### Overview

The image shows a search interface for a system named Blacklight. The interface allows filtering results by various fields such as Author, Language, Location, Year, Month, Hashtags, Mentions, and Type. The search results are displayed in a list format with detailed information for each entry including the Author, Time, Text, Collection, Location, Hashtags, Mentions, and a link to More Like This. Example entries are shown:

1. **user4**
   - **Author:** user4
   - **Time:** 2016-11-12T23:42:38Z
   - **Text:** @user1 @user2 George Clooney is so #sexy.

2. **user3**
   - **Author:** user3
   - **Time:** 2016-11-08T23:42:38Z
What we accomplished this semester

- Choose a front end development framework that would work for our needs
- Build a knowledge base for Rails and Blacklight
- Learn and finalize the data
- Build the front end for an information retrieval system from scratch which displays accurate data in an efficient way
Verification of Tools and Platforms

- GETAR needs a front end to be a viable information retrieval system.
- Many user tasks that must be included in the verification process: query, refine search, browse, visualize, analyze, etc.
- The front end is the communications channel between the IR system backend and the user.

Previous Semester:
- HUE

Recommended:
- Blacklight

Others:
- Elasticsearch
- Kibana
- Fusion
- Custom using Solr API
<table>
<thead>
<tr>
<th></th>
<th>HUE</th>
<th>Blacklight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
<td>Tool / GUI</td>
<td>Framework / GUI</td>
</tr>
<tr>
<td><strong>Audience</strong></td>
<td>Collection admins, Data scientists</td>
<td>Depends on chosen design</td>
</tr>
<tr>
<td><strong>Goals</strong></td>
<td>Collection visualization, advanced analysis.</td>
<td>Search engine, Document viewing, Query results</td>
</tr>
<tr>
<td><strong>Coupling</strong></td>
<td>Integrated with backend (Cloudera Search)</td>
<td>Low Coupling. Multiple instances possible.</td>
</tr>
</tbody>
</table>
Ruby on Rails

- **Pros:**
  - Same web application can have different environment with different gems.
  - Different gems can be added in anytime
- **Cons:**
  - Rails is a framework.

Server-side web application MVC framework for Ruby
Rails Architecture

Browser ➔ Web Server ➔ Public ➔ Routing

Rails Framework

Controller ➔ View ➔ Model

Database
Rails Architecture
Ruby gems

- Bundle-gem: Clean way to manage different gems and their dependencies.
- Rsolr: A Ruby client for Apache Solr.
- Blacklight: Provides a basic discovery interface for searching an Apache Solr index.
- Date range limit: Integer range limiting
- Devise: Flexible authentication solution
- Blacklight advanced search for implementing more like this search handler
- GeoBlacklight: Discovery and access for geospatial data
- D3 on rails for visualization using AJAX
Blacklight as a Gem

Features:

- Stable URLs
- Provides JSON, RSS, and Atom (XML) responses.
- Faceted searching
- Search queries can have different sets of fields
- Results sorting
- Records can be shared via email, SMS, or exported as formatted citation

Pros:
- Does not require local access to Cloudera
- Works with any version of Apache Solr
- Has facet search result
- Has easy CRUD functionality

