EE/CprE/SE 491 - sddec22-19
Text extractions from documents into Elasticsearch
Week 5 Report

2/2822 - 3/6/22
Client: Kayla Gilleland, Buildertrend
Faculty Advisor: Goce Trajcevski

Team Members:

Bruce Bitwayiki
Jared Hayashi
Rushal Sohal
Tiffany Mayberry

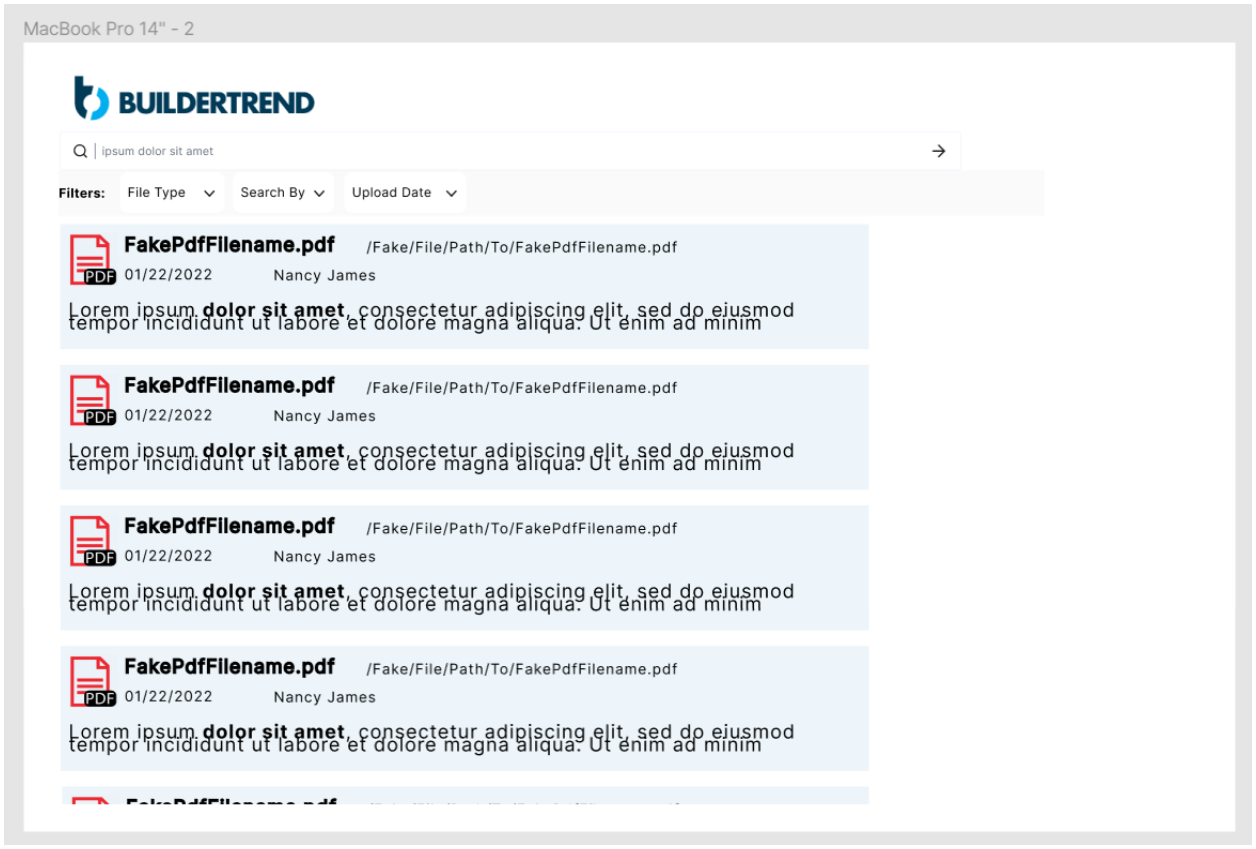
Weekly Summary

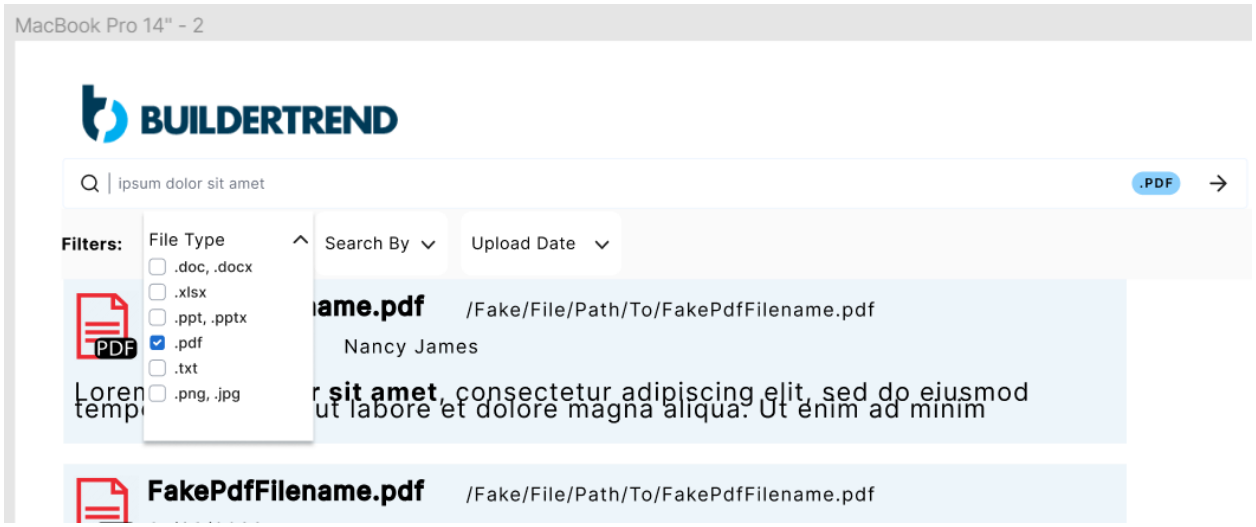
This week, we were mainly able to get our problem statement done with the needed requirements and constraints. We mapped the answers from our client to the desired sections. We now have a better overall understanding of our project. Besides, we were able to get some work started on the frontend and backend.

Past Week Accomplishments

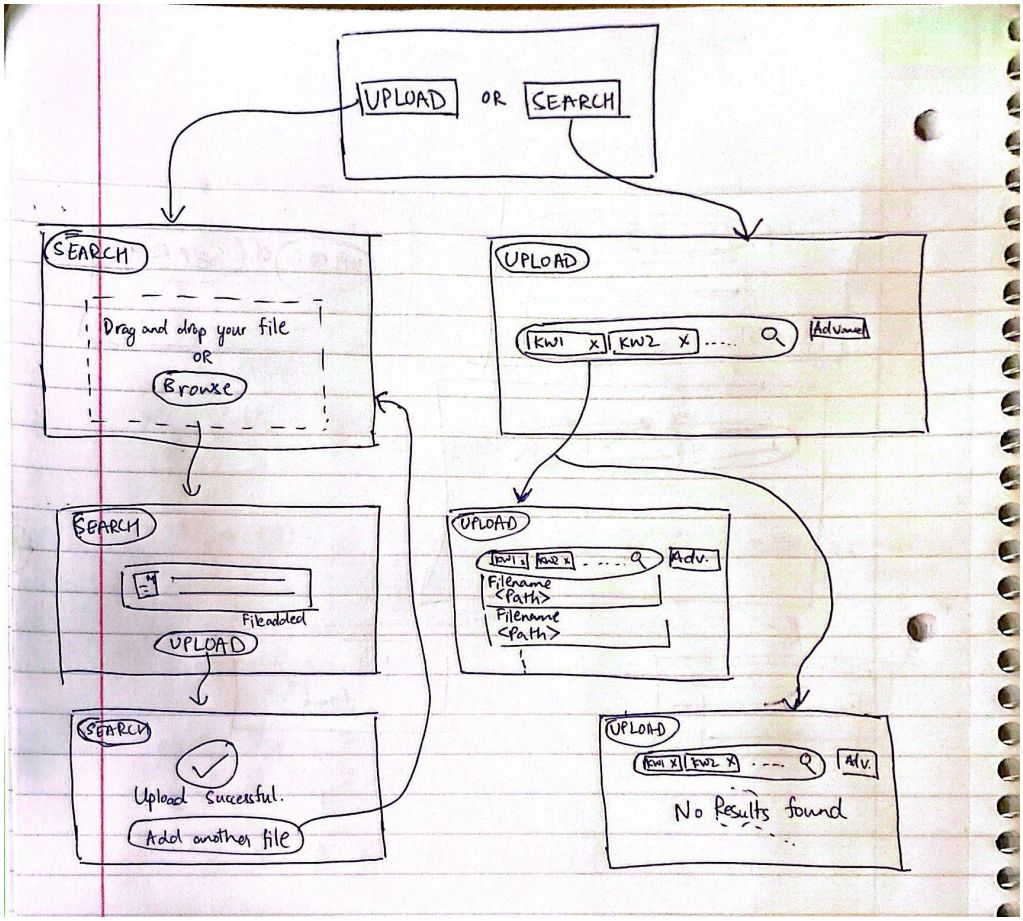
- Mapped answers to questions asked last week to requirements.
- Created an initial mockup of the user interface. See screenshots in the “UI Mockup” section.
- Done some basic UI/UX sketches for our webapp.
- Created an example app to extract data from the example files provided by the client using Tika. See example output
- Brainstorm and experiment with Tika-Elasticsearch architecture

UI Mockup


MacBook Pro 14" - 2
BUILDERTREND
ipsum dolor sit amet
Filters:
File Type
Search By
Upload Date
FakePdfFilename.pdf
/Fake/File/Path/To/FakePdfFilename.pdf
01/22/2022
Nancy James
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim
FakePdfFilename.pdf
/Fake/File/Path/To/FakePdfFilename.pdf
01/22/2022
Nancy James
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim
FakePdfFilename.pdf
/Fake/File/Path/To/FakePdfFilename.pdf
01/22/2022
Nancy James
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim
FakePdfFilename.pdf
/Fake/File/Path/To/FakePdfFilename.pdf
01/22/2022
Nancy James
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim


MacBook Pro 14" - 2
BUILDERTREND
ipsum dolor sit amet
.PDF
Filters:
File Type
.doc, .docx
.xlsx
.ppt, .pptx
.pdf
.txt
.png, .jpg
Search By
Upload Date
/Fake/File/Path/To/FakePdfFilename.pdf
Nancy James
FakePdfFilename.pdf
/Fake/File/Path/To/FakePdfFilename.pdf

Basic UI/UX sketches



Example Output

Name	File Type
architecture-first-floor-plan	image/jpeg
architecture-second-floor-plan	image/jpeg
bachelor-view-sample-house-construction-document-by-scott-harris-1920x1920	image/jpeg
bill-of-sale	application/vnd.openxmlformats-officedocument.wordprocessingml.document
BillsPOs	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
BTCostCodes_20220218	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
commercial-lease-agreement	application/vnd.openxmlformats-officedocument.wordprocessingml.document
construction-contract	application/vnd.openxmlformats-officedocument.wordprocessingml.document
Construction-Document-1	image/jpeg
Construction-Document-2	image/jpeg
Construction-Document-site-plan	image/jpeg
EstimateReport	application/vnd.ms-excel
Job Details - Buildertrend	application/pdf
Owner Invoice Print - 1	application/pdf
Owner Invoice Print - 2	application/pdf
Owner Invoice Print - 3 Voided	application/pdf
Owner Invoice Print - 3	application/pdf
ProposalPrint	application/pdf
receipt-template	application/vnd.openxmlformats-officedocument.wordprocessingml.document
Schedule List 2	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
Schedule_List_2	application/pdf
Schedule_List_2	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
ShortShoppingList	text/plain; charset=windows-1252
To-Do Print	application/pdf

Individuals Contributions

Name	Individual Contributions	Hours This Week	Hours Cumulative
Bruce Bitwayiki	Additional work on Tika, explore Tika and Elasticsearch architecture	3	10
Jared Hayashi	Researched Apache Tika and created sample code to extract file metadata	4	11
Rushal Sohal	Basic UI/UX sketches, looked up some development tools	4	8
Tiffany Mayberry	Created UI mockup	3	16

Plans for the Upcoming Week

- Brainstorm project plan document (milestones, gantt chart, project management method)
- Continue research on Tika and Elasticsearch for use cases
- Request a virtual machine from ETG

Summary of Weekly Advisor Meeting

- Reviewed requirements document
- Received suggestions for engineering standards
- Reviewed and added use cases to use case diagram
- Ask client if we should prioritize some files over others