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 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
<table>
<thead>
<tr>
<th>Document Type</th>
<th>Value</th>
<th>Availability</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td></td>
<td>Search</td>
<td></td>
</tr>
<tr>
<td>letter</td>
<td>751019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>note</td>
<td>387542</td>
<td></td>
<td></td>
</tr>
<tr>
<td>report</td>
<td>361207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>memo</td>
<td>318395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>email; letter</td>
<td>223885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>email</td>
<td>213103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>form</td>
<td>110836</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EFFECT OF THE RELATIVE HUMIDITY OF INSPIRED AIR ON THE MORPHOLOGICAL APPEARANCE OF THE RAT LARYNX. SHORT TITLE: CONTROLLED HUMIDITY.**

by
Pub: 1989-02-06
IGTB; 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 1RF1/US COMPREHENSIVE REQUEST 448;US COMPREHENSIVE REQUEST 201;US COMPREHENSIVE REQ 1RF9/VERMONT REQ22;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; BLYN

**VORIDIAN SAICCOR PROGRAM.**

by
Pub: 2003-05-28
VERMONT REQ18/VERMONT REQ4/VERMONT REQ23

**PROJECT ALTERT. RJR DATABASE MARKETING. SALEM MAIL FILES 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 HUMIDITY.**

by
Pub: 1989-02-06
IGTB; RJR; AYRES PH; AVALOS JT; GERALD L; DELTECH ENGINEERING; CHARLES RIVER LABORATORIES; XYBION MEDICAL SYSTEMS; VERITAS LABORATORIES; PRI
Searching requests

- Requests for auto-completion
**Searching requests**

- Requests for searching results

**General**
- Request Method: POST
- Status Code: 200 OK
- 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

Webpage

Search Request
Response

Elasticsearch

User hit event/ Search
response success

Flask log

Search response
fail

Request Error
Exception

Persist logs

Ceph
Log System -- Example

Search log example

```
{
    "_index": "tobacco_search_log",
    "_type": "_doc",
    "_id": "uSmx42BPKxZBTomqZU",
    "_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-b7m48cQ9g6KYs_kcl",
    "_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

---

(Article content can be anything you want it to be, but it should be related to the examples provided.)

---

(Article content can be anything you want it to be, but it should be related to the examples provided.)

---

(Article content can be anything you want it to be, but it should be related to the examples provided.)
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
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
Future Work

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

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