Augmenting TV Newscasts via Entity Expansion

José Luis Redondo García[2], Michiel Hildebrand[1], Lilia Perez Romero[1], Raphaël Troncy[2]

[1] CWI, Amsterdam, The Netherlands,
{M.Hildebrand, L.Perez}@cwi.nl
[2] EURECOM, Sophia Antipolis, France,
{redondo, raphael.troncy}@eurecom.fr

Abstract. We present an approach that leverages on the knowledge present on the Web for identifying and enriching relevant items inside a News video and displaying them in a timely and user friendly fashion. This second screen prototype (i) collects and offers information about persons, locations, organizations and concepts occurring in the newscast, and (ii) combines them for enriching the underlying story along five main dimensions: expert's opinions, timeline, in depth, in other sources, and geo-localized comments from other viewers. Starting from preliminary insights coming from the named entities spotted on the subtitles, we expand this initial context to a broader event representation by relying in the knowledge of other Web documents talking about the same fact. An online demo of the proposed solution is available at http://linkedtv.project.cwi.nl/news/

Keywords: Video Annotation, Entity Expansion, News Enrichment

1 Introduction

Second screen applications are a popular approach to enrich the TV viewing experience. Within the LinkedTV project we developed a second screen application for News broadcasts. The functionality and the design of this application is the result of a user-centred design process involving focus groups, interviews, iterative design and evaluation [2]. The initial design and evaluation of the prototype was done with manually curated content. This was deliberately chosen, as we did not want the quality of the content impact the evaluation of the design. However in practice, the manual curation of the content is not feasible, as it is too time consuming for a broadcaster.

This paper provides a solution to automatically generate the content for the LinkedTV News second screen application. The application supports two modes of interaction. The passive mode gives the user access to factual information about persons, locations and organisations that are related to a news item, as shown in the screenshot on the left side of Figure 1. To generate the content for this passive mode we therefore need a logic able to identify entities that could be highly relevant to the user. A typical approach is to use Named Entity Recognition (NER) over the textual information attached to particular video fragment.

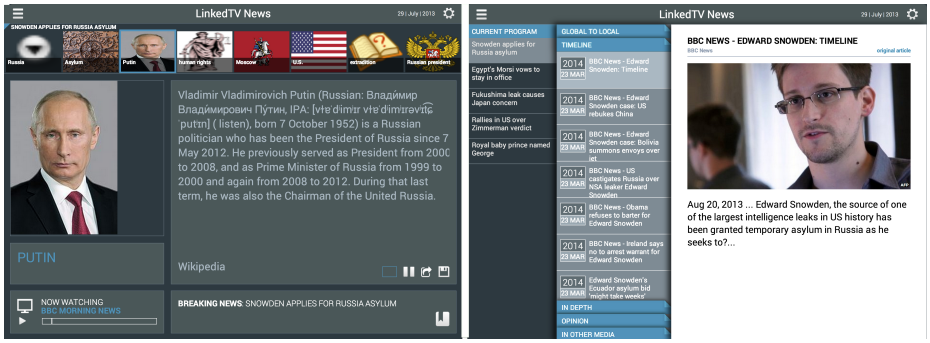

Fig. 1: Demo screen captures corresponding to the (1) passive mode and the (2) active mode. Access the demo at `http://linkedtv.project.cwi.nl/news/` for more details.

By linking the entities to real world objects using web identifiers (Named Entity Disambiguation) we get access to the factual information about these entities. A growing number of APIs provide such a service, like AlchemyAPI[3] or DBpedia Spotlight[4]. In the manual curation of the content we, however, experienced that relevant entities do not only occur in the subtitles, but can also be related to the video indirectly. Therefore, we propose an extension of the traditional entity extraction method on the subtitles that studies additional Web documents related to the main video. In particular, our solution uses a representative set of concepts found in the subtitles to retrieve extra Web documents that are annotated as well, in order to extend and better rank the initial set of entities and propose the main N candidates to be displayed in the interface.

