



Project Proposal (Team Eccentrics)

CSE 299

Junior Design

Section 04

Spring 2022

North South University

Submitted To: Prof. Intisar Tahmid Naheen (ITN)

Name	ID	Email
Mohammad Nabiluzzaman Nelay	1831251042	mohammad.nabiluzzaman@northsouth.edu
Rafid Ahmed	1831395642	rafid.ahmed@northsouth.edu

Project Name - Spitrack

Introduction

Spitrack is an online system that enables researchers to collect necessary documents and information from a given cluster of URLs, and then perform some research on the collected information using a specific search engine. Spittract provides the ability of parsing data from every document type like PDF, Word, PowerPoint, Text and HTML Page. The parsed data is stored in a database, so that it can be used for research purposes. The collection of data and documents will not be only from the given URLs in the cluster that will be created by the user, but also, the user will have the option to provide some depth of “the URLs” up to which the extraction of data needs to be done from a specific given URL. Spitrack will be available for both the Web and Mobile (Android) platforms.

Possible Features

- 1. User Log In** – Users will be able to Log In into the system using their registered “username” and password. Sign In using Google will be an additional feature.
- 2. User Sign Up** – Users will be able to register themselves into the system by providing some necessary information like: First Name, Last Name, Email, Gender, Username and Password.
- 3. Dashboard** – It will show all the basic information about the user, the number of clusters he/she has created and the recent searches that have been performed by them.
- 4. Cluster Creation** – Users will be able to create a cluster of three different URLs, provide some depth up to which the data needs to be collected, set the type of documents the cluster should collect, and give the cluster a name as per their choice. Once the cluster is set for creation, data is scrapped and saved into the database. Email notification on scrapping completion will be an additional feature.

5. Clusters – Users will be able to see the clusters that they have created. Each cluster will show the follow information: Name of the Cluster, Date of Cluster Creation, URLs, Selected types of Document, Depth.

6. Search – Users will be able search on the data that has been scrapped by selecting some pre-made clusters, and providing some keywords for the search. Each search results will show some title, a shortened content, type of document from where it has been extracted, the root and current URL of the document, and the depth where the document has been found.

7. Voice Search – Users will be able to search on the scrapped data through voice commands. This is an additional feature.

8. Admin Log In – Admins will be able to Log In to their respective admin accounts using their “username” and “password”.

9. Admin Dashboard– It will show all the basic information about the admin.

10. Add New Admin – Existing admins will be able to add new admins to the system. The information required for the new admin sign up are: Full Name, Email, Gender, Username, and Password.

11. Control Scrapping Strategies – Admins will be able to set the type of documents that can be scrapped by the users.

Software

1. UI Design – Adobe XD
2. Version Control System – Git
3. Repository Hosting Service – Bitbucket
4. Task Management – Trello
5. Communication Platform – Slack

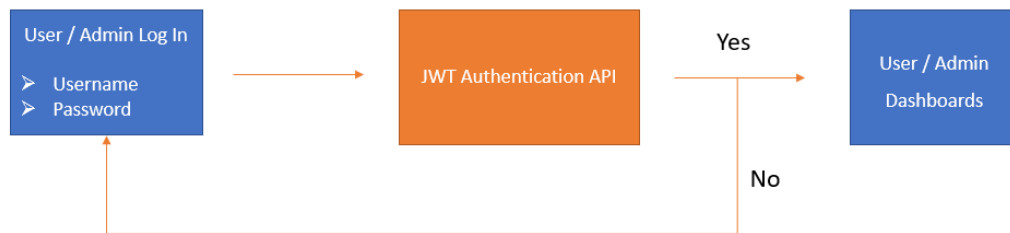
6. Backend Development Using Python (IDE) – JetBrains PyCharm
7. Frontend Development (IDE) – Visual Studio Code
8. Android Development – Android Studio
9. Web Server (SQL Database) – Apache XAMPP

