

EE/CprE/SE 491 - sddec22-19

Text extractions from documents into Elasticsearch

Week 1 Report

1/25/22 - 2/6/22

Client: Kayla Gilleland, Buildertrend

Faculty Advisor: Goce Trajcevski

Team Members:

Bruce Bitwayiki
Jared Hayashi
Rushal Sohal
Tiffany Mayberry

Weekly Summary

The main goal for this week was to do preliminary research into Buildertrend and into tools described in the goals and stretch goals of the project. We prepared a set of questions to gain more information for the kickoff meeting this upcoming Monday.

Past Week Accomplishments

- Research into [Apache Tika](#) - Everyone
 - Open-source toolkit that detects and extracts metadata and text from numerous file types.
- Research into [Elasticsearch](#) - Everyone
 - Open-source search and analytics engine that can process nearly all kinds of data.
- Explore Potential Alternatives/Competitors to Apache Tika - Everyone
 - Apache Solr - Functionality of Lucene (indexing/searching based on text) with database support

Individuals Contributions

Name	Individual Contributions	Hours This Week	Hours Cumulative
Bruce Bitwayiki	Research client & proposed tech, Meeting notes	1	1
Jared Hayashi	Coordinated with client, Apache Tika research	1	1
Rushal Sohal	Initial meeting, some researching	1	1
Tiffany Mayberry	Coordinated with advisor, Research	1	1

Plans for the Upcoming Week

- Kickoff meeting with client on Monday
- Continue researching tools/frameworks for project (Apache Tika/Elasticsearch)

Summary of Weekly Advisor Meeting

- Reviewed project goals and stretch goals
- Brainstormed questions to ask for kickoff meeting