

FEK Team Final Presentation

CS 5604 Information Storage and Retrieval, Dr. Edwards Fox
TA: Ziqian Song

Eddy Powell, Chao Xu, Han Liu, Rong
Huang, Yanshen Sun

Dec 10, 2019
Virginia Tech, Blacksburg, VA 24061

Outline

1. Introduction
2. Tools and Platforms
3. What We Have Achieved
4. Future Work
5. Conclusion





Introduction

- Provide an interface for the work completed this semester by all teams
- View the Tobacco and ETD datasets
- Use Kibana to manipulate and visualize the data



Tools and Platforms

- Elasticsearch, Kibana
- Node.js, Python, HTML, Javascript, CSS, MySQL, Reactivesearch
- Postman and Jupyter Notebook
- Ceph

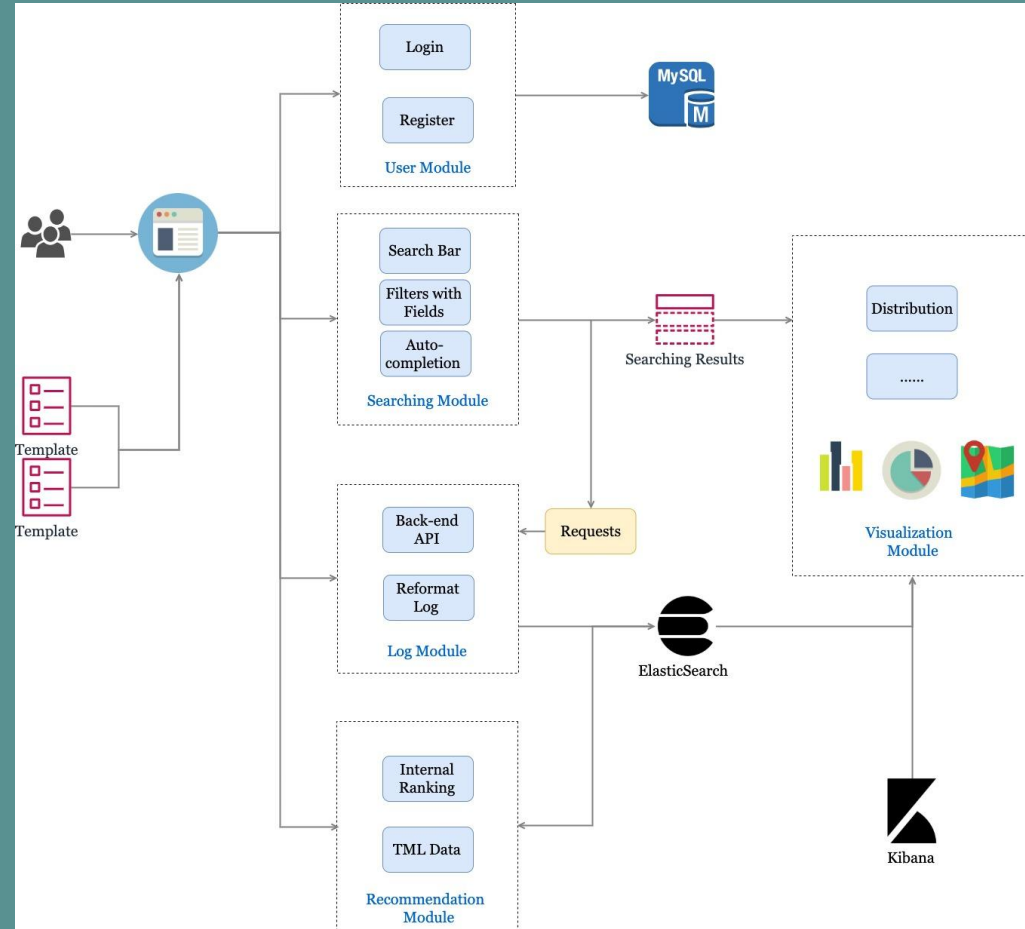


Achievements

- Instruction for using Kibana
- Instruction for using Postman (mainly for developers)
- **Build a user-friendly website with all functionalities**

Achievements

- User Module
- Searching Module
- Log Module
- Visualization Module
- *Recommendation Module





