
Front End Final Presentation

CS 5604 Information Storage and Retrieval, Dr. Edward Fox

Rachel Kohler, Patrick Sullivan, Reza Tasooji

Dec. 6, 2016

Virginia Tech, Blacksburg, VA 24061

Overview

The screenshot shows a web browser window with the Blacklight search interface. The browser's address bar contains a search query: `search_field=_ss&q=*`. The Blacklight logo is visible in the top left, and a "History" link is in the top right. Below the logo is a search bar with a dropdown menu set to "All Fields" and a "Search" button. A "Limit your search" sidebar on the left lists various filters: Author, Language, Location, Year, Month, Hashtags, Mentions, Event, and Type, each with a right-pointing arrow. The main content area shows the search results. At the top, it says "You searched for: * ✕" with a "Start Over" link. Below this is a pagination bar: "« Previous | 1 - 10 of 21 | Next »" and two buttons: "Sort by Time" and "10 per page". The first result is "1. user4", which is a tweet. The tweet details are: Author: user4, Time: 2016-11-12T23:42:38Z, Text: @user1 @user2 George Clooney is so #sexy., Collection: Collection4, Location: Nags Head, Hashtags: sexy, Mentions: user1 and user2, and a "More Like This" link with the text "Click Here". The second result is "2. user3", with details: 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**

Blacklight versus HUE

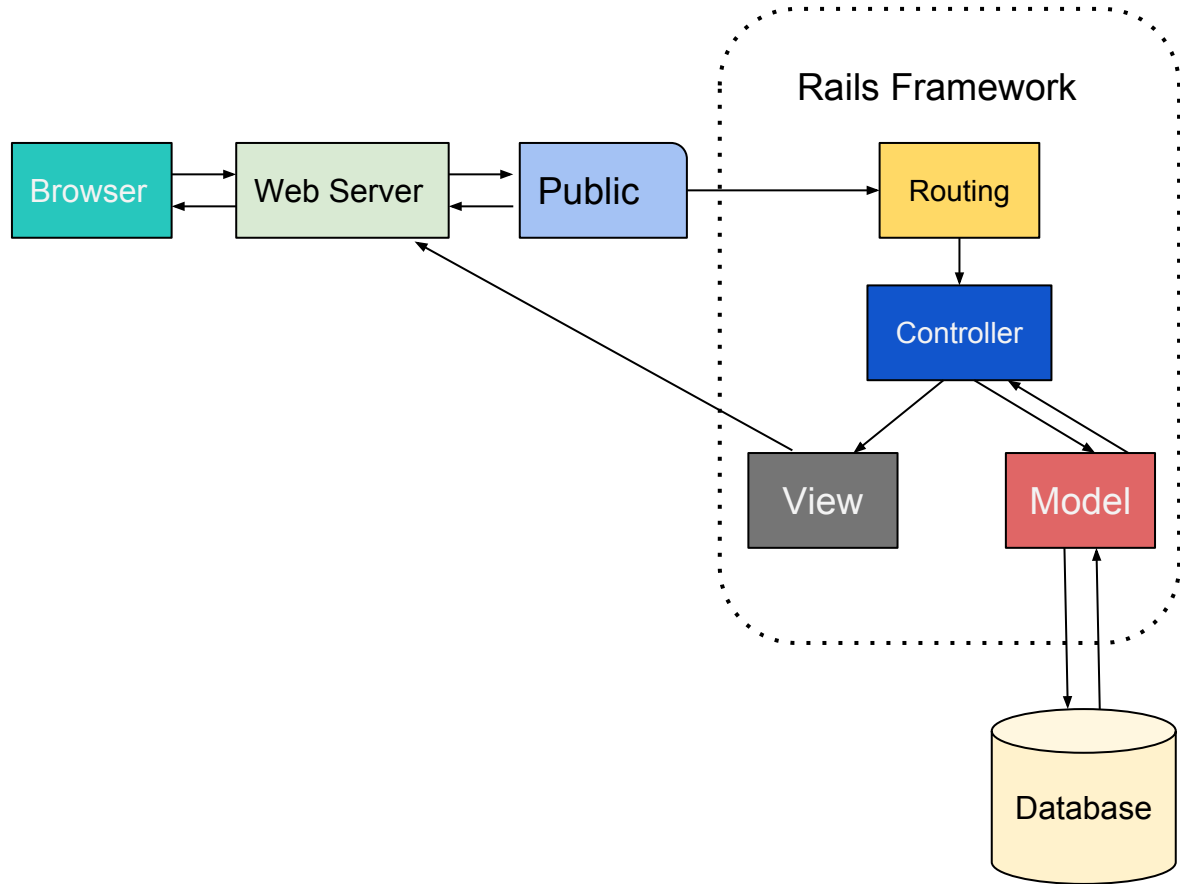
	HUE	Blacklight
Category	Tool / GUI	Framework / GUI
Audience	Collection admins, Data scientists	Depends on chosen design
Goals	Collection visualization, advanced analysis.	Search engine, Document viewing, Query results
Coupling	Integrated with backend (Cloudera Search)	Low Coupling. Multiple instances possible.

