

EE/CprE/SE 492 - sddec22-19

Text extractions from documents into Elasticsearch

Week 10 Report

8/22/22 - 9/15/22

Client: Kayla Gilleland, Buildertrend

Faculty Advisor: Goce Trajcevski

Team Members:

Bruce Bitwayiki

Jared Hayashi

Rushal Sohal

Tiffany Mayberry

Weekly Summary

In the first few weeks of the semester we have reviewed our project schedule and began working on the scheduled tasks. On the frontend, we began work on the search bar and result view components. On the backend, we setup Elasticsearch and refactored out Tika implementation.

Past Week Accomplishments

- Met with client to discuss current progress and plans moving forward
- Began frontend/backend implementations and testing framework

Individuals Contributions

Name	Individual Contributions	Hours This Week	Hours Cumulative
Bruce Bitwayiki	Get ElasticSearch cluster up and running ready to process data/queries. Brainstorm ways to parse various files to Json for indexing	5	25
Jared Hayashi	Refactored Apache Tika implementation, created methods to extract text from text-based documents (e.g. docx, pdf, txt)	7	27
Rushal Sohal	Working on the result card and view to hold all the cards (learning js and css with it)	5	22
Tiffany Mayberry	Created search bar component and different	6	39

	styles for result card		
--	------------------------	--	--

Plans for the Upcoming Week

- Develop JSON format for communication between components
- Set up a timeline for milestone demo meetings with the client

Summary of Weekly Advisor Meeting

- Discussed about testing scenarios for the project - use cases, unit tests, etc.
- Broadening the scope (but still evaluating if we can achieve it)