Admin

- Can monitor users on dashboard
- Perform CRUD operations for current users



Searching

Technique

Functionality

Demo

Explanation



Searching Components

- **JavaScript**
- **Create-react-app**: initialize the react application
- **Reactivesearch**: build UI components and connect Elasticsearch
- **Fancybox**: displays the searching page
- **HashRouter**: build multiple pages and routes in one app
- **Axios**: promise based HTTP client for the browser and node.js
- **Filepond**: support uploading files with fancy boxes



Searching Functionalities

- Ability to search on the ETD and Tobacco Datasets
- Support multiple filters with metadata and date
- Auto-suggestion in search bar
- Highlighting in results
- Customization of queries
 - Use of “&&” and “:” to search in specific few fields
 - Auto-suggestion starts from 3rd characters in search bar

Searching demo

FEK Home Visualization Datasets Uploads CiteSpace Profile Logout

Search for ETD

Search ETD

10000 results found in 7788ms

Type

Search

Thesis

Dissertation

Degree Level

Search

masters

doctoral

Date Issued

Start - End c

[Carve That Opossum and Plucky, Ducky UnderwearA Narrative Inquiry of Laughter in a Preschool Classroom](#)

by Smidl, Sarah Lynn , Stremmel, Andrew J. , undefined, Hill, Lynn T. ,Garrison, James W. ,Fu, Victoria R.
Pub: 2003-05-06

Carve That Opossum and Plucky, Ducky UnderwearA Narrative Inquiry of Laughter in a Preschool Classroom Sarah L. Smidl (ABSTRACT) This thesis is a narrative inquiry of laughter in a U the many situations in which children laugh as well as laughter's importance for the children, for me, and for all of us as a whole within the context of our classroom. To date, there is a pa young children. The majority of research that has been conducted has been quantitative in nature, with few attempts to comprehensively describe the many situations in which laughter oc and describe preschoolers' laughter, taking into consideration the many facets of their school day including free play, story time, playground time, and snack time. My sample included all c to 4 years, 4 months at the outset of the study. I also deemed it important to look at what these laughter-producing situations meant to me and the children in my classroom, including wha the classroom, and how it helped me to grow both personally and professionally through my research.

narrative inquiry ,laughter ,preschool
Thesis

[A Usability Problem Inspection ToolDevelopment and Formative Evaluation](#)

by Colaso, Vikrant , Hartson, H. Rex , undefined, McCrickard, Donald Scott ,Pérez-Quñones, Manuel A.
Pub: 2003-04-18

Usability inspection methods of user interaction designs have gained importance as an alternative to traditional laboratory-based testing methods because of their cost-effectiveness. How lacking a theoretical foundation. Other, more formal approaches like the cognitive walkthrough are tedious to perform and operate at a high-level, making it difficult to sub-classify problem formative evaluation of the Usability Problem Inspection tool â a cost-effective, structured, flexible usability inspection tool that uses the User Action Framework as an underlying knowledg particular task or a combination of tasks. It is also possible to limit the scope of inspection by applying filters or abstracting lower level details.

UAF ,UJEM ,Usability evaluation method ,UPI ,UPI tool
Thesis

[The Sorption of Roxarsone, an Organoarsenical Animal Feed Additive](#)

by



Searching demo

FEK Home Visualization Datasets Upload Q&A Profile Logout

Search for Tobacco

Search Tobacco

10000 results found in 2277ms

Document Type

Search

<input type="checkbox"/> letter	751019
<input type="checkbox"/> note	387542
<input type="checkbox"/> report	361207
<input type="checkbox"/> memo	318395
<input type="checkbox"/> email; letter	223885
<input type="checkbox"/> email	213103
<input type="checkbox"/> form	110836
—	

Availability

Search

<input type="checkbox"/> public	4708973
<input type="checkbox"/> restricted	886963



[EFFECT OF THE RELATIVE HUMIDITY OF INSPIRED AIR ON THE MORPHOLOGICAL APPEARANCE OF THE RAT LARYNX. SHORT TITLE: CONTROLLED H](#)
by
Pub: 1989-02-06
I GTB; RJR; AYRES PH; AVALOS JT; GERALD L; DELTECH ENGINEERING; CHARLES RIVER LABORATORIES; XYBION MEDICAL SYSTEMS; VERITAS LABORATORIES; PRI

