

EE/CprE/SE 491 - sddec22-19

Text extractions from documents into Elasticsearch

Week 8 Report

3/28/22 - 4/3/22

Client: Kayla Gilleland, Buildertrend

Faculty Advisor: Goce Trajcevski

Team Members:

Bruce Bitwayiki

Jared Hayashi

Rushal Sohal

Tiffany Mayberry

Weekly Summary

This week our focus was refining our projects testing plan and gave an in-class presentation on our design plan. As far as our testing plan we researched and selected testing libraries we are going to use, and thought through different types of tests (unit, interface, integration, system, regression, and acceptance testing). Our presentation on our design received good feedback from our class.

Past Week Accomplishments




- Overview of our testing plan:
 - Critical Components:
 - Web application
 - Elasticsearch Handler
 - Tika text extractor
 - Types of tests:
 - Unit
 - Interface
 - Integration
 - System
 - Regression
 - Acceptance
- UI Prototype
 - We created a rough prototype of the UI to show what it will look like for our presentation. See UI Prototype Screenshots to get a sense of what it looks like. It has limited functionality (shows a static set of results and filters files based on if it's a pdf) but is hosted [here](#).
- Elasticsearch Setup
 - We setup elasticsearch on our server

- Install Docker
- Install Elasticsearch and kibana endpoints

UI Prototype Screenshots

Q abc co FILTER(S) SEARCH UPLOAD FILE

4 results (0.7 seconds)

-  **fake_pdf_file.pdf**
/User/Documents/Clients/
2 matches: "ABC Co" "PDF"
-  **fake_pdf_file_v2.pdf**
/User/Documents/Clients_v2/
2 matches: "ABC Company" "PDF"
-  **TheVeryImportantDocument.docx**
/User/Downloads/Clients/
1 match: "ABC Co"

Q abc co FILTER(S) SEARCH UPLOAD FILE

4 results (0.7 seconds)

Filters

File types:
 .pdf .txt .docx .pptx .xlsx .png
 .jpg

Author:





Upload date:

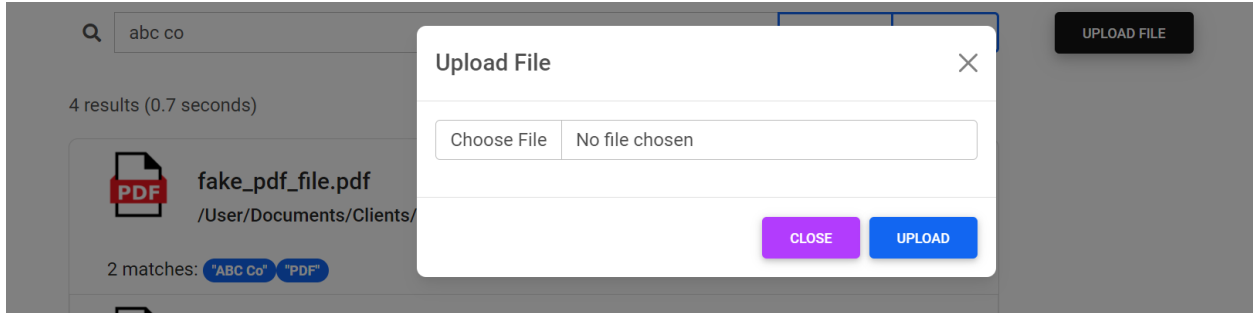
File created date:

Last Updated Date:

Containing Folder:

CLOSE APPLY

-  **fake_pdf_file.pdf**
/User/Documents/Clients
2 matches: "ABC Co" "PDF"
-  **fake_pdf_file_v2.pdf**
/User/Documents/Clients
2 matches: "ABC Company" "PDF"
-  **TheVeryImportantDocument.docx**
/User/Downloads/Clients
1 match: "ABC Co"
-  **anotherTextFile.txt**
/User/Documents/Rando



Individuals Contributions

| Name | Individual Contributions | Hours This Week | Hours Cumulative |
|------------------|--|-----------------|------------------|
| Bruce Bitwayiki | Elasticsearch research and setup | 3 | 18 |
| Jared Hayashi | Presentation, Elasticsearch research | 3 | 18 |
| Rushal Sohal | Presentation, testing plan | 3 | 16 |
| Tiffany Mayberry | Created rough prototype of UI in react | 10 | 31 |

Plans for the Upcoming Week

- Begin composing final design document
- Start planning faculty presentation
- Add more functionality to the prototype
- Begin work on Elasticsearch prototype

Summary of Weekly Advisor Meeting

- Reviewed project testing document and received advice for all of the testing sections
- Began talking about expectations for our final presentation