a potentially global diffusion in almost zero time. However, different notions of virality exist in terms of speed, outreach, etc. with respect to the "importance" of a news article. While "importance" is still highly subjective and context - respectively - community dependent, the actors (named entities) involved are valuable indicators of the content's "inherent semantics". For instance, a report about the BREXIT and its consequences for Britain and its European partners is likely to contain named entities, such as politicians like `Theresa May` or `Emmanuel Macron`, organizations such as the `European Banking Authority` and the `European Commission` as well as cities respectively countries such as `Frankfurt`and `Luxemburg`. While country or city names are straightforward indicators for the importance of an article with respect to the mentioned place, it requires semantics to derive this information for persons or institutions. With the emergence of knowledge bases (KBs) such as DBpedia [1] or YAGO [5,8] and methods for named entity disambiguation [6] we are able to exploit semantics of Web contents automatically and interpret them accordingly.

In this demonstration paper we introduce ELEVATE-live, an extension of our ELEVATE framework [2], providing a Web-based user interface allowing its users an entity-level assessment and visualization in order to explore the

interdependencies between Web news articles and geo-locations. To this end, our paper makes the following contributions by:

- incorporating the ELEVATE-framework and raising Web contents to the entity-level for semantic analytics;
- exploiting KBs in order to reveal non-trivial interdependencies between named entities contained and associated countries;
- providing a Web interface to study the "virality" of news articles with respect to countries concerned and vice versa.

2 Overview on ELEVATE-Live

ELEVATE-live is a conceptual enhancement of the ELEVATE framework [2] allowing the assessment and visualization of Web data by the example of online news articles. To this end, we "semantify" Web contents by harnessing location information associated with named entities and aggregating them for further analytics. The ultimate step is an analytics interface that allows exploring the "virality" w.r.t. the associated countries. Figure 1 highlights the five steps of data processing in ELEVATE-live, which will be explained subsequently.
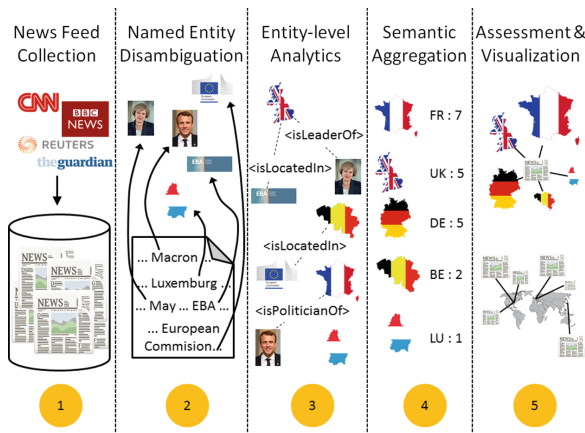

Fig. 1. Conceptual approach of the ELEVATE-live pipeline illustrated by a "BREXIT" related news article

(1) News Feed Collection
In an initial step, ELEVATE-live monitors the feeds of various online news agencies such as CNN, BBC, Reuters, etc. The Web contents are then preprocessed in order to obtain the plain news articles.
(2) Named Entity Extraction and Disambiguation
Subsequently, we employ AIDA [6] in order to reveal the named entities contained in a news article. By doing so, we raise each article to the entity-level.

(3) Entity-level Analytics

Next, the named entities contained in a news article are analyzed in order to gather location related information. To accomplish this, we utilize country- and organization-centric YAGO relations, such as `isLocatedIn`, `livesIn`, `worksAt`, etc. As there are potentially many countries associated with a named entity (via different relations), we pursue two strategies of knowledge base discovery:

  (1) Breadth-first-search: stopping when "hitting" an entity of type `country`
  (2) Depth-first-search: revealing all `countries` associated with a named entity

(4) Semantic Aggregation

After that, we aggregate the geo-centric entity information derived from the previous step. Depending on the chosen exploitation strategy (DFS or BFS) we obtain a set of associated `countries` associated with each article. Since, there are (usually) multiple relations associated with each named entity, this might lead to one (in the case of BFS) or - potentially - many (in the case of DFS) associated countries per entity.

(5) Countries Prediction

In the final step, we provide a Web interface to assess and visualize the news articles. To this end, we utilize the extracted geo-information in order to rank and present the articles based on their correlation to specific countries <u>or</u> allow a country-based exploration of the most relevant articles.