[ECLIPSE/DTC. DIRECT MARKETING STATUS DECEMBER 13, 2000 \(20001213\) STATUS MEETING. DECEMBER 13: 34 WEEKS POST LAUNCH.](#)
by
Pub: 2000-12-13
VERMONT REQ25;US COMPREHENSIVE REQUEST 443;WALLACE 1RFP1;US COMPREHENSIVE REQUEST 448;US COMPREHENSIVE REQUEST 201;US COMPREHENSIVE REQ
1RFP9;VERMONT REQ23;VERMONT REQ18;VERMONT REQ15;VERMONT REQ7;VERMONT REQ17
MCGUIRE; PARADE; TV GUIDE; BRATTON D; DEPT OF JUSTICE; LABRECQUE M; RJR; JENNINGS C; SEVERANCE E; WEST MARKETING; BATTLE E; COLEMAN; ZIPPO; BLYNN C

[VORIDIAN SAICCOR PROGRAM.](#)
by
Pub: 2003-05-28
VERMONT REQ18;VERMONT REQ4;VERMONT REQ23

[PROJECT ALERT. RJR DATABASE MARKETING. SALEM MAILFILES TO YOUNG AMERICA.](#)
by
Pub: 1995-01-26
RJR; DRUM L; YOUNG AMERICA; SIMON H

[EFFECT OF THE RELATIVE HUMIDITY OF INSPIRED AIR ON THE MORPHOLOGICAL APPEARANCE OF THE RAT LARYNX. SHORT TITLE: CONTROLLED H](#)
by
Pub: 1989-02-06

Copyright © 2019 Made with  by FEK Team 

Searching requests

- Requests for auto-completion

▼ General

Request URL: http://2001.0468.0c80.6102.0001.7015.40b4.a1fb.ip6.name:9200/tobacco/_msearch?

Request Method: POST

Status Code: ● 200 OK

■ _msearch

■ emitlogs

■ **_msearch**

■ emitlogs

■ _msearch

■ emitlogs

■ _msearch

■ emitlogs

▼ Request Payload

```
{"preference": "search"}
{"query": {"bool": {"must": [{"bool": {"must": [{"bool": {"should": [{"multi_match": {"query": "stat", "fields": ["Brands^1", "Witness_Name^3", "Person_Mentioned^1", "Organization_Mentioned^1", "Title^5", "Topic^1"], "type": "best_fields", "operator": "or", "fuzziness": 0}}, {"multi_match": {"query": "stat", "fields": ["Brands^1", "Witness_Name^3", "Person_Mentioned^1", "Organization_Mentioned^1", "Title^5", "Topic^1"], "type": "phrase_prefix", "operator": "or"}]}]}]}]}}, {"minimum_should_matc
```

Searching requests

- Requests for searching results

▼ General

Request URL: http://2001.0468.0c80.6102.0001.7015.40b4.a1fb.ip6.name:9200/tobacco/_msearch?

Request Method: POST

Status Code: ● 200 OK

Remote Address: [2001:468:c80:6102:1:7015:40b4:a1fb]:9200

Referrer Policy: no-referrer-when-downgrade

▼ Request Payload

```
{"preference": "List"}
```

```
 {"query": {"bool": {"must": [{"bool": {"must": [{"multi_match": {"query": "status", "fields": ["Brands", "Witness_Name", "Person_Mentioned", "Organization_Mentioned", "Title", "Topic"]}}]}]}]}, "highlight": {"pre_tags": ["<mark>"], "post_tags": ["</mark>"], "fields": {"Brands": {}, "Witness_Name": {}, "Person_Mentioned": {}, "Organization_Mentioned": {}, "Title": {}, "Topic": {}}, "size": 5, "_source": {"includes": ["*"], "excludes": [], "from": 0}}
```