Cons:
- Learning curve of Rails architecture and framework
**Blacklight**

Router:
- Blacklight.yml
- Router.rb

Controllers:
- Catalog_controller.rb
<table>
<thead>
<tr>
<th>Column Family</th>
<th>Column Name</th>
<th>Index</th>
<th>UI Element</th>
<th>Facet</th>
<th>Search Filter</th>
<th>List of Results</th>
<th>Single Result Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tweet archivesource</td>
<td>archive_source_s</td>
<td>--</td>
<td>Author</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tweet from_user</td>
<td>author_s</td>
<td>--</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tweet time</td>
<td>created_time_dt</td>
<td>--</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage clean-text-proanity</td>
<td>text_txt</td>
<td>--</td>
<td>Text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet readable-lang</td>
<td>language_s</td>
<td>--</td>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage language</td>
<td>collection_name_s</td>
<td>--</td>
<td>Collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet readable-collection</td>
<td>location_ss</td>
<td>--</td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet created-year</td>
<td>t_year_i</td>
<td>--</td>
<td>Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet created-month</td>
<td>t_month_i</td>
<td>--</td>
<td>Month</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet hashtags</td>
<td>hashtags_ss</td>
<td>--</td>
<td>Hashtag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet mentions</td>
<td>mentions_ss</td>
<td>--</td>
<td>Mention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet classification-label</td>
<td>classification_s</td>
<td>--</td>
<td>Classification?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet real-world-events</td>
<td>events_s</td>
<td>--</td>
<td>Event</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet sner-people</td>
<td>sner_people_s</td>
<td>--</td>
<td>People Mentioned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet sner-organizations</td>
<td>sner_org_s</td>
<td>--</td>
<td>Organization Mentioned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet sner-locations</td>
<td>sner_loc_s</td>
<td>--</td>
<td>Location Mentioned</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clean-tweet tweet-importance</td>
<td>t_importance_f</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage title</td>
<td>title_s</td>
<td>--</td>
<td>Title</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage author-publisher</td>
<td>author_s</td>
<td>--</td>
<td>Author</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage sub-urls</td>
<td>sub_urls_s</td>
<td>--</td>
<td>URLs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage domain-location</td>
<td>location_s</td>
<td>--</td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage organization-name</td>
<td>organization_s</td>
<td>--</td>
<td>Organization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>webpage webpage-importance</td>
<td>w_importance_f</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>doc-type doc-type</td>
<td>doc_type_s</td>
<td>--</td>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>doc-topic label-list</td>
<td>topic_label_s</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>doc-topic doc-probability</td>
<td>topic_probability_s</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>doc-cluster cluster-label</td>
<td>cluster_label_s</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>doc-cluster cluster-probability</td>
<td>cluster_probability_s</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Query
And
Results

Author: user1
Time: 2016-11-08T23:42:13Z
Text: I cannot believe that this collection is fake!
Collection: Collection1
Location: Blacksburg and Virginia
Hashtags: fake
More Like This: Click Here
Faceted Search
Additional Information for a Document

user1

- **Author:** user1
- **Time:** 2016-11-08T23:42:13Z
- **Text:** I cannot believe that this collection is #fake!
- **Collection:** Collection1
- **Location:** Blacksburg and Virginia
- **Hashtags:** fake
- **Classification:** FraudulentActivity
- **Event:** CollectionFraud and Fall16TweetFaking
- **Type:** tweet
User Authentication (UA) and Activity Management

- User signup using email
- User login with email and password
- User can save documents and searches
- User activity is logged in server log
Blacklight can use many different user authentication management systems, but is designed for Devise UA.

Users Table
- Username, password, email, type, IP address
- Useful for user studies

Searches Table
- User ID, user type, query parameters, time
- Useful for planning new features

Bookmarks Table
- User ID, user type, document ID, title
- Useful for document impact and importance

Tables can be modified, created as needed.
Future Work

User Authentication
GeoBlacklight and Leaflet
Visualizations
User Authentication, Activity, and Management

Improvements from the users:
- Security from outside IP addresses
- User types can have specialized views

Improvements from the queries / searches:
- Analyzing user activity, experience
- Plan areas for additional features

Improvements from the documents:
- Bookmarked documents indicate overall impact
- Relevance feedback for particular query
GeoBlacklight and Leaflet

Implemented for Solr jetty

For GETAR and IDEAL:

- Requires Solr 4.7+ (Current version 4.1)
- schema.xml
- solrconfig.xml

Better tool to explore in future: Leaflet for Ruby
D3 for rails

Json in Blacklight
- Using Json result from Blacklight for visualization in D3

Search result
- /catalog.json?search_field=all_fields&q=auckland

Facet list
- /catalog/facet/subject_topic_facet.json

Our Current progress
- Implemented D3 for simple Json data
Conclusion

Special thanks to
Dr. Fox
Sunshin Lee
(congratulations on defense!)
SOLR Team
Acknowledgements

IDEAL\textsuperscript{1} project

NSF Grants:

IIS-1319578

GETAR\textsuperscript{2} project

NSF Grant:

IIS-1619028

1. http://vtechworks.lib.vt.edu/handle/10919/47942
Questions?
Some functions for Blacklight in Catalog_controller.rb:

- config.add_facet_field
- config.add_index_field
- config.add_show_field
- Config.add_search_field