Ruby on Rails

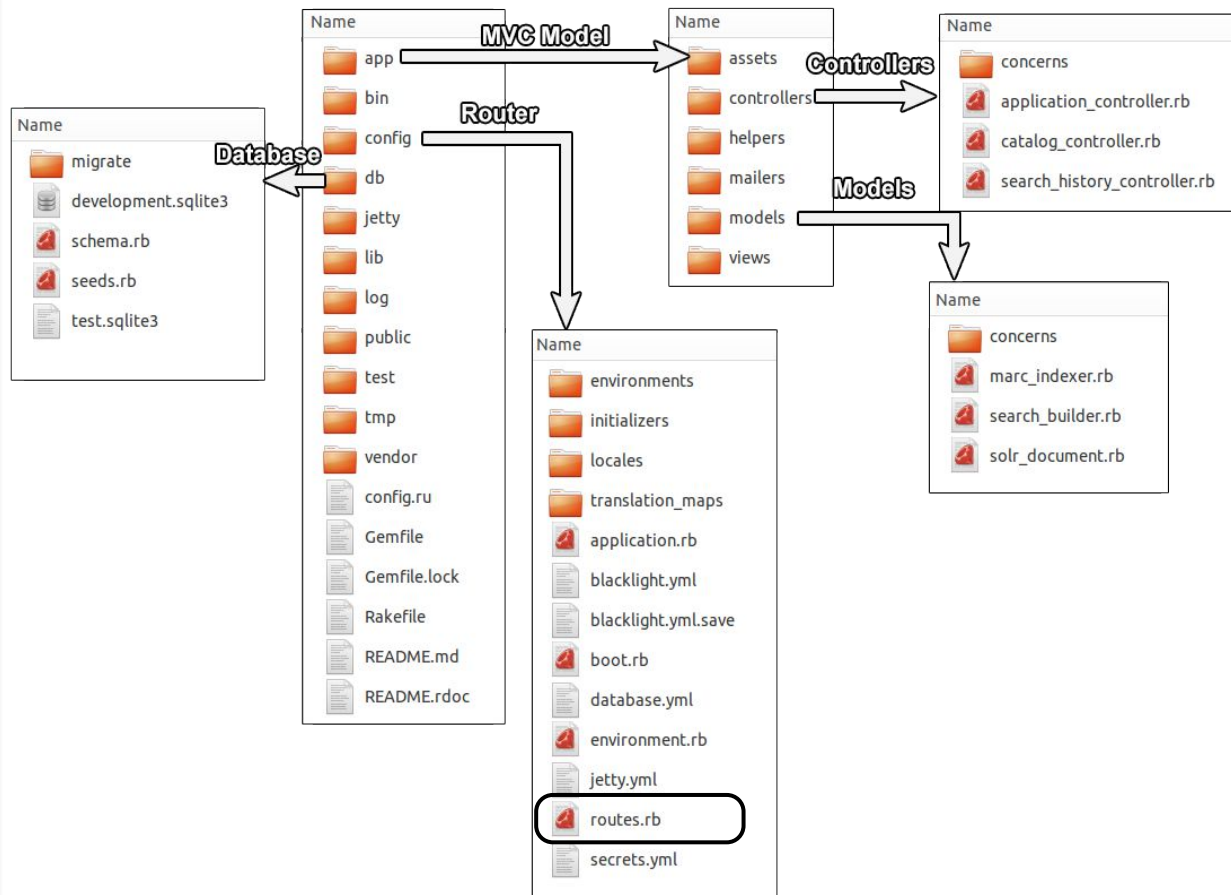
- Pros:
 - Same web application can have different environment with different gems.
 - Different gems can be added in anytime
- Cons:
 - Rails is a framework.