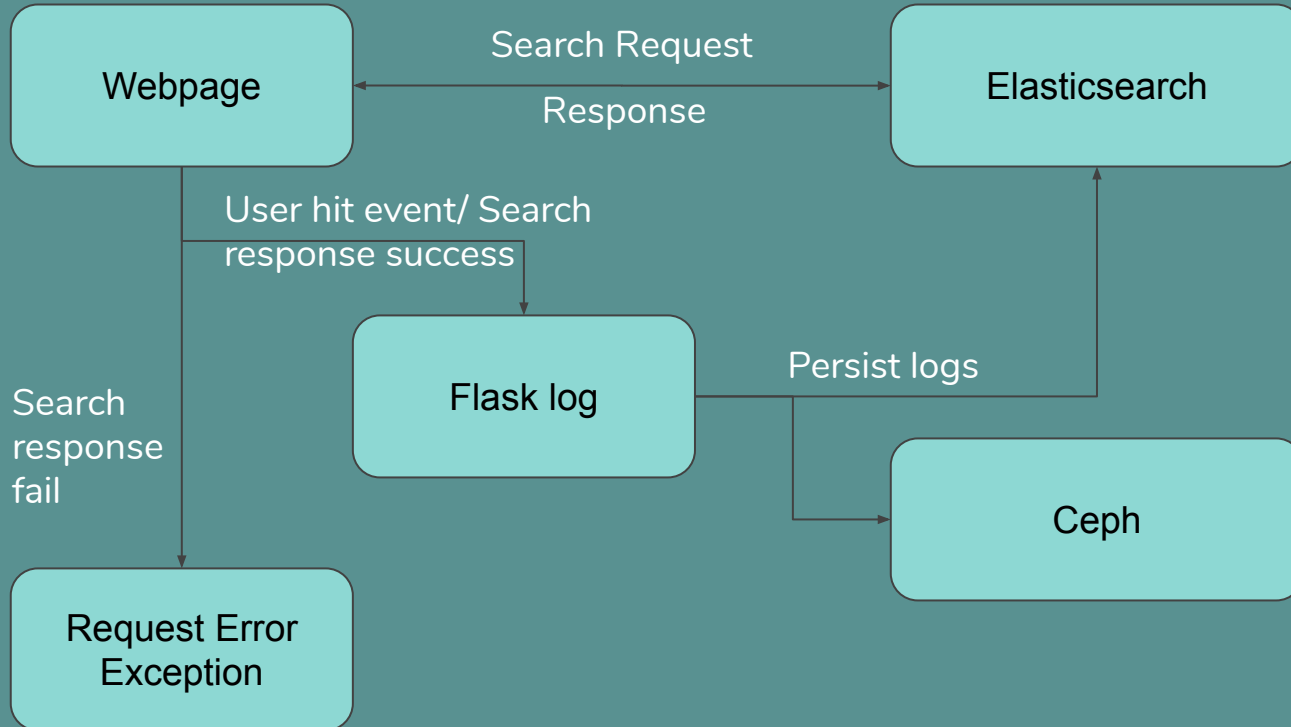


Log System

- A custom record of search query, filters applied, user information, and hitting events
- Saved both on Ceph and Elasticsearch

Log System

-- Data flow through requirements



Log System -- Example

Search log example

```
{
  "_index" : "tobacco_search_log",
  "_type" : "_doc",
  "_id" : "uSmx424BPKzKBT0mQcZU",
  "_score" : 1.0,
  "_source" : {
    "status" : 200,
    "message" : "Success",
    "data" : {
      "user" : {
        "username" : "test",
        "email" : "no email given"
      },
      "activity" : {
        "url" : "http://localhost:9200/tobacco3/_msearch?",
        "search_text" : "Deposition of THOMAS RICHARD ADAMS, March 15, 2000, WHITELEY v.
          RAYBESTOS-MANHATTAN INC.",
        "filters" : {
          "availability.keyword" : [
            "public"
          ]
        }
      }
    },
    "dataset" : "tobacco",
    "time" : "2019-12-07 23:07:54.939386",
    "ip" : "127.0.0.1"
  }
}
```

