

EE/CprE/SE 492 - sddec22-19

Text extractions from documents into Elasticsearch

Report 14

11/09/22 - 11/22/22

Client: Kayla Gilleland, Buildertrend

Faculty Advisor: Goce Trajcevski

Team Members:

Bruce Bitwayiki

Jared Hayashi

Rushal Sohal

Tiffany Mayberry

Weekly Summary

This week the focus was on cleaning up parts of the UI, combining components, and diving into text extraction from images.

Past Week Accomplishments

- Expanded queries by including optional text-based filters & fixed UI of entering those filters (ex: can't add the same text for the same filter)
- Combined both parts of the UI (Upload & Search / Results) together
- Backend can now extract text from images and store them in json files

Individuals Contributions

Name	Individual Contributions	BiWeekly Hours	Hours Cumulative
Bruce Bitwayiki	ElasticSearch Analyzer implementation; **limited bandwidth this week due to other classes, assignments**	5	55
Jared Hayashi	Worked on Tika OCR to extract text from images	5	57
Rushal Sohal	pushing spring server for uploads to vm, looking at test cases	4	58
Tiffany Mayberry	Extended search querying to be able to filter by different text filters & update UI of selecting different text filters	6	77

Plans for the Upcoming Weeks

- Continue Writing Tests
- Finish integrating components together (Upload + Tika)
- Begin working on our final presentation, poster, and report

Summary of Weekly Advisor Meeting

- Cancelled due to fall break