EE/CprE/SE 492 - sddec22-19

Text extractions from documents into Elasticsearch

Report 12

9/30/22 - 10/11/22
Client: Kayla Gilleland, Buildertrend
Faculty Advisor: Goce Trajcevski

Team Members:

Bruce Bitwayiki
Jared Hayashi
Rushal Sohal
Tiffany Mayberry

Weekly Summary

Our team was able to focus on intercomponent communication the past two weeks. We were able to get both the UI and the text extractor components to make requests to our Elasticsearch instance. We also began writing tests.

Past Week Accomplishments

- Tika is now able to send http requests via the elastic dependency
- Elasticsearch Handler is able to connect our instance of Elasticsearch to our UI when ran locally (localhost). Currently has two endpoints working: one for searching content and another for searching for filepath.
- Began initial setup and creation for UI tests.

Individuals Contributions

Name	Individual Contributions	BiWeekly Hours	Hours Cumulative
Bruce Bitwayiki	Add more sample docs to ElasticSearch for testing. Analyze various querying options and find best/most optimized option.	5	35
Jared Hayashi	Migrated project over to main directory as a maven project. Developed code for communication between Tika and Elasticsearch, refactored code, and wrote test cases.	8	40

Rushal Sohal	N/A (had an emergency at home)	0	32
Tiffany Mayberry	Developed Elasticsearch Handler to handle the connection between Elasticsearch API and our UI. Began work on hosting the Elasticsearch Handler on our server. Began writing basic UI tests.	8	57

Plans for the Upcoming Weeks

- Setup Spring server for Tika so the frontend can send files to it
- Working on uploading files on the frontend side
- Open port for elasticsearch communication or look for alternative communication method
- Looking into OCR for images

Summary of Weekly Advisor Meeting

- Start documenting test cases including trivial ones
- Ask client for more feedback and about test cases (and then document them for acceptance testing)
- Ask client in depth the limits they want of explore (for tika extractor)
- Ask client for some sample data, if needed
- Deviations from the design documentation should be included in the final documentation