Hit log example

```
{
  "_index" : "tobacco_hit_log",
  "_type" : "_doc",
  "_id" : "g_-b7m4BcQ9g6KYS_kc1",
  "_score" : 1.0,
  "_source" : {
    "status" : 200,
    "message" : "Success",
    "data" : {
      "user" : {
        "username" : "test",
        "email" : "no email given"
      },
      "dataset" : "tobacco",
      "time" : "2019-12-10 02:00:30.580931",
      "ip" : "127.0.0.1",
      "hit" : { }
    }
  }
}
```

Original
document record



Visualizations for ETD and Tobacco

Goals:

- 1) Visualize the data with charts, maps, tables
- 2) Build user-friendly interfaces to display visualizations

Approaches:

- 1) Python Packages: matplotlib, pyecharts
- 2) Kibana

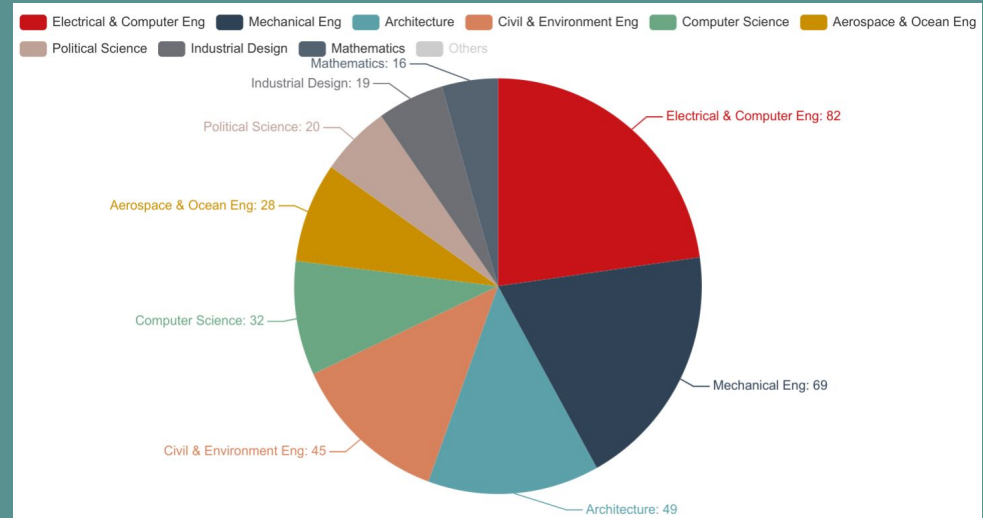
Visualizations - Python Packages

Advantages:

- 1) More flexibility of graph types
- 2) Allow us to process contents
- 3) Allow users to interact with data

Disadvantages:

- 1) Take too much time to clean and process the data
- 2) Hard to make it dynamic



Kibana Visualizations-Tobacco Settlement Documents

Types of visualizations include:
DataTable, Tag graph, Pie chart,
Area Graph, Gauge

The keywords utilized mainly
include: brands, cases, languages,
topics



A demonstration of Tag Graph

Kibana Visualizations-ETDs

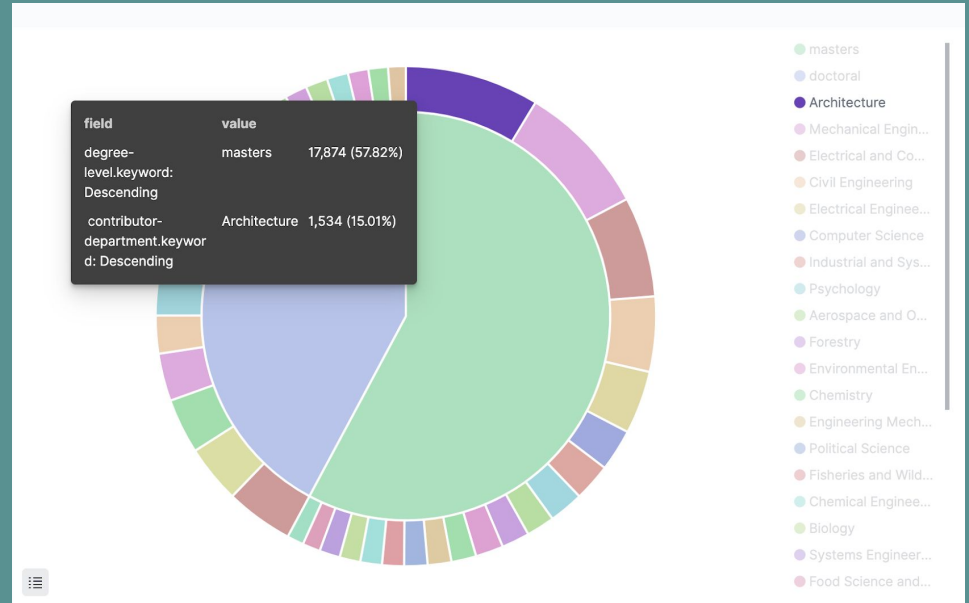
Kibana is used to create a series of visualizations for users to understand the ETD dataset.

