Team FE
Final Presentation

Virginia Tech, Blacksburg, VA, 24060

CS 5604
Fall 2017
12/12/2017

Haitao Wang
Yali Bian
Shuo Niu
Jieun Chon
Table of contents

1. GETAR Portal Front-end
   • Project Workflow
   • Frontend Modules
   • Core Metadata
   • Schema
   • Functionality
   • Future Work

2. GETAR Visualization
   • Functions
   • Framework
   • Search
   • TweetList
   • SNS
   • User Information
   • Timeline
   • Views
GETAR Portal
Front-end Project Workflow

(w/ GeoBlacklight)
GETAR Portal - Frontend Modules

WEB PORTAL

(GEO) BLACKLIGHT SCHEMA

DATA QUERY

SOLR REST API

Text Search Module
Map View Module
Visualization Module
User Management Module

Solr

MySQL/PostGIS SQL
## GETAR Portal - Core Metadata

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document ID</td>
<td>id</td>
<td>Unique document identifier</td>
<td>899664353946996738</td>
</tr>
<tr>
<td>Author</td>
<td>dc_creator_sm</td>
<td>Document author</td>
<td>jhdee</td>
</tr>
<tr>
<td>Year</td>
<td>solr_year_i</td>
<td>Year when publishing the document</td>
<td>2017</td>
</tr>
<tr>
<td>Spatial coverage</td>
<td>dct.spatial_sm</td>
<td>Interpreted location as place name</td>
<td>Radford, virginia</td>
</tr>
<tr>
<td>Institution</td>
<td>dct_provenance_s</td>
<td>Institution of data collection holder</td>
<td>Virginia Tech</td>
</tr>
<tr>
<td>Spatial bounds</td>
<td>dct Bounds_s</td>
<td>Bounding box for S,W,N,E of interpreted location</td>
<td>ENVELOPE (34.0, -118.2, 36.5, -117.4)</td>
</tr>
<tr>
<td>URL</td>
<td>dct_url_s</td>
<td>URL of orginal documents</td>
<td><a href="http://xxxx.yy.zz">http://xxxx.yy.zz</a></td>
</tr>
<tr>
<td>Full text</td>
<td>dc_description_s</td>
<td>Full text of tweet / webpage</td>
<td>Full text</td>
</tr>
<tr>
<td>Topics</td>
<td>dc_topics_s</td>
<td>Related topic name</td>
<td>&quot;midflight experience&quot;</td>
</tr>
<tr>
<td>Collections</td>
<td>dct_isPartOf_sm</td>
<td>Related collection name</td>
<td>&quot;Solar eclipse&quot;</td>
</tr>
<tr>
<td>Geometry type</td>
<td>dc_geotype_s</td>
<td>Point or polygon</td>
<td>&quot;point&quot;</td>
</tr>
<tr>
<td>Document type</td>
<td>dc_format_s</td>
<td>Tweet or webpage</td>
<td>&quot;webpage&quot;</td>
</tr>
<tr>
<td>Hashtag</td>
<td>dct_hashtag_s</td>
<td>Hashtags of tweets</td>
<td>#solareclipse</td>
</tr>
</tbody>
</table>
### GETAR Portal - Schema

#### for text search
- dc_creator_sm
- dc_description_s
- dc_format_s
- dc_identifier_s
- dc_publisher_s
- dc_rights_s
- dc_subject_sm
- dc_title_s
- dct_isPartOf_sm
- dct_provenance_s
- dct_spatial_sm
- layer_geom_type_s
- layer_slug_s

#### for sorting text fields
- dc_publisher_s
- dc_title_s
- dct_provenance_s

#### for spell checking
- dc_creator_sm
- dc_publisher_s
- dc_subject_sm
- dc_title_s
- dct_provenance_s
- dct_spatial_sm

#### for suggestions
- dc_creator_sm
- dc_publisher_s
- dc_subject_sm
- dc_title_s
- dct_provenance_s
- dct_spatial_sm
GETAR Portal - Functionality

Home Page

http://mule.dlib.vt.edu:3033
GETAR Portal - Functionality

Facet Search
GETAR Portal - Functionality

Spatial Search
GETAR Portal - Functionality

Document Details

UnitAPI

Source Type: tweet
Subject(s): experience
Collection: WatchEclipse
Author(s): UnitAPI
Publisher: https://twitter.com
Place(s): America
Data & Time: 2017-08-21T23:47:03Z
Hashtags: #SolarEclipse2017
Document URL: https://twitter.com/UnitAPI/status/899779917046583297
GETAR Portal
- Functionality

Search History

Successfully saved your search.