Server-side
web
application
MVC
framework for
Ruby

Rails Architecture



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

- 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

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

Blacklight

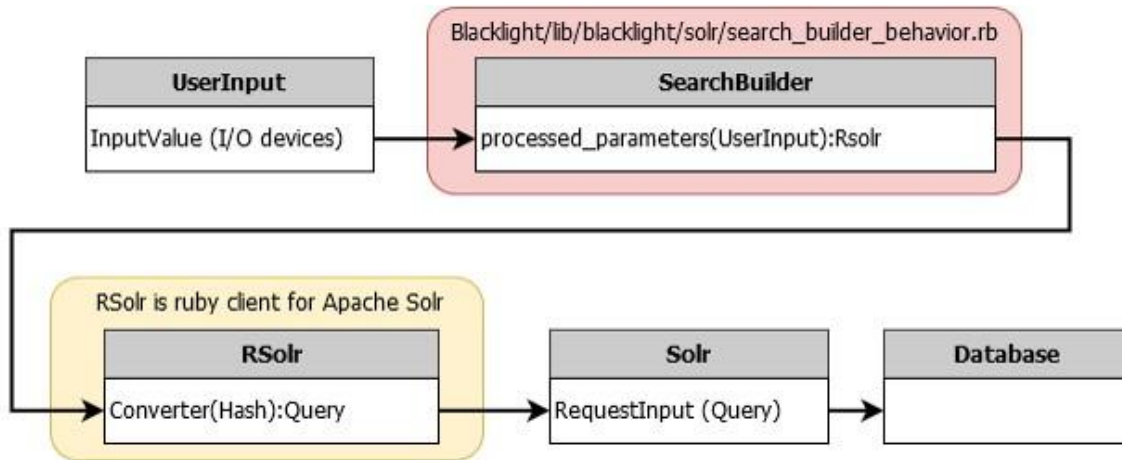
Router:

Blacklight.yml

Router.rb

Controllers:

Catalog_controller.rb



The Data

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

The Interface

The screenshot displays the Blacklight search interface. At the top, the browser's address bar shows the URL `search_field=_ss&q=*`. The Blacklight logo is on the left, and a 'History' link is on the right. Below the logo is a search bar with 'All Fields' selected and a search button. A 'Limit your search' sidebar on the left lists filters: Author, Language, Location, Year, Month, Hashtags, Mentions, Event, and Type. The main content area shows search results. The first result is '1. user4', with details: Author: user4, Time: 2016-11-12T23:42:38Z, Text: @user1 @user2 George Clooney is so #sexy., Collection: Collection4, Location: Nags Head, Hashtags: sexy, Mentions: user1 and user2, and a 'More Like This: Click Here' link. The second result is '2. user3', with details: Author: user3, Time: 2016-11-08T23:42:38Z, and Text: s*** why wont politics stop. Navigation links include 'You searched for: * x', 'Start Over', and '« Previous | 1 - 10 of 21 | Next »'. Sorting and pagination options are 'Sort by Time' and '10 per page'.

