# cEE/CprE/SE 492 - sddec22-19 Text extractions from documents into Elasticsearch Report 10

9/16/22 - 9/30/22 Client: Kayla Gilleland, Buildertrend Faculty Advisor: Goce Trajcevski

#### **Team Members:**

Bruce Bitwayiki Jared Hayashi Rushal Sohal Tiffany Mayberry

## Weekly Summary

We worked on our presentation for the first PIRM (Peer/Instructor Review Meetings) and also started working on communication between our components. The backend is now able to send data between the two components and the frontend is now able to query through http requests.

#### Past Week Accomplishments

- Created JSON format protocol for communication between components
- Apache Tika sending post requests to Elasticsearch
- Frontend experiments of sending get requests to Elasticsearch
- Revised Search Bar and Result Card Style
- Reflected on current progress status and struggles in preparation for PIRM

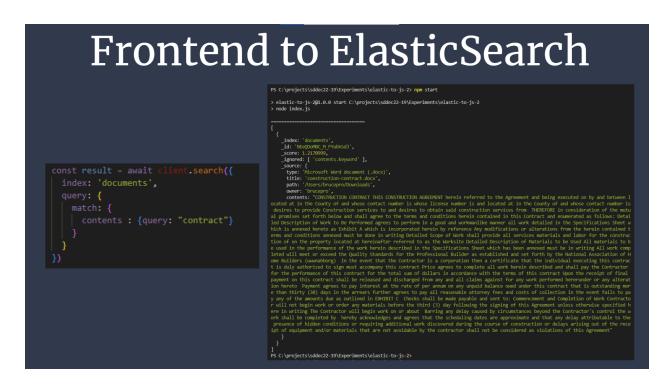
## Individuals Contributions

Name	Individual Contributions	BiWeekly Hours	Hours Cumulative
Bruce Bitwayiki	Create a data index and upload some sample documents to ElasticSearch for testing. Exploring various method of querying the data.	5	30
Jared Hayashi	Worked on Apache Tika conversion into JSON files as well as communication between Tika and Elasticsearch.	5	32
Rushal Sohal	Worked on the result view area and tried some styling to the card	10	32

Tiffany Mayberry Experimented with the frontend connection to Elasticsearch (trial and error), Edited the style of the result cards / search bar	10	49
--	----	----