Search History

- school shooting
  - Subject: experience
  - Data type: Point
  - Bounding Box: 171.9140624, 39.0229 - 20.12553, 852597

- school shooting
  - Author: booboo666
  - Subject: experience
  - Data type: Point
  - Bounding Box: 171.9140624, 39.0229 - 28.12553, 852597

- solar usa view
  - Author: booboo666
  - Subject: experience
  - Data type: Point
  - Bounding Box: 171.9140624, 39.0229 - 20.12553, 852597

- solar usa view
  - Author: booboo666
  - Subject: experience
  - Data type: Point
  - Bounding Box: 171.9140624, 39.0229 - 28.12553, 852597

- Author: wamaznews
  - Subject: experience
  - Data type: Point

- Author: wamaznews
  - Subject: experience
  - Data type: Point

- Author: wamaznews
  - Subject: experience
  - Data type: Point

- Author: wamaznews
  - Subject: experience
  - Data type: Point

- Author: wamaznews
  - Subject: experience
  - Data type: Point

- Author: wamaznews
  - Subject: experience
  - Data type: Point
GETAR Portal - Functionality

User login

Log in

Email
Password
Remember me
Log in

Sign up

Email
g471000@vt.edu
Password (6 characters minimum)
Password confirmation
Log in

Edit User

Email
g471000@vt.edu
Password (leave blank if you don’t want to change it)
6 characters minimum
Password confirmation
Current password (we need your current password to confirm your changes)
Update

Cancel my account

Unhappy?
Cancel my account

Back
GETAR Portal - Future Work

- Spatial and temporal representation visualization of data
- Integration of GETAR Portal with GETAR Visualization
- More Customization
- Performance Analysis
- Usability Testing
- Documentation for the future users/developers
GETAR Visualization

Functions:

- Search
- View Tweets
- Social Network
- User Information
- Timeline
- Keywords
- Geo-location
GETAR Visualization

Functions:

- Search
- View Tweets
- Social Network
- User Information
- Timeline
- Keywords
- Geo-location

http://mule.dlib.vt.edu/cs5604f17_fe/TweetBank/src/
GETAR Visualization - Project Framework

User input → Solr Query → PHP → Solr/Jena → JSON → Decoded Values → D3
GETAR Visualization - Framework

Feature 1 - Flexible
a. Easy to add new data types
b. Easy to extend new views

Feature 2 - Interactive
a. overview and detail
b. brushing and linking
c. exploratory searching
Search

Keyword:
- Single word
- Multiple words

Search fields:
- Tweet
- User name
- Hashtag

Cluster:
- DiamondRing
- WatchEclipse
- NoEnglish
- ...

Time range:
12/31/2016 19:00 - 12/31/2017 19:00

User info:
- Follower #
- Friends #
- Tweet #
- Favorite #
- User lang

User info range:
- 0-10
- 10-100
- 1000-10000
- ...

Num of friends:
100 TO 999
GETAR Visualization - Views: TweetList

1. Embed all tweet metadata (Query from SOLR Team)
   a. UTF-8
   b. Links
   c. Emoji

2. Highlighted with keywords (NER, POS from Tweet Team)
   a. HASHTAG
   b. Place
   c. Organization
   d. Time
   e. Number

3. Interactions:
   a. Search and highlight, rerank
   b. Search and update tag cloud
   c. highlight/unhighlight
GETAR Visualization - Social Network

Nodes: individual accounts
Links: mentions
Interaction: hover to show user name
GETAR Visualization - User information

- **Num of followers** dropdown
- **Bar chart**
- **Pie chart**
- **Off**

**Search History**

**Oregon**
- **9,635 tweeters**
- **Num of users**
- **Bar chart**
- **Pie chart**
- **Off**

- **Num of users**
- **10M - 100M**
- **1M - 10M**
- **0.1M - 1M**
- **10K - 100K**
- **1K - 10K**
- **100 - 999**
- **10 - 99**
- **0 - 9**

**California**
- **24,942 tweeters**
- **Num of users**
- **Bar chart**
- **Pie chart**
- **On**

- **Num of users**
- **10M - 100M**
- **1M - 10M**
- **0.1M - 1M**
- **10K - 100K**
- **1K - 10K**
- **100 - 999**
- **10 - 99**
- **0 - 9**

**Search History**
Timeline

Search History

3,793 tweets

Word 'Idaho' appears 286 times at 8/21/2017 12:24:00 pm
Visualization Views - Tag Cloud

D3-cloud: tag cloud generator for D3.js

1. Dynamic Update through Interactions
   a. Summarization of current search results.
   b. Help users narrow down tweet list
   c. Exploratory search

2. Generalization and Emotional Expression
   a. Key Words from POS/NER
   b. Emoji

3. Interaction Reaction
   a. Searching: update based on tweet-list
   b. Highlight: update based on filtered tweet-list
Visualization Views - Geo Map 1

Different Levels
- World map, US map, State map
Visualization Views - Geo Map 2

Overview and Details
- Zoom in and zoom out
- Click Countries/States to zoom in
- Click white parts (sea) to zoom out
Visualization Views - Geo Map 3

Extended based on TopoJSON
- More detailed data can be added
- Cities.JSON for all cities in CA
QUESTIONS?

Thanks go to the US National Science Foundation for supporting Global Event and Trend Archive Research (GETAR) through IIS-1619028.

Thank you!