Text Extractions from Documents into Elasticsearch

Team: sddec 22-19

Client: Buildertrend Advisor: Goce Trajcevski

Team Members: Bruce Bitwayiki, Jared Hayashi,

Rushal Sohal, Tiffany Mayberry

http://sddec22-19.sd.ece.iastate.edu/

Overview

The Client

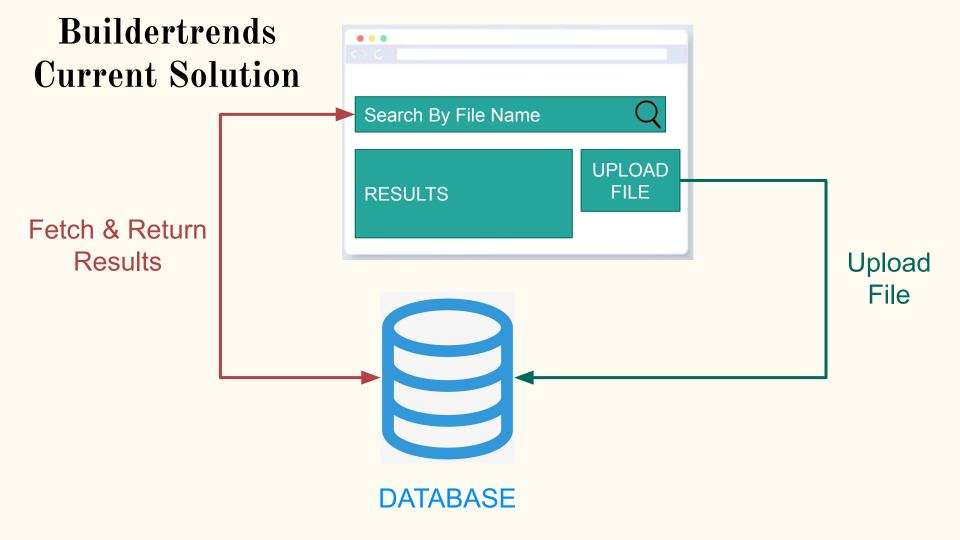
- Industry-Leader in Construction Management Software
- An all-in-one platform: set a budget, communicate with clients, schedule work and keep tabs on job site progress
 - Presales, Project Management, Financial Tools,
 Customer Management
- 27,000+ construction companies use Buildertrend

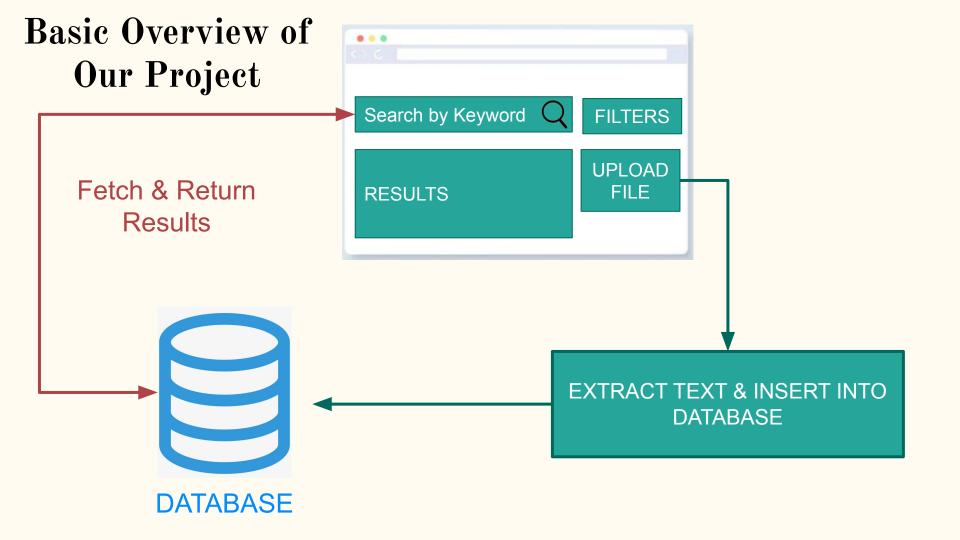












Task Responsibility

FRONT-END

BACK-END

Tiffany Mayberry (SE)