Types of visualizations mainly include:

Table, Charts, Maps

The keywords utilized mainly include:

Level of Degree, Department, Discipline,
Issue Date



A demonstration of a pie chart

Kibana Visualizations

Add Data to Kibana
Use these solutions to quickly turn your data into pre-built dashboards and monitoring systems.

- Dashboard** (highlighted)
- APM**: APM automatically collects in-depth performance metrics and errors from inside your applications. [Add APM](#)
- Logging**: Ingest logs from popular data sources and easily visualize in preconfigured dashboards. [Add log data](#)
- Metrics**: Collect metrics from the operating system and services running on your servers. [Add metric data](#)
- SIEM**: Centralize security events for interactive investigation in ready-to-go visualizations. [Add security events](#)

Dashboards

[Create new dashboard](#)

Title	Description	Actions
ETD Dashboard		
Tobacco Dashboard		
[Flights] Global Flight Dashboard	Analyze mock flight data for ES-Air, Logstash Airways, Kibana Airlines and JetBeats	
[Logs] Web Traffic	Analyze mock web traffic log data for Elastic's website	
[eCommerce] Revenue Dashboard	Analyze mock eCommerce orders and revenue	

Rows per page: 10

Add panels

Search... Sort Types 3

- [eCommerce] Average Sales Per Region
- [eCommerce] Average Sales Price
- [eCommerce] Average Sold Quantity
- [eCommerce] Controls
- [eCommerce] Markdown
- ✓ Create new ...
- [Create new visualization](#)
- Create new ...

New Visualization

Filter

Pie
Compare parts of a whole

Visualizations: Gauge, Goal, Heat Map, Horizontal Bar, Line, Markdown, Metric, **Pie**, Region Map, TSVB, Tag Cloud, Timelion

