

NewsStand: An interactive news organizer

Group 7: Hossein Esfandiari¹, Nikolaos Kofinas, Mahyar Najibi², and Hong Wei
{hossein, nkofinas, najibi, hyw} @cs.umd.edu

February 26, 2015

1. Introduction

Nowadays, there is a great amount of information available and generally it is difficult to locate a website that contains local news from a city, or even a small village that you are interested in. As an example, suppose that someone is going to move to a new area and he is interested in reading the local news to find out if it is a good neighborhood. An effective way is to search for news websites that are localized for the city of interest and look for articles that are focused on the city.



Figure 1: NewsStand interactive map.

2. Problem Definition

This paper proposes a new, easy and user-friendly way to search for news based on areas. Newsstand's goal is to create a map, in which the user can zoom in and out, and pin various

¹www.cs.umd.edu/~hossein

² www.umiacs.umd.edu/~najibi

news into the focused areas of the map [1]. The idea behind this approach is that the users will have an easy way to navigate through the map and find news for specific areas.

3. Challenges

There are several challenges towards designing a news analyzer like NewsStand. One of the most important challenges is scalability. News articles are being produced at a fast rate and NewsStand should be able to process all these new articles.

Moreover, determining the geographical information of news documents is another important challenge. To deal with that, one needs to extract words which refer to geographical locations [4]. However, there are many words which refer to both geographical and non-geographical entities, words which refer to more than one location, and words which all refer to a single location. Resolving these ambiguities needs a sophisticated inference approach.

4. General Approach

Generally, NewsStand monitors RSS feeds from online news sources and retrieves articles within minutes of publication. It then extracts geographic content from articles using a custom-built geotagger and groups articles into story clusters using a fast online clustering algorithm.

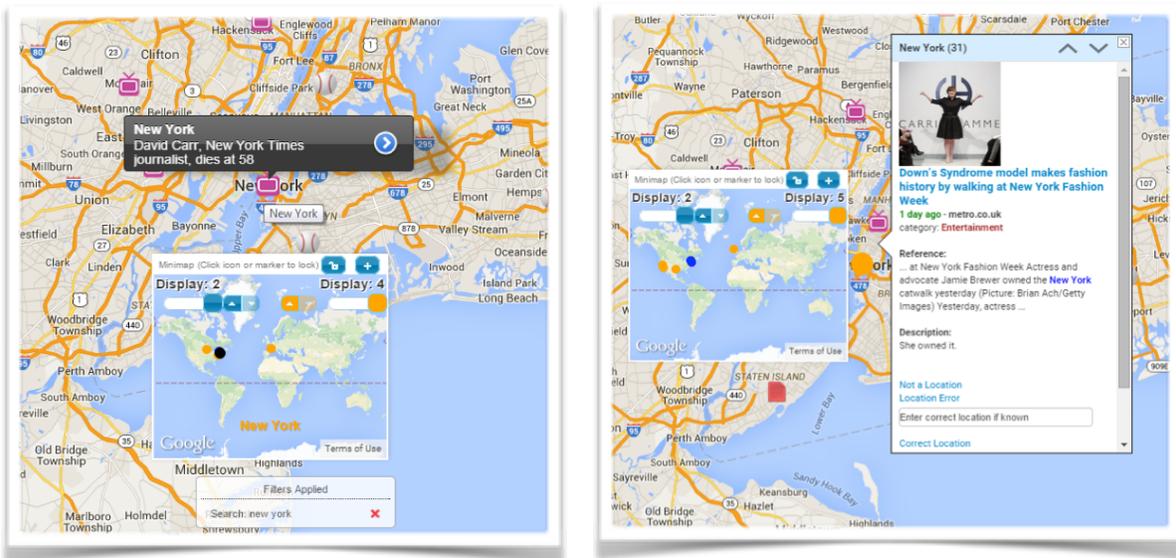
- A. **News Crawling:** NewsStand uses RSS feeds as its source of data. It exploits a simple intuitive way to extract story content by retrieving the largest sections found containing no markup tags.
- B. **Geotagging:** NewsStand's geotagging module includes four stages: **1) Entity Feature Vector Extraction:** In this phase, the authors use Name-Entity Recognition from Natural Language Processing (NLP) to extract phrases that are most likely to be geographic locations called Entity Feature Vector (EFV); **2) Gazetteer Record Assignment:** After extracting the EFVs, NewsStand uses a gazetteer GeoName to find geographic features, which appear in both gazetteer and EFVs. **3) Geographic Name Disambiguation:** NewsStand then associates each geographic feature with several matching locations from gazetteer and uses multiple heuristic filters such as geographic distance to resolve ambiguities. **4) Geographic Focus Determination:** The geotagger

then ranks these locations (based on such as frequency) to distinguish the importance of geotagged locations to form geographic focus (main location).

- C. **Online Clustering:** All news articles describing the same news event are then grouped together to form a story cluster. Two features are involved: time and Term Frequency-Inverse Document Frequency (TF-IDF).
- D. **Cluster Focus:** At last, the paper computes focus (main locations) for each cluster and display them on the map associating with corresponding lat/long coordinates.

5. User Interface

[NewsStand](#) [4] is a web application that you can use on your laptop, tablet and smartphone [5]. In fact, the most important difference between NewsStand and other news services is that in NewsStand, geographical location of the news is highlighted. At the first look at NewsStand web service, you see some news distributed on the map (see Figure 1). As you zoom in on some specific location, you see more news about that location (see Figure 2, part A). By zooming out, less important news disappears and you only see important news. Moreover, you may chose to see classified news, related to business, science and technology, entertainment, health and sports. Interestingly, you can, also, search for specific news and see, locations related to that news (see Figure 2, part B). You can find an interesting introduction video by the designers here: <http://vimeo.com/106352925>.



A) Looking at the title of a news in New York. B) Looking at a specific news and its locations.

Figure 2: Looking at local news in NewsStand website.

References:

- [1] Samet, Hanan, Jagan Sankaranarayanan, Michael D. Lieberman, Marco D. Adelfio, Brendan C. Fruin, Jack M. Lotkowski, Daniele Panozzo, Jon Sperling, and Benjamin E. Teitler. "Reading news with maps by exploiting spatial synonyms." *Communications of the ACM* 57, no. 10 (2014): 64-77.
- [2] Teitler, Benjamin E., Michael D. Lieberman, Daniele Panozzo, Jagan Sankaranarayanan, Hanan Samet, and Jon Sperling. "NewsStand: A new view on news." In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, p. 18. ACM, 2008.
- [3] Lieberman, Michael D., Hanan Samet, and Jagan Sankaranarayanan. "Geotagging with local lexicons to build indexes for textually-specified spatial data." *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*. IEEE, 2010.
- [4] Samet, Hanan. "UMD NewsStand." UMD NewsStand. Web. 14 Feb. 2015. <<http://newsstand.umiacs.umd.edu/web/>>.
- [5] Samet, Hanan, Benjamin E. Teitler, Marco D. Adelfio, and Michael D. Lieberman. "Adapting a map query interface for a gesturing touch screen interface." In *Proceedings of the 20th international conference companion on World wide web*, pp. 257-260. ACM, 2011.