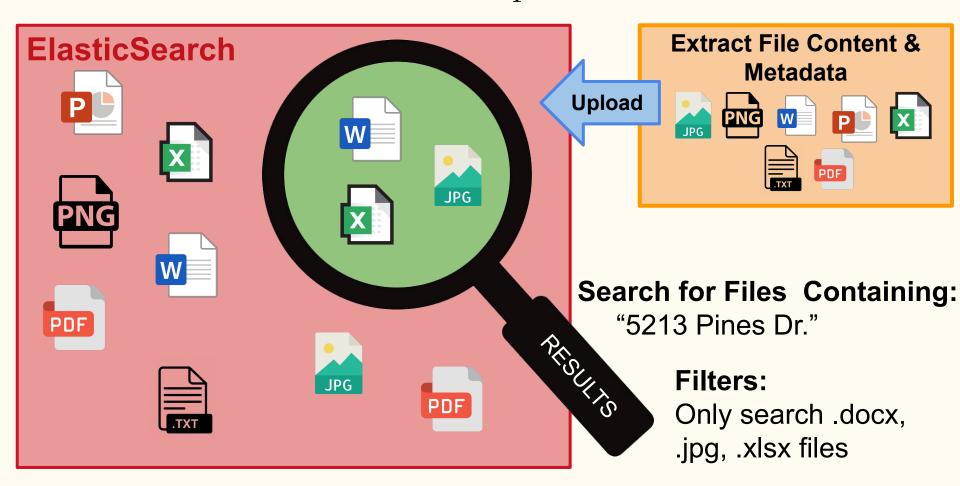
Rushal Sohal (CPR E)

Bruce Bitwayiki (CPR E)

Jared Hayashi (SE)

Requirements & Stuff

Functional Requirements



Non-Functional Requirements



Easy to use UI



Return result within 10 seconds



Easily modifiable and scalable for future functionality

Constraints and Considerations







Frontend should be a web or desktop application

Use **Elasticsearch** to store the metadata

Application should **limit** the returned results

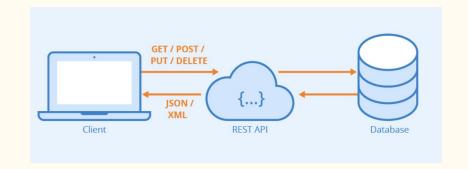
Elasticsearch

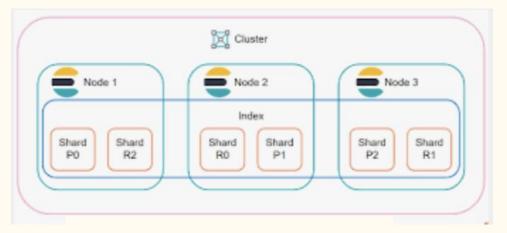
- Free and open search and analytics engine
- Based on the Lucene library
- Elastic Stack
- At core process JSON requests and return JSON data
- Index based searching
- Structure based on documents
- Netflix, Ebay, Walmart, etc.



Risk Management & Mitigation

- Project Planning
- UI Integration
- REST API
- Elasticsearch Node reliability
- Text Extraction handler
- Testing





(© Elasticsearch Structure)