Query And Results

The screenshot shows a web browser window with the Blacklight search interface. The search bar is highlighted with a red box, showing the search term 'user1' and a 'Search' button. Below the search bar, there are filters for 'Limit your search' and a search history bar. The search results section shows '1 entry found' and a list of results. The first result is highlighted with a red box, showing the following details:

1. [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
- More Like This: [Click Here](#)

Faceted Search

The screenshot displays a Blacklight search interface. At the top, the browser tabs and address bar are visible. The main header includes the Blacklight logo and a search bar with the text "All Fields" and a search button. Below the search bar, a "Limit your search" sidebar is highlighted with a red box. This sidebar lists various facets: Author, Language, Location, Year, Month, Hashtags, and Mentions. The "Hashtags" facet is expanded, showing a list of tags and their counts: GOOGLE (1), dreams (1), fake (1), inane (1), magic (1), moneymoneymoney (1), sexy (1), and soFun (1). The main search results area shows "You searched for: *". Below this, there are navigation links for "Previous | 1 - 10 of 21 | Next »" and "Sort by Time" and "10 per page" options. The results list includes two entries: "1. user4" and "2. user3". Each entry displays metadata such as Author, Time, Text, Collection, Location, and Hashtags, along with a "More Like This" link.

blacklight project - Go... x blacklight/documen... x your bundle is locke... x Your Gemfile.lock i... x rtasooji/CETARFron... x +

8=/&search_field=ss&q=* Search

blacklight History

All Fields * Search

Limit your search

Author >

Language >

Location >

Year >

Month >

Hashtags ▾

GOOGLE 1

dreams 1

fake 1

inane 1

magic 1

moneymoneymoney 1

sexy 1

soFun 1

Mentions >

You searched for: * x Start Over

Previous | 1 - 10 of 21 | Next » Sort by Time - 10 per page -

1. user4

Author: user4

Time: 2016-11-12T23:42:38Z

Text: @user1 @user2 George Clooney is so #sexy.

Collection: Collection4

Location: Nags Head

Hashtags: sexy

Mentions: user1 and user2

More Like This: [Click Here](#)

2. user3

Author: user3

Time: 2016-11-08T23:42:38Z

Text: s*** why wont politics stop

Collection: Collection3

Location: New York

More Like This: [Click Here](#)

Additional Information for a Document

The screenshot shows a web browser window displaying a Blacklight search results page. The browser's address bar shows 'g/1-100' and the search bar contains 'Search'. The Blacklight logo is visible in the top left, and 'History' is in the top right. Below the search bar, there are links for 'Back to Search' and 'Start Over'. The main content area displays the document title 'user1' and a list of metadata fields. A red rectangular box highlights the following metadata:

- 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

To the right of the document details is a 'Tools' sidebar with the following options: Email, SMS This, and Cite.

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

UA with Devise

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

IDEAL¹ project

NSF Grants:

IIS-1319578

GETAR² project

NSF Grant:

IIS-1619028

1. <http://vtechworks.lib.vt.edu/handle/10919/47942>
2. <http://www.ctrnet.net/sites/default/files/GETARsummaryWeb.pdf>

Questions?

Appendix

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

Appendix

Router.rb:

```
config.ru | routes.rb | test.log | Telemetry Consent | Welcome
Rails.application.routes.draw do

  concern :range_searchable, BlacklightRangeLimit::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

Term	Description	Example
uuid	Unique identifier	http://purl.stanford.edu/vr593vj7147 http://ark.cdlib.org/ark:/28722/bk0012h535q urn:geodata.tufts.edu:Tufts.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
dc: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

Appendix

GeoBlacklight Schema:

- Geospatial features:

Term	Description	Example
georss:box	Bounding box as maximum values for S W N E	12.6 -119.4 19.9 84.8
georss:point	Point representation for layer as y, x — i.e., centroid	12.6 -119.4
georss:polygon	Shape of the layer as a Polygon in the form S W N W N E S E S W	12.6 -119.4 19.9 -119.4 19.9 84.8 12.6 84.8 12.6 -119.4

- Layers information:

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