* Problem (1)

1. For text analysis methods:
   1. For temporal trends:
      1. Assumption: “Create\_at” contains the date and time stamp of the tweet post.
      2. Classify the tweets based on the [created\_at] and be able to filter the tweets that are posted for the last minute, hour, day, or year.
   2. For spatial distribution:
      1. Classify the tweets based on the [geo-latitude, geo-longitude] and grouping them according to continents, countries, cities, and landmarks. [reverse geocoding]
   3. For Semantic classification:
      1. Start by topic based text summarization methods such as “Topic Modeling” [2] and employ the “Topic Rose Tree” [1] to organize topics into a hierarchical structure in which similar topics are in the same branch
2. Interface Sketch: (visualization designs, interactive features)

**Filter by:**

Cities

Countries

continents

**Filter by:**

Last hour

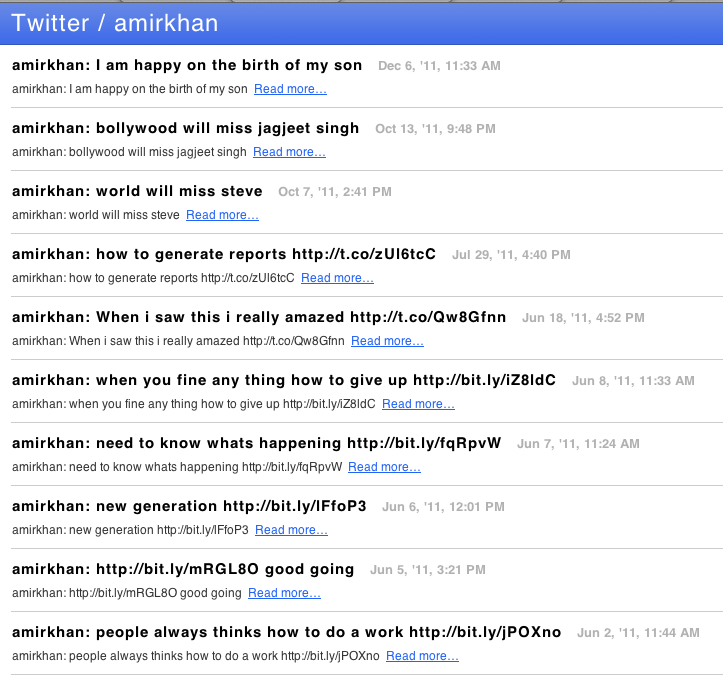
Last day

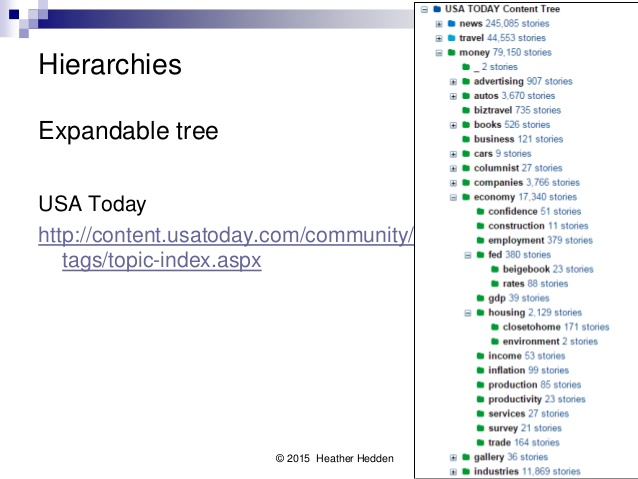
Last year

Filter

Filter

Filter





Topics:

TimeLine:



Load From:

* Load From Button: to load the dataset
* Topics Expandable List: list all the topics that was discovered from the tweets’ text, and once press on one of the sub-topic in the list, the list of related tweets are displayed
* Map window
* TimeLine window: present a histogram (number of tweets are posted according to time stamp)
* Example of scenario:
  + Load dataset of tweet messages
  + Press on sub-topic “actors”
  + Then, a list of tweets related to this topic is listed on a small window near the topics, also pin points are placed on top of the map to display the tweets –related to this sub-topic – locations
  + Also, a histogram of number of tweets related to this sub-topic are presented on the timeline through the last month for example
  + User can draw a boundary on the map and then press filter button
  + User can filter the outputs based on :
    - For Map: based on cities, countries, or continents.
    - For timeline: based on last hour, day, or year.

1. System architecture design:



the system starts by reading the raw data (tweets data) then analyzing them by filtering them with Topics. Save the filtered data in the intermediate database. When the interface is running, it retrieves the data from the database, also, it displays the histogram of the tweets according to the time stamps in the timeline. When the user wants to retrieve the data within a boundary on the map, it will filter the data according to the geometry attributes and show them on the map

References:

[1]

W.Dou, L.Yu, X.Wang, Z.Ma, and W. Ribarsky. Hierarchical topics: Visually exploring large text collections using topic hierarchies. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2002–2011, Dec 2013.

[2]

X. Wang, W. Dou, Z. Ma, J. Villalobos, Y. Chen, T. Kraft, and W. Ribarsky. I-si: Scalable architecture for analyzing latent topical-level information from social media data. In *Computer Graphics Forum*, volume 31, pages 1275–1284. Wiley Online Library, 2012.