## Visuals of Current Progress from Presentation

👹 React App	× +				- 🗆 ×
$\leftarrow \rightarrow G$	O D localhost:3000			☆	
<b>Q</b> ABC co	þ				
100×100 fake_file	-1.pdf /nloads/Clients/				
3 matches: ABC					
o materies.					
fake_file	e_2.pdf ownloads/Clients/				
2 matches: ABC					
2 matches.					
fake_file	e_1.pdf pwnloads/Clients/				
1 match: "ABC co"					
T Match. ABC co					
Doc	ument	[s -> Ti]	ka -> El	asticSe	earch
Doc	ument	$s \rightarrow Ti$	ka -> El	asticSe	earch
Doc	with with with the second sec	na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Property Value Origin	asticSe	earch
Doc	with with with the second sec	<u>s</u> -> Ti	Property Value Crigin Arroy Monsell Office Live Lat send by Monsell Office Live Revision under 3	asticSe	earch
Doc	with with with the second sec	na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Property Value Origin Origin <u>Reference</u> Last son-city Mound Offers Lear Last son-city Mound Offers Lear Vessen norther Program norther Company	asticSe	earch
Doc	ument	na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Property Welle Origin Origin Monsel Office Stare Laf survey by Monsel Office Stare Previous number Program may Monsel Office Wood Company Manager Content contents 4/2/2019 5/37 PM Date Set Wave 4/2/2019 5/37 PM	asticSe	earch
Doc	with with with the second sec	na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Property Value Origin Coligin	asticSe	earch
Doc		na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Property         Value         Property           Origin         Monuel (Stice Isser           Last survel by         Monuel (Stice Isser           Last survel by         Monuel (Stice Isser           Revision number         3           Version number         3           Revision number         3           Company         Manager           Context status         4/20/2019 5.25 FM           Data fording time (Volume 1)         5/2019 5.25 FM           Context status         Context status           Context status         Context status           Context status         2/2/2019 5.25 FM           Up (Status 2)         Status 2)           Value (Status 2)         Status 2)           Context status         Context status           Context status         2/2/2019 5.25 FM           Up (Status 2)         Status 2)           Context status         Context status           Context status         2/2/2019 5.25 FM           Up (Status 2)         Status 2)           Context status         2/2/2019 5.25 FM           Up (Status 2)         Status 2)           Context status         2/2/2019 5.25 FM           Up (Status 2)         Status 2)	asticSe	earch
		na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Property         Value         A           Origin         Color         Monitor Office New           Lat survey         Monitor Office New         Monitor Office New           Lat survey         Monitor Office New         Monitor Office New           Program mare         Monitor Office New         Monitor Office New           Constrainting         A         Monitor Office New           Data Inst survey         A/2.2019 5.57 PM         Data Inst survey           Constrainting         Monitor 00.00         Constrainting           Context start         Context start         Context start           Context start         Monitor Not open withmast effic         Pager		
Doc	<pre>value value v</pre>	na (ba) (ba) Sadah-Marah (10) 750 Matrixan	Popenty         Value         Origin           Origin         Monut Offers for           Lat surved by         Monut Offers for           Lat surved by         Monut Offers for           Program mare         Monut Offers for           Manager         Offers for           Content constant of 202019.527 PM         Date list samed           Date list samed         4/24/2019.527 PM           Date list samed         4/24/2019.527 PM           Content samed         Content samed           Content samed         Content samed           Violations         275           Word count         275           Violations         375           1 - 1         1		earch
Settings Help POST document { "types": PD "title": "D	s/_doc F (_pdp)*, do betatils2.pdf*,		Property         Value         Origin           Origin         Moreal Office Iser           Lat surved by         Moreal Office Iser           Lat surved by         Moreal Office Iser           Review number         3           Yearson number         3           Data lat surved         4/23/2019 547 PM           Data lat surved         4/23/2019 557 PM           Data lat surved         4/23/2019 52 PM           Lat prived         Tatal editory late (Lat Surved Area)           Content status         Content status           Content status         2010           Content status         2013		
Settings Help POST document { "type" = TO "title": "0 "poth" = "A "content": "b "content":	<pre>s/_doc  f(_cpdf)*,     do Details_pdf*,     services.setup.its PM Job Details - Buili     receives - Builings     receive</pre>	<pre>etrtrend localhost/summaryGrid.aspx</pre>	Property     Value     Origin       Origin     Moreal Cores have       Lat survel by     Moreal Cores have       Lat survel by     Moreal Cores have       Lat survel by     Moreal Cores have       Tatis and the core of th		
Settings Help POST document { type::-po titie::- poter:- "contents": ?chonQues that you" started	<pre>s/_doc F (_pdf)", c (pdf)", c (</pre>	dertrend localhost/summeryGrid.aspx we'll go through 3 steps to help you get start on inporting existing data , and	Property     Value       Origin     Grant Color Line       Lat sevel by     Monot Color Line       Lat sevel by     Monot Color Line       Lat sevel by     Monot Color Line       Property     Monot Color Line       Dele List used     4/2/2019.55 FM       Lat profile     Color Line       Content station     Monot Color Line       Content station     Monot Color Line       Content station     Color Line       Content station     Contents station       Version cont     2033       Version cont     2033       Version cont     1,       Contents static     Contents station       Content static     Contents static       Content static     Contentsta		