3   Demonstration

ELEVATE-live facilitates the user to assess and visualize the virality of news articles (cf. https://elevate.greyc.fr). In the following, we describe the two main use-cases of our system: event assessment and exploration.
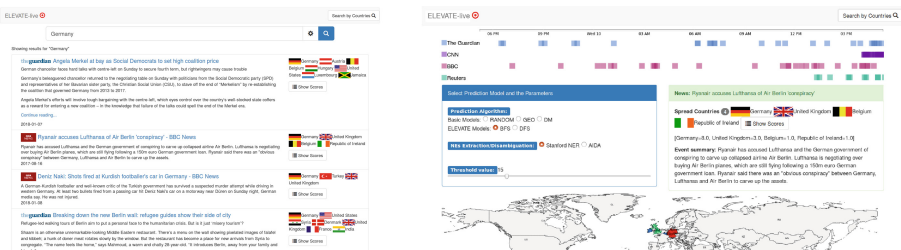

Fig. 2. Country-specific news assessment [left] and exploration [right] (Color figure online)

Assessing Viral News Stories by Country

In our first use-case, news contents can be searched by their relevance for a country based on the named entities contained in the article. Further options allow the user, e.g., to investigate the underlying models (BFS and DFS). In addition, temporal constraints can be defined in order to focus the query onto a certain

time-interval. The document ordering is then done based on the aggregated score derived from the semantically enriched documents (cf. Fig. 2 [left]).

Assessing Virality from semantically enriched Web Contents
In our second use-case, the most recent news articles from these news sources are mapped on a zoomable timeline. Each news article is represented by a colored square. This news timeline allows the user to navigate through the news stories along the temporal dimension by summarizing the story in focus. The user can further explore the countries in which a story has the potential to become viral based on the entities contained. As before, the user may also explore the differences of BFS- and DFS-based semantic enrichment. The results in terms of countries "affected by the virus" are highlighted on an interactive world map, with a color coding from blue to red representing the degree of virality (cf. Fig. 2 [right]).

4   Related Work

There are only a few related works addressing the aspect news virality in association with the countries involved. A study on the virality of tweets has been investigated by Hansen et al. without considering the aspect of the named entities involved [3]. STICS, on the contrary, provides a search engine which employs entity-level analytics to search documents, without providing country-specific analytics [4]. Jenders et al. introduce an approach in order to discover viral tweets, again without the notion of country-specific aspects [7]. Thus, ELEVATE-live is unique in interlinking news articles and associated countries via a geo-temporal enabled search interface.

5   Conclusions and Outlook

We presented a novel Web-based tool that exploits entity information from online news in order to assess and visualize its virality. The originality of our approach stems from the analytics on the entity-level. In future work, we intend to pursue our studies on arbitrary Web contents and to investigate recurring patterns.

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., et al. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_52
2. Govind, Spaniol, M.: ELEVATE: a framework for entity-level event diffusion prediction into foreign language communities. In: Proceedings of the 9th International ACM Web Science Conference (WebSci 2017), pp. 111–120 (2017)
3. Hansen, L.K., Arvidsson, A., Nielsen, F.A., Colleoni, E., Etter, M.: Good friends, bad news - affect and virality in twitter. In: Park, J.J., Yang, L.T., Lee, C. (eds.) FutureTech 2011. CCIS, vol. 185, pp. 34–43. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22309-9_5

4. Hoffart, J., Milchevski, D., Weikum, G.: STICS: searching with strings, things, and cats. In: Proceedings of the 37th International ACM SIGIR Conference on Research & #38; Development in Information Retrieval, SIGIR 2014, pp. 1247–1248. ACM, New York (2014)
5. Hoffart, J., Suchanek, F.M., Berberich, K., Weikum, G.: YAGO2: a spatially and temporally enhanced knowledge base from wikipedia. Artif. Intell. **194**, 28–61 (2013)
6. Hoffart, J., Yosef, M.A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., Weikum, G.: Robust disambiguation of named entities in text. In: Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, pp. 782–792 (2011)
7. Jenders, M., Kasneci, G., Naumann, F.: Analyzing and predicting viral tweets. In: Proceedings of the 22nd International Conference on World Wide Web, WWW 2013 Companion, pp. 657–664. ACM, New York (2013)
8. Suchanek, F.M., Kasneci, G., Weikum, G.: YAGO: A core of semantic knowledge - unifying wordnet and wikipedia. In: 16th International World Wide Web Conference (WWW 2007), pp. 697–706. ACM (2007)