Technologies

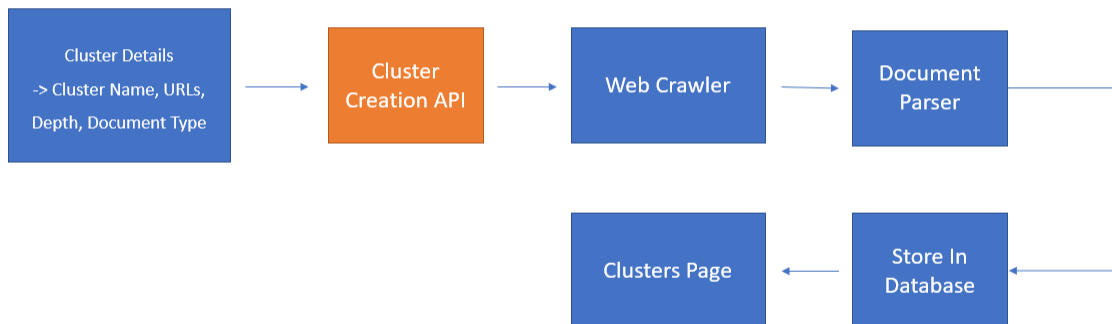
1. Django
2. React JS
3. Beautiful Soup
4. HTML
5. CSS
6. Bootstrap
7. JavaScript
8. Python
9. XML
10. Java
11. MySQL
12. Django REST Framework

Flow Chart or Architecture Diagram

Authentication



Cluster Creation

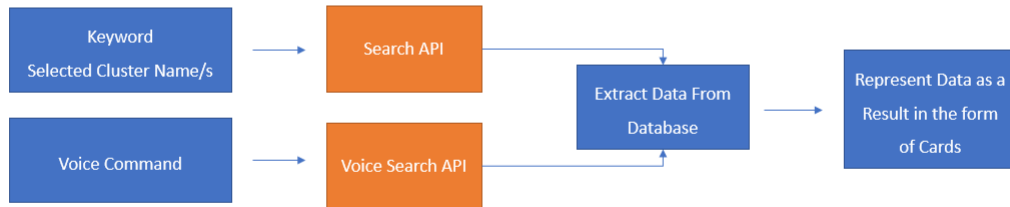


Search and Voice Search 🔍

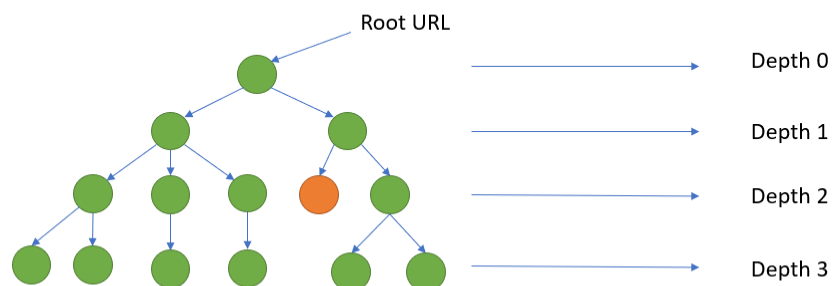
Each result card will show the following:

Title, Content, Depth, Current URL, Root URL, Type of Document.

Note that Current URL is the URL where the actual document or information was found!

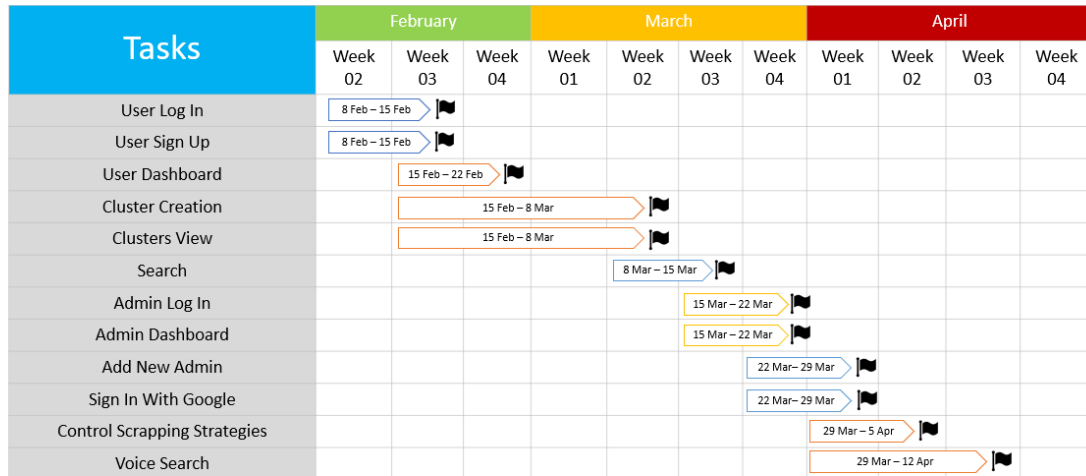


Depth (URL Tree)



Gantt Charts

Gantt Chart (Web-Application) - Neloy



Gantt Chart (Mobile-Application) - Rafid