Settings Help POST document {"types": PD "title:" "onner:" ? raho@ues that you" customize Builder's Builder's Pr efia 3	s/_doc s/_doc	dertrend localhost/summoryGrid.aspx bb Select adMelcome to Buildertrend Now we'll go through 3 steps to help you get thermol User Subs/V endors Options ses ScheduleJob Name * Example JobsitaJob There are no items to select Add Edit	Property     Value       Origin     Moreout Office Law       Last succity     Moreout Office Law       Last succity     Moreout Office Law       Last succity     Moreout Office Value       Progen name     Moreout Office Value       Data ist succity     Moreout Office Value       Concernery     Moreout Office Value       Data ist succity     Moreout Office Value       Concernery     Moreout Office Value       Total defice gives     00000       Context succity     Moreout 2575       Owneart count?     2033       1 - {     *		
Settings Help POST document { "type": "PD "title: "D "poth": "A "contents": ? RihowQues that you" started - Guildor's Pr efix J Street A 2022Actua Marcoon Mo	<pre>s/_doc f (_pdf)", db Details_pdf', services_nget a lange s/_doc f (_pdf)", db Details_pdf', services_nget a lange f (_pdf)", f (_pdf)", f (_pdf)", f (_pdf)", f (_pdf) f (_pdf</pre>	dertrend localhost/summaryGrid.aspx bb Select edMelCome to Buildertrend Now we'll go through 3 steps to help you get we'll go through 3 steps to help you get stert on importing existing data , and they are no items to select Add Edit Zip Gode Gills Projected Start Feb 28, 2020 For en on items to select Add Edit Zip Gode Gills Projected Start Feb 28, 2020 Actual Completion Job Schedule Color uraday FridgyStatus Open Contract Feb 28,	Property     Value       Origin     Grand Color Line       Lat sevel by     Monot Color Line       Lat sevel by     Monot Color Line       Lat sevel by     Monot Color Line       Density     Monot Color Line       Total solution     Account of the Line       Content and Accounts of the Line     Content and Accounts of the Line       Content and Color Line     Content and Color		
Settings Help POST document { type::po title::p romer:: romer:	<pre>s/_doc  F (_pdf)*, do Details_pdf, intervention address F (_pdf)*, do Details_pdf, intervention F (_pdf)*, f (_pdf</pre>	dertrend localhost/summaryGrid.aspx bo Select definition existing adds, and setter on ignoring existing data, and tetran localhost/summaryGrid.aspx bo Select definition existing data, and tetrant liga through 3 steps to help you get start on inporting existing data, and tetrant Users Subs/V endors Options ess Schedul-Do Name "Example JobsiteDo There are no items to select Add Edit 2016 Gode 6312 Projected Subrer Feb 28, 2020 Fund Completion Job Scheder Price re Feet 6800 permit & Juch Trob Builder's is a savings of up to 50% off for	Popenty       Value         Origin       Monitol Office law         Last sorted by       Monitol Office law         Last sorted by       Monitol Office law         Data isotation of a sorted office Vision       Monitol Office law         Data isotation of a sorted office Vision       Monitol Office law         Data isotation of a sorted office Vision       Monitol Office Vision         Data isotation of a sorted office Vision       Context and Conte		
Settings Help POST document { "types": PD "path": "/U "contents": ?ahoQues that you" cutemiz Builder's Builder's D Street Ad 2022Actuu Marcon Wo SGS9,000, R Ris Insus Builderth Duilderth Builderth Street Ad 2022Actuu Marcon Wo SGS9,000, R Ris Insus Builderth Duilderth Duilderth Duilderth Street Ad Street	<pre>s/_doc s/_doc s/_d</pre>	dertrend localhost/summaryGrid.aspx b Select edWelcome to Buildertrend Now we'll go through 3 steps to help you get start on importing existing data , and there are no liters to select Add Edit New Ye'll go through 3 steps to Add Select Subs/V andros Dytions ass ScheduleJob Name * Example JobsiteJob There are no liters to select Add Edit 2022Actual Completion Job Schedule Color uraday FridayStusus Open Contract Pr ice re Feet 6000 Permit # Lot Info Builder's so ut how mach you can save on Builder's	Property     Value       Origin     Normal Cline Line       Lat surved by     Monost Cline Line       Lat surved by     Monost Cline Line       Property     Normal Cline Line       Context stand     Context stand       Context stand     Standstan Vid SpenneflowatedIL.       Context stand     Standstan Vid SpenneflowatedIL. <t< td=""><td></td><td></td></t<>		
Settings Help POST document { "types": PD "path": "/U "contents": ?ahoQues that you" cutemiz Builder's Builder's D Street Ad 2022Actuu Marcon Wo SGS9,000, R Ris Insus Builderth Duilderth Builderth Street Ad 2022Actuu Marcon Wo SGS9,000, R Ris Insus Builderth Duilderth Duilderth Duilderth Street Ad Street	<pre>s/_doc  F (_pdf)",</pre>	dertrend localhost/summeryGrid.aspx bo Select define the select Add Edit action of the select Add Edit action of the select Add Edit 2017 Gode Bills Projected Schedule Color marky FridgSTP Projected Start Feb 28, 2012 Gode Bills For Start Feb 28, 2013 Gode Bills For Start Feb 28, 2013 Gode Bills For Start Feb 28, 2014	Property     Value       Origin     Normal Cline Line       Lat surved by     Monost Cline Line       Lat surved by     Monost Cline Line       Property     Normal Cline Line       Context stand     Context stand       Context stand     Standstan Vid SpenneflowatedIL.       Context stand     Standstan Vid SpenneflowatedIL. <t< td=""><td></td><td></td></t<>		



## Plans for the Upcoming Weeks

- Setup Spring server for Tika so the frontend can send files to it
- Use the findings of the frontend experiments to implement the connections between the frontend and Elasticsearch
- Write test cases for the UI
- Begin implementing the filter bar

## Summary of Weekly Advisor Meeting

• Deferred due to PIRM meeting