The active mode of the LinkedTV news second screen application allows the user to explore background and related information to a news item, as shown on the right side of Figure 2. The interface provides 5 browsing dimensions that will be populated by relying on a Google Custom Search Engine[5] and a method to generate the queries that are tailored to specific requirements of the content in the different dimensions.

2 LinkedTV News companion application

In the passive mode the LinkedTV news application operates as a second screen that is synced with the TV program. This mode supports the user with looking up factual information about the entities that occur in the news. The screenshot on the left side of Figure 1 shows entities per news item and is refreshed when the next news item start. The interface contains three parts. At the top it contains a carousel of the entities related to the news item. The middle of the interface shows the entity slide with information about the active entity, taken from DBPedia.

[3] `http://www.alchemyapi.com/`
[4] `http://spotlight.dbpedia.org/`
[5] `https://www.google.com/cse/`

The bottom part of the interface provides information about the current news item, controls for the video and a button to bookmark the current news item.

The bookmark button in the passive mode provides a simple form of interaction that allows the user to store news item to watch later, when passing to the active mode and exploring the bookmarked news items. The screenshot on the right side of Figure 1 shows the interface of the active mode. The left column of the interface shows a list of bookmarked items, the item on Edward Snowden is selected. The left column shows the article that is currently selected. The middle column shows the five browsing dimensions based on information needs from end users: the timeline gives an overview of past events, the opinion dimension gives access to articles in which journalists express their opinion about the topic, in other sources different perspectives on the news can be found as expressed by different news providers, in depth dimensions allows exploration of more detailed background, and the geo-localized dimension shows how people in different places in the world respond to the news item.

2.1 Entity extraction and expansion for the Passive Mode

For the passive mode we need to find the elements of the news, for example, basic information about the location or people involved. The logic that feeds the information displayed in this passive mode is crucial not only because it complements what the user is watching in every moment, but also because it generates the concepts that launch further enrichments in the active mode. To reconstruct the semantic context associated with one particular news video, we perform named-entity recognition over the corresponding subtitles using the NERD framework [3]. The output of this phase is a collection of entities annotated using the NERD Ontology[6]. This set includes a list of ranked entities that are explicitly mentioned during the video.

The set of entities obtained from a traditional named entity extraction operation is normally insufficient and incomplete for expressing the context of a news event. Sometimes the relevant concepts are simply not mentioned in the transcripts while being crucial for understanding the story. We perform then a process named *entity expansion*, which relies on the idea of retrieving and analyzing additional documents from the Web where the same event is also described. By increasing the size of set of documents to analyse, we increase the completeness of the context and the representativeness of the list of entities, reinforcing relevant entities and finding new ones that are potentially interesting regarding that news item. The entire logic is illustrated in in Figure 2.

Query Generation The *Five W's* is a popular concept of information gathering in journalistic reporting. It captures the main aspects of a story: who, when, what, where, and why [1]. We try to represent the news item in terms of four of those five W's in order to generate a query that retrieves documents associated to the same event. In order to achieve this, the original entities are mapped to the NERD Core ontology. From those ten different categories, we generalize to

[6] `http://nerd.eurecom.fr/ontology/nerd-v0.5.n3`

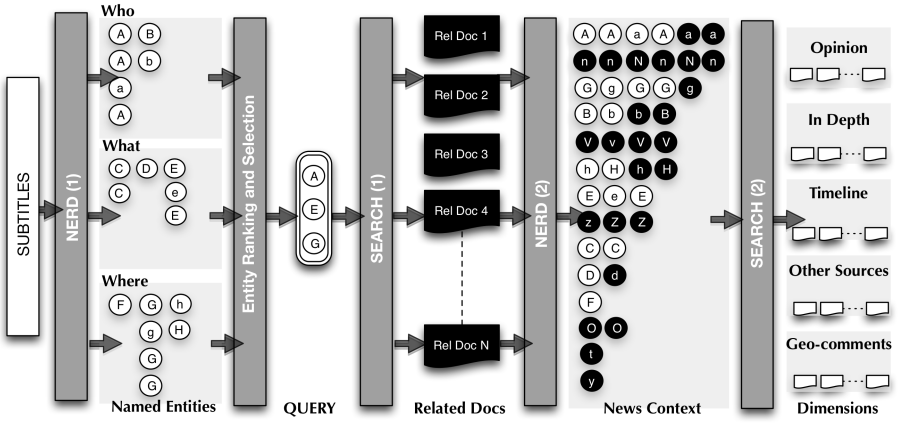

Fig. 2: Schema of Named Entity Expansion Algorithm.

