Global Event Detector

Final Project Presentation
Multimedia, Hypertext, and Information Access
4/27/2017
Blacksburg, VA 24061
Emma Manchester, Alec Masterson,
Ravi Srinivasan, Harrison Grinnan,
Sean Patrick Crenshaw
Project Systems

Data Acquisition
- Polling
- Databases
- Parsing

Data Analysis

Data Visualization
Polling

Collects data from Reddit WorldNews Subreddit

Driver script runs every 12 hours

Processes news articles

Stores articles in database
Databases

Stores data into raw, processed, cluster, and clusterPast tables

Retrieving articles more than once

Data used for visualizations
<table>
<thead>
<tr>
<th>rawld</th>
<th>url</th>
<th>title</th>
<th>content</th>
<th>datePosted</th>
</tr>
</thead>
<tbody>
<tr>
<td>63hj1t</td>
<td><a href="http://www.reuters.com/article/north-korea-">http://www.reuters.com/article/north-korea-</a></td>
<td>North Korea fires projectile into sea of Japan</td>
<td>SEOUL North Korea test-fire...</td>
<td>2017-04-04 18:27:05</td>
</tr>
<tr>
<td>63f13y</td>
<td><a href="http://www.independent.co.uk/">http://www.independent.co.uk/</a></td>
<td>Donald Trump administration blames North Korea for...</td>
<td>US President says the attack...</td>
<td>2017-04-04 12:05:27</td>
</tr>
<tr>
<td>63cpx1</td>
<td><a href="http://www.bbc.com/news/world/">http://www.bbc.com/news/world/</a></td>
<td>At least 18 people killed in suspected...</td>
<td>Share this with EmailFacebook...</td>
<td>2017-04-04 03:33:59</td>
</tr>
<tr>
<td>63e5a5</td>
<td><a href="http://www.dw.com/en/eu-to-cut-gas-dependency">http://www.dw.com/en/eu-to-cut-gas-dependency</a></td>
<td>EU to cut gas dependency on Russia</td>
<td>Israel and several EU nations...</td>
<td>2017-04-04 09:38:06</td>
</tr>
<tr>
<td>63cz8z</td>
<td><a href="http://www.independent.co.uk/">http://www.independent.co.uk/</a></td>
<td>Junior doctor whose Facebook post o...</td>
<td>Police say Rebecca Ovenden...</td>
<td>2017-04-04 04:57:16</td>
</tr>
<tr>
<td>63cz5i</td>
<td><a href="https://www.yahoo.com/new/">https://www.yahoo.com/new/</a></td>
<td>Hero dog dies tackling female suicide...</td>
<td>A dog has died after attacking...</td>
<td>2017-04-04 04:56:25</td>
</tr>
</tbody>
</table>
Parsing - BeautifulSoup

['Four', 'More', 'Secret', 'Jails', 'Illegally
being', 'illegally', 'detained', 'in', 'at',
the', 'Novaya', 'Gazeta', 'newspaper', 'has',
that hundreds', 'of', 'men', 'were', 'being',
agovernment-backed', 'crackdown', 'on', 'the',
jails', 'in', 'theChechen', 'villages', 'of',
claimed', 'that', 'atleast', 'four', 'more',
theirsexual', 'orientation. The', 'men', ',',
', 'are', 'only', 'released', 'after', 'their',
the', 'outlet', 'wrote. The', 'newspaper', 'also',
about', 'the', 'crackdown', ',', 'despite', 'The',
exist. In a meeting, with Putin,
Parsing - Removing Stopwords

Parsing - Filtering Content

Parsing - Stemming Content

Project Systems

Data Acquisition → Data Analysis (Seed Clustering) → Data Visualization
Stanford Named Entity Recognizer

3 Class Models used

- Location
- Person
- Organization

SNER demo
Seed Extraction

Seed - a series of entities
Attached to article object
Inserted into Database
North Korea says ready to strike U.S. aircraft carrier
North Korea 'will test missiles weekly,' senior official tells BBC
North Korea: 'US has now gone seriously mad' -- US strike group to
the North is ramping up for a sixth nuclear test.
US military considers shooting down North Korea missile tests
Putin sends troops to Russia's border with North Korea
Chances of imminent war with North Korea 'wildly overblown,' U.S. e
North Korea warns Australia of 'blindly toeing US line', warns of r
Kim Jong-un is starting to get 'very paranoid', UN ambassador warns
South Korea Tells Trump It's Actually Never Been a Part of China
North Korea warns it will 'wipe America off face of the Earth' after
Data Visualization

- A PHP-based single-page Web-Application
- Used to display Cluster Data
Global Event Crawler and Seed Generator for GETAR
Article Carousel

For Passive Viewing

Sources: US prepares charges against WikiLeaks' Assange

2017-04-20 18:00:24

Seeds: US, WikiLeaks, Evon Perez, Pamela Brown, Shimon

WikiLeaks releases more top-secret CIA docs as U.S. considers charges: WikiLeaks claims the 31-page user guide for a CIA device code-named Weeping Angel can turn some Samsung TVs into surveillance t

Venezuela crisis: Teenager and woman shot dead at anti-government protests

Seeds: Venezuela, Venezuela Nicolas, Caracas, San Cristobal, Colombian

Venezuelan president Nicolas Maduro donates 500,000 to Trump fund despite economic woes

Seeds: Venezuelan, Nicolas Maduro, Trump, Venezuela, Venezuela Donald Trump
Deliverables

Effective Clustering of Articles from Reddit

Storage of Entities in Database

Website

Visualization of Clusters

Article Carousel
Lessons Learned

- Ask for help early
- Project specifications change over time
- Most things have a built-in Python library - save time!
- Frequent, regularly scheduled meetings keep things moving
Questions?
References

References