[eCommerce] Orders

[Flights] Flight Log

- 30k**
- 30k_etd_metadata
- a*metadata
- accounts

Visualize Create

Save Share Inspect Refresh

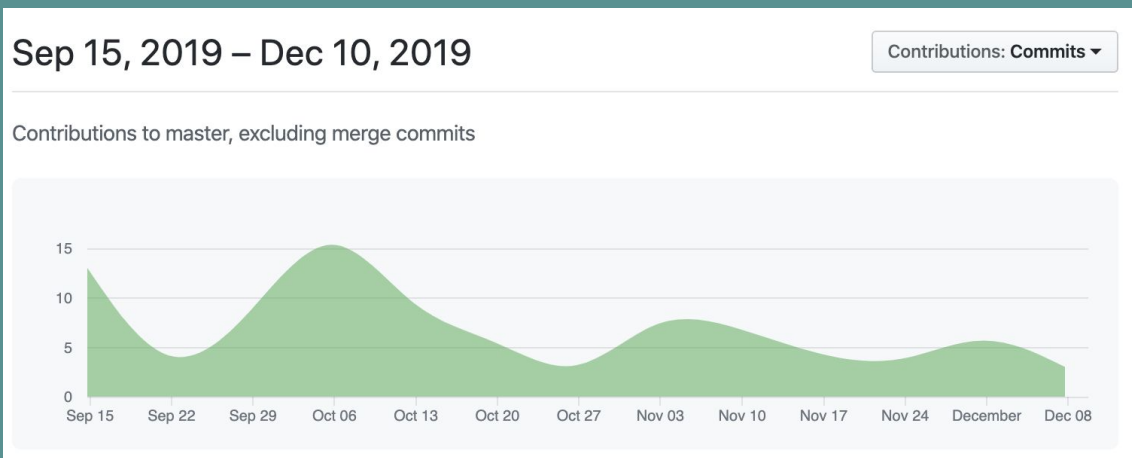
30k

Metrics: **ADD_BUCKET**, Split slices, Split chart

Buckets: Add

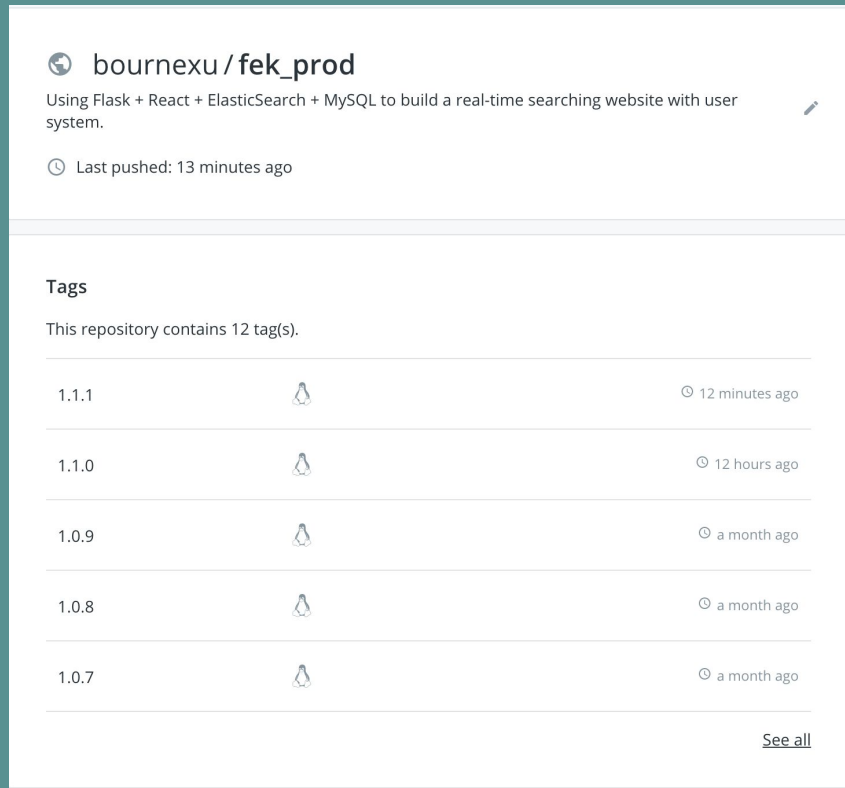
Summary

- Github 2 repos, each has **110+** commits:
 - 'master' for dev and test locally
 - 'prod' for cloud deployment








Summary

- We have released **10+ versions**
- Now, we are at version **fek_prod 1.1.1**



The screenshot shows the GitHub repository page for `bournexu/fek_prod`. The repository description is "Using Flask + React + ElasticSearch + MySQL to build a real-time searching website with user system." and it was last pushed 13 minutes ago. The "Tags" section lists 12 tags, with the following visible:

Tag	Icon	Time
1.1.1		12 minutes ago
1.1.0		12 hours ago
1.0.9		a month ago
1.0.8		a month ago
1.0.7		a month ago

[See all](#)



Future Work

- Complete unit tests for INT's CI/CD.
- Implement the TML team's recommendation module.
- Chapter 9 Section 5: Evaluation¹
 - Chapter 8 & Chapter 19 in textbook
- Welcome to our Github to report issues and give feedback in the future

1. Zhai, C., & Massung, S. (2016). *Text data management and analysis: a practical introduction to information retrieval and text mining*. Morgan & Claypool.



Conclusion

Special thanks for Dr.Fox and the CME, CMT, ELS, TML, INT groups.

Without the help from all of you guys, we couldn't have achieved as much!!!

Funding: IMLS LG-37-19-0078-19



Live Demo

<http://2001.0468.0c80.6102.0001.7015.b2eb.3731.ip6.name:3000/>

Our website must be run through the VT network. This requires being physically in range or using [VT's VPN service](#)