three classes: the Who from `nerd:Person` and `nerd:Organization`, the Where from `nerd:Location`, and the What from the rest of NERD types after discarding `nerd:Time` and `nerd:Amount`. The When or so-called temporal dimension does not need to be computed since it is considered to be provided by the video publisher. The final query is the result of concatenating the labels of the most relevant entities in the sets Who, What, Where in that particular order, for a given time period t. This query will be injected into a document search engine (in our case, the Google) where additional descriptions about the news event can be found.

Entity Clustering In this phase, the additional documents which have just been retrieved are processed and analyzed in order to extend and re-rank the original set of entities and consequently get a better insight about the event. They are again analyzed by the NERD framework in order to extract more named entities. Once finished, we performed a centroid-based clustering operation over the entities retrieved, considering as centroid the entity with the most frequent disambiguation URL's and most repeated label. The output of this phase is a list of clusters containing different instances of the same concept.

Entity Ranking The final step of the expansion consists of ranking the different named entities obtained so far according to its relative frequency in the transcripts of the event video, relative frequency over the additional document; and average relevance according to the named entity extractors. The final output of the entity expansion operation is a list of entities together with their ranking score and the frequency in both the main video and in the collected documents retrieved from the search engine.

2.2 Related content for the Active Mode

The active mode of the application acts as a hub where the viewers can access extra documents for complementing what is being told in the main news video. A

similar logic to the one explained in Section 2.1 is applied over the main entities coming from the expansion process for building custom queries in Google CSE, but relaxing or emphasizing some particular *W's* and operating over particular lists of Web resources. For the **Timeline** we rely on a query created by including only the most relevant entity inside the pattern "The" + entity + "case", without any time constraint. For example, in news item about the recent polemic of Edward Snowden we would have the text "The Edward Snowden case". For **In other sources**, we launch a query generated from the set of expanded entities over a curated list of resources including mainly journals and broadcasters Web sites. In our example the query looked like "Asylum Snowden Russia". The **Opinion** dimension includes a list of documents obtained by executing the same query in the previous dimension, but operating over a different list of curated sources that considers only subdomains specialized in opinion documents. The **Geo-localized** data is obtained by launching the same query into the Twitter API, filtering by location and reducing the temporal dimension t to the last 7 days. And the **In depth** section gathers documents obtained by attaching to the label of the most relevant entity from the entity expansion results the keyword "in depth", and removing any temporal restriction or Web source limitation. In our example the query would be"Edward Snowden in depth".

3 Discussion

The preliminary results indicate that we are able to offer to the viewer a relevant set of entities that expands the initial concepts detected by traditional named entity recognition approaches. Many details not explicitly mentioned in the video but present in the context of the newscasts are now available to the viewer to be consumed, either while he is watching the news in passive mode or he browses additional insights in the active mode. We are currently working on formally evaluating the entity expansion method and relevance ranking algorithms.

Acknowledgments

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

References

1. L.-J. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recognition. In *IEEE 11th International Conference on Computer Vision (ICCV'07)*, pages 1–8, 2007.
2. L. Perez Romero, R. Ahn, and L. Hardman. Linkedtv news: designing a second screen companion for web-enriched news broadcasts. Technical report, Technical Report, Eindhoven University of Technology, 2013.
3. G. Rizzo and R. Troncy. NERD: A Framework for Unifying Named Entity Recognition and Disambiguation Extraction Tools. In *13th European Chapter of the Association for Computational Linguistics (EACL'12)*, Avignon, France, 2012.