NRV Tweets

- Midterm Presentation
- VT CS4624, Blacksburg, VA
- Sponsors: Mohamed Magdy, Dr. Andrea Kavanaugh, Ji Wang
- Ben Roble, Justin Cheng, Marwan Sbitani
Virtual Town Square data

- Tweets
- Facebook Groups
- Websites
- RSS Feeds
Events and tags

Data from Tweets/FB/RSS within the NRV

- Scanned for events
- Tweets/posts/messages are grouped together if they’re related
- Tagged with keywords
Categorizing Data

Tagging
- Machine learning algorithms
- etcML
- Latent Dirichlet Allocation (LDA)

Searching
- Solr
Example Data Schema

- **topics**
  - id INT(11)
  - title VARCHAR(255)
  - description TEXT
  - created_at DATETIME
  - updated_at DATETIME
  - user_id INT(11)
  - keywords VARCHAR(255)
  - day DATETIME

- **topic_keywords**
  - id INT(11)
  - name VARCHAR(255)
  - distribution DECIMAL(10,0)
  - created_at DATETIME
  - updated_at DATETIME
  - topic_id INT(11)

- **stories**
  - id INT(11)
  - title VARCHAR(255)
  - description TEXT
  - url VARCHAR(255)
  - views INT(11)
  - created_at DATETIME
  - updated_at DATETIME
  - user_id INT(11)
  - topic_id INT(11)
  - kind INT(11)

- **tags**
  - id INT(11)
  - name VARCHAR(255)
Data Example

```json
{
    'id' : 4898,
    'title' : 'University Distinguished Awards.....',
    'description' : 'BLACKSBURG May 10....'
    'views' : 76,
    'updated_at' : '2014-02-09 10:10:53',
    'topic_id' : 3729,
    ...
}
```
Twitter query: "IHOP"

**Top predictions: negative**

- **96%** RT @antvilleda95: "@PaigeOlvera: So the wait for IHop is 15 minutes, but chances are it will die down by the time we get there xD!"ight than

- **94%** "@PaigeOlvera: So the wait for IHop is 15 minutes, but chances are it will die down by the time we get there xDight than

- **92%** Working on free pancake day at IHop is the worst

**Top predictions: positive**

- **97%** RT @TheRebusAgency: @HOUcheap @houstonsocial @IHOP Thanks for posting! We shared this on our FB page! Have a great week.

- **96%** @IHOP I love you and thank you for free pancakes once a year

- **95%** @IHOP HOORAY! Can't Wait :)
Solr

- Powerful search engine platform
- Allows queries of large amount of Media data that we’ve collected.
Id: 438370322849345536
Tweet: Pix of my garden, after winter-storm "Pax" left his calling card.
url: http://t.co/jG4QP6hZG
Hashtag: Winter Storm
Type: post
Date: 2014-02-25T17:50:04Z

Id: 438418813462523904
Tweet: #Augusta cleanup from Winter Storm Pax to cost more than $8 million
url: http://t.co/bd00ACN32P
Hashtag: Winter Storm
Type: post
Date: 2014-02-25T17:50:04Z

Id: 438334465358561283
Tweet: Volunteering for Winter Storm PAX clean-up information is now available online including a downloadable volunteer sheet. More information...
Hashtag: Winter Storm
Type: post
Date: 2014-02-25T12:33:04Z

Id: 438328462177497088
Tweet: #NYC Winter Storm Pax Northeast, Middle Atlantic Forecast: Washington, Philadelphia ... - The Weather ... #NYCJobs
url: http://t.co/allW5TBv
Hashtag: Winter Storm
Type: post
Date: 2014-02-22T17:50:04Z

Id: 438290609074626560
Tweet: RT @heretohelp: We are monitoring Winter Storm #Pax Call 1-800-243-5860 24/7 if you need to file a claim. Tips from @RedCross http://t.co/c...
Data Aggregation

- Database tables
- Hadoop (HDFS)
- Virtual Town Square (VTS)
- Trend Timeline from Tweets
Today’s
This Week’s
Collected From
All Stories

Ben Norris to travel around the world in pursuit of 2013 Special Olympics World Games gold medal
BLACKSBURG Va. Jan. 25 2013 – Just as students and instructors return to campus for the start of the...
Current progress

- Data access progress
- Solr progress (news stories)
- Tweet tagging progress
Future progress & Deliverables

- Integration with VTS
- News searchable
- Tweets tagged, searchable
- Facebook info tagged, searchable
Possible uses

- Single source for local news
- Historical reference of events
- Tracking of event social impact
- Researching of trends
- Forecasting of event responses
Timeline

- Feb 6 - Initial Meeting
- Feb 19 - Raw tweet data accessed, Solr tool reviewed
- March 5 - Initial research completed on associating news stories with topics
- March 19 - Initial research completed on associating Tweets with topics
- April 9 - Best topical modeling tool selected
- April 23 - Tweets and news stories associated with topics and searchable with Solr
- May 7 - Final report delivered