To display the result correctly, each field needs to have their own flags in schema.xml
Router.rb:

```ruby
Rails.application.routes.draw do
  concern :range_searchable, Blacklight::RangeLimit::Routes::RangeSearchable.new
  mount Blacklight::Engine => '/'
  Blacklight::Marc.add_routes(self)
  root to: "catalog#index"
  concern :searchable, Blacklight::Routes::Searchable.new

  resource :catalog, only: [:index], as: 'catalog', path: '/catalog', controller: 'catalog' do
    concerns :searchable
    concerns :range_searchable
  end

  concern :exportable, Blacklight::Routes::Exportable.new

  resources :solr_documents, only: [:show], path: '/catalog', controller: 'catalog' do
    concerns :exportable
  end

  resources :bookmarks do
    concerns :exportable

    collection do
      delete 'clear'
    end
  end
end
```
Appendix

GeoBlacklight Schema:

- Dublin Core Metadata Initiative

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
</table>
| uuid        | Unique identifier                                                          | http://purl.stanford.edu/vr593v7147 http://ark.ckan.org/ar.k/28722/bk0012h535q
|             |                                                                             | url:geodata.tufts.edu:Tuffs.CambridgeGrid100_04                        |
| dc:identifier | Unique identifier                                                          | Often the same as uuid but may be an alternate identifier              |
| dc:title    | Title                                                                       | Digital Map of Village Boundaries of Andhra Pradesh, India, 2001        |
| dc:description | Description                                                              | Village boundaries of Andhra Pradesh linked to Census 2001. Includes data for 26041 villages, 237 towns, 23 districts and 1 state. This layer is part of the Village Map of India which includes socio-demographic and economic Census data for 2001 at the Village level. |
| dc:rights   | Rights for access                                                          | Restricted, Public                                                     |
| dct:provenance | Source institution                                                        | Berkeley, Harvard, MassGIS, MIT, Stanford, Tufts                       |
| dct:references | URLs to related services                                                  | See Example 1. Uses JSON-LD syntax (http://www.w3.org/TR/json-ld) with role values based in part on CatInterop (https://github.com/OSGeo/Cat-Interop) link properties. |
| dc:creator* | Author(s)                                                                  | George Washington                                                     |
| dc:format   | File format of layer data                                                  | GeoTIFF, Shapefile                                                    |
| dc:language | Language                                                                   | English                                                                |
| dc:publisher | Publisher                                                                  | ML InfoMap                                                            |
| dc:relation* | URLs to related resources                                                 | http://sws.geonames.org/1252881/about.rdf                              |
| dc:subject* | Subject                                                                    | Census, Human settlements                                            |
| dct:type    | Resource type                                                               | Dataset, Image, PhysicalObject                                        |
| dct:spatial* | Spatial coverage and place names                                           | Paris, France                                                         |
| dct:temporal* | Years                                                                    | 2010                                                                  |
| dct:issued  | Date issued                                                                | 2/18/2008                                                             |
| dct:isPartOf* | Holding dataset                                                          | Village Maps of India                                                 |

Source: http://journal.code4lib.org/articles/9710
GeoBlacklight Schema:

- **Geospatial features:**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>georss:box</td>
<td>Bounding box as maximum values for SW NE</td>
<td>12.6 -119.4 19.9 84.8</td>
</tr>
<tr>
<td>georss:point</td>
<td>Point representation for layer as y, x — i.e., centroid</td>
<td>12.6 -119.4</td>
</tr>
<tr>
<td>georss:polyon</td>
<td>Shape of the layer as a Polygon in the form SW NW NE SE SW</td>
<td>12.6 -119.4 19.9 -119.4 19.9 84.8 12.6 84.8 12.6 -119.4</td>
</tr>
</tbody>
</table>

- **Layers information:**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>layer:id</td>
<td>The complete identifier for the WMS/WFS/WCS layer</td>
<td>druid:vr593vj7147</td>
</tr>
<tr>
<td>layer:geom_type</td>
<td>Geometry type for layer data.</td>
<td>Point, Line, Polygon, and Raster</td>
</tr>
<tr>
<td>layer:modified</td>
<td>Last modification date for the metadata record</td>
<td>2014-04-30T13:48:51Z</td>
</tr>
<tr>
<td>layer:slug</td>
<td>Unique identifier visible to the user, used for Permalinks</td>
<td>stanford-vr593vj7147</td>
</tr>
</tbody>
</table>

Source: http://journal.code4lib.org/articles/9710