Resource and Cost Estimate



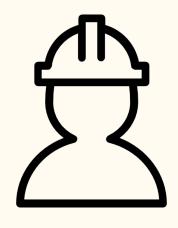


ETG Server

Open Source Packages

System Design

Users & Needs





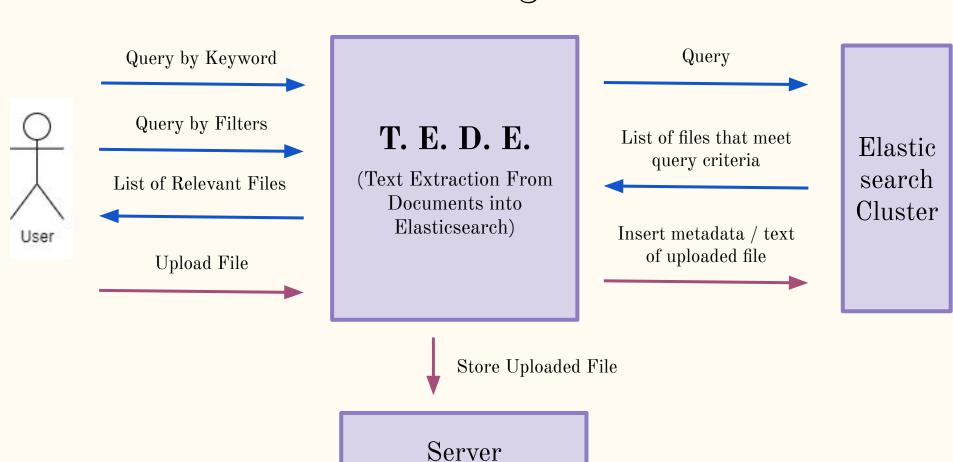


Homeowners
Contracts,
estimates/quotes,
specifications, notes

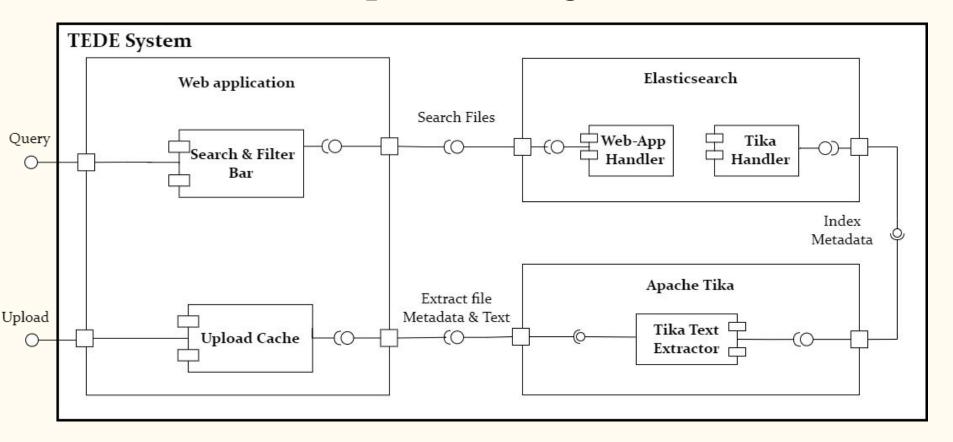


Buildertrend Staff Support

Context Diagram



Component Diagram



Technology Stack

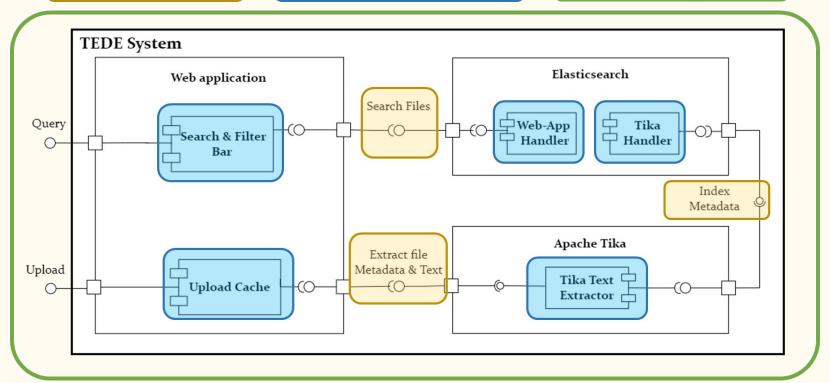
Web Application	Database	Text Extractor
React	{ REST:API }	0010111100001100 1100001011001010 1110000101101
JS	elasticsearch	Java

Testing Plan

Interface Testing
Postman

Unit Testing
Junit, Jest, React Test Library

Integration & System Testing



Project Plan

Project Milestones and Schedule

GOAL: Complete functionality before fall break

		-	JU	JA	LL	. (/111	hι	eu	ĴΙ	uI.	ıcı	υIO)11č	ıIII	y bere	JΓ	Ι	all	ŊJ	l e	aĸ											
Tasks	01/25/22	1/31/22	2/7/22	2/14/22	2/21/22	2/28/22	3/7/22	3/14/22	3/21/22	3/28/22	4/4/22	4/11/22	4/18/22	4/25/22	5/2/22	5/9/22	Break	8/22/22	8/29/22	9/5/22	9/12/22	9/19/22	9/26/22	10/3/22	10/10/22	10/17/22	10/24/22	10/31/22	11/7/22	11/14/22	11/21/22	11/28/22	12/5/22	12/12/22
1. Project Planning																																		
Complete Final Design Document														Х																				
Final Presentation		A												X																				
2 Develop Web Application UI						A																												
UI Prototype is Accepted by Client														Х																				
Complete Search Bar Implementation																				X														
Complete Filter Bar Implementation								¥								*	*					Х												¥
Comple File Upload Implementation								Break								Week	Summer Break					X									reak			Finals Week
3. Develop REST API						A										S	Jer														<u>m</u>			S
Complete Frontend-to-Elasticsearch								Spring								Finals	Ē							X							Fall			ina
Complete Frontend-to-Tika								S								ш	Su							X										
Complete Tika-to-Elasticsearch																								Х										
4. Setup Elasticsearch																																		
Installed Elasticsearch on Server											Х																							
5. Develop File Text Extractor																																		
Extract Text from All File Types																											Х							
6. Application Testing																																		
Client Acceptance					47	A																										X		

Project Schedule

Tasks/Subtasks	01/25/22	1/31/22	2/7/22	2/14/22	2/21/22	2/28/22	3/7/22	3/14/22	3/21/22	3/28/22	4/4/22	4/11/22	4/18/22	4/25/22	5/2/22	5/9/22	Break	8/22/22	8/29/22	9/5/22	9/12/22	9/19/22	9/26/22	10/3/22	10/10/22	10/17/22	10/24/22	10/31/22	11/7/22	11/14/22	11/21/22	11/28/22	12/5/22	12/12/22	Completed
1. Project Planning																																			
Team Building																																			
Team Initiation Assignment								ak								×	eak														Ų.			×	
Professionalism Assignment								Break								Finals Week	Summer Break														Break			Finals Week	
Team Website								ng								SIS	mer														<u>m</u>			SIS	
Requirements & Standards Planning								Spring								E	Ē														Fall			Ein	
Design Planning								o,									رة.																		
Testing Planning																																			
Project Planning																																			
2 Develop Web Application UI																																			31
Design core components																																			
Create UI Mockup								높								×	ak																	×	
Implement Prototype								Break								Finals Week	Summer Break														Break			Finals Week	
Implement Search Bar and Results View	/							ng E								Is	ner														<u>—</u>			Is	
Implement Filter Bar								Spring								ing	Ę														Fall			ing	
Implement File Uploader								0)								_	ũ	ш																	
Integrate REST API Calls																																			
Revise Implementation																																			
3. Develop REST API								¥								×	ak																	×	
Design Calls between Web App and Ela	stics	sear	ch					3re								Wee	Bre														Break			Wee	
Design Calls between Text Extractor and	d Ela	astic	sea	rch				ng E								S	ner														B			Is	
Implement Web App API Calls								Spring Break								Finals Week	Summer Break														Fall			Finals Week	
Implement Text Extractor API Calls								S								-	งี																	_	

Tasks/Subtasks	01/25/22	1/31/22	277722	2/14/22	2/21/22	2/28/22	3/7/22	3/14/22	3/21/22	3/28/22	4/4/22	4/11/22	4/18/22	4/25/22	5/2/22	5/9/22	Break	8/22/22	8/29/22	9/5/22	9/12/22	9/19/22	9/26/22	10/3/22	10/10/22	10/17/22	10/24/22	10/31/22	11/7/22	11/14/22	11/21/22	11/28/22	12/5/22	12/12/22	Completed
4. Setup Elasticsearch																																			
Request server to host								Break								Finals	Break														Break			Finals	
Install / Setup Elasticsearch on server								Ä								뜶	ă														ă			ᇤ	
Setup Elasticsearch Node																	7000																		
Develop File Text Extractor																																			
Design component interactions																																			
Install Apache Tika on server																																			
Implement file type detection																																			
Implement file meta data detection																	~																		
Implement .docx text extraction								eak								ek	Summer Break														¥			Week	
Implement .txt text extraction								Spring Break								Finals Week	8														Fall Break			¥.	
Implement .xlsx text extraction								ring								als	шe														all E			Finals	
Implement .pptx text extraction								S								直	JIN JIN														ш			亡	
Implement .pdf text extraction																																			
Implement .jpg text extraction																																			
Implement .png text extraction																																			
Integrate REST API Calls																																			
Revise Implementation																																			
6. Application Testing																																			
Create frontend unit tests																	~																		
Create text extractor unit tests																ek ek	rea																	Week	
Create elasticsearch handler unit tests								Break								Finals Week	Summer Break														Break			3	
Integration tests								ā								Jals	Ĕ														ā			Finals	
System tests																直	Sun																	直	
Regression tests																	(18,60)																		
User acceptance tests																																			

DEMO