CS699 Project Proposal

Parismita Das
Roll No: 22M0815

Vaibhav Singh
Roll No: 22M0827

24 September 2022

Drugpedia

Contents

1 Abstract

According to one's definition, the purpose of medicine is "to cure sometimes, to relieve often, to comfort always."[1]. Every medicine has its ingredient list thoroughly checked and verified, but despite that, some detrimental ingredients may be present that the patient is unaware of, or the patient is allergic to. One can look up these ingredients online, doing so will take a lot of efforts as there aren't many platforms that give this data collectively and one might not be able to find any certified information while surfing through the net. Our aim is to implement a platform where a user can find all the details related to the prescribed medicine and its ingredient compositions and their potential roles.

2 Methodology

We will incorporate web scraping to collect the data from the two health-care platforms 'WebMD' and 'Tata 1mg' to retrieve composition summary and medicine composition respectively. And we intend to use *Beautiful Soup* to scrape the data using python. We will build the web application using *Flask*, *PostgreSQL* as database along with the frond-end, based on *HTML & CSS*.

3 Milestones

3.1 Part 1 : First Half

Module 1: Employing web scraping to display information on medications and their ingredients based on user searches. (For example: Displaying summary of 'Pan-40' medicine based on user search)
Relevant information to be shown: **Medicine usage, potential side effects and its composition.**
Website to be used for scraping: **Tata 1mg (`https://www.1mg.com/`)**

3.2 Part 2 : Second Half

Module 2: Employing web scraping to display summary of medicine ingredient (For example: Summary of the key ingredient of 'Pan-40' medication i.e. pantoprazole).
Relevant information to be shown: summary of the medicine's composition.
Website to be used for scraping: **WebMD (`https://www.webmd.com/drugs/2/index`)**

Module 3: Fix bugs and Test the platform.

4 Stack Used

- Python - using libraries such as Flask for APIs and Beautiful Soup for Web Scrapping.
- Git - via GitHub
- PostgreSQL - using psycopg2
- Makefile
- HTML and CSS

5 Update on project status - 12 October, 2022

We have initialized the database on local along with the basic models of Medicines and Ingredients which are implemented using *Flask-SqlAlchemy* which is the ORM we are using.

The github repository is initialized, including Github Actions for CI/CD. We have made the dependency graph and the Makefile for dependencies, linting, auto formatting of python code('pep8' standard). We have structured the code into Model, View, Controller, Service and Routes. The boilerplate code is ready which ensures the basic rendering of HTML pages and connection with Server.

The task of scrapping 'TATA 1mg' for medicine list and its link to the description page is complete. The medicine details are fetched successfully using BeautifulSoup4.

The task of scrapping 'WebMD' for ingredient list and their links to the description page is complete. Summary of ingredient(uses, side effects, precautions) is fetched successfully using BeautifulSoup4.

We are planning to start working on database designing and APIs for the upcoming project update.

Github: `https://github.com/parismita/Drugpedia/`

References

[1] Edward Livingston Trudeau, To Comfort Always, Independently Published, