



Figure 1: *a)* Graph View of the *named entity* based clustering operation on the collected items. *b)* Visual summary of the cluster “Beyonce”: the introduction is fetched through disambiguating the selected topic on DBpedia and it is followed by the storyfication of microposts and media items.

3. MEDIA FINDER

MediaFinder aims to automatically generate visual summaries using the results of MediaServer. This framework is composed of the following components:

- Near de-duplication detection: The same media item can occur in multiple microposts, typically when a re-tweet or re-share operations is performed. State of the art techniques in content-based image retrieval are used to detect near-duplicate images.
- Named Entity Extraction: For each micropost, we extract named entities using the NERD framework [4]. A multi-lingual entity extraction is performed and the entities extracted are typed using the NERD Ontology⁶.
- Topic Generation: We identify topics using four clustering operations using the textual features of the microposts (LDA, named entities, etc.).

Once the clustering operations are completed, the entity which best represents a set is selected. We call this process *topic generation*. For each cluster, a Bag of Entities (BOE) is computed and the most representative entity is selected to be the cluster topic. We disambiguate this topic using a DBpedia URI⁷. Consequently, the output of this component is a set of clusters (limited to ten for visualization optimization) that corresponds to topics extracted from the data collected.

The visualization emphasizes the different aspects of storyboards. The graph view shows the relationships between microposts and topics, 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

⁶<http://nerd.eurecom.fr/ontology/nerd-v0.5.n3>

⁷<http://dbpedia.org>

easy sharing possible. Figure 1 illustrates MediaFinder with the 2013 Superbowl event. One of the cluster (labeled as “Beyonce”) brings the attention of the user. By simply clicking on the cluster, an illustrative storyboard is depicted mixing microposts and social media shared by the crowd.

Acknowledgments

This work was partially supported by the European Union’s 7th Framework Programme via the projects LinkedTV (GA 287911).

4. REFERENCES

- [1] A. Joly, J. Champ, P. Letessier, N. Hervé, O. Buisson, and M. Viaud. Visual-Based Transmedia Events Detection. In *20th ACM international conference on Multimedia (MM’12)*, Nara, Japan, 2012.
- [2] V. Milicic, J. L. Redondo García, G. Rizzo, and R. Troncy. Tracking and Analyzing The 2013 Italian Election. In *10th Extended Semantic Web Conference (ESWC’13), Demo Session*, Montpellier, France, 2013.
- [3] G. Rizzo, T. Steiner, R. Troncy, R. Verborgh, J. L. Redondo García, and R. V. de Walle. What Fresh Media Are You Looking For? Retrieving Media Items from Multiple Social Networks. In *International Workshop on Socially-aware multimedia (SAM’12)*, Nara, Japan, 2012.
- [4] G. Rizzo and R. Troncy. NERD: A Framework for Unifying Named Entity Recognition and Disambiguation Extraction Tools. In *13th Conference of the European Chapter of the Association for Computational Linguistics (EACL’12)*, Avignon, France, 2012.
- [5] T. Steiner. A Meteoroid on Steroids: Ranking Media Items Stemming from Multiple Social Networks. In *22nd World Wide Web Conference (WWW’13), Demo Session*, Rio de Janeiro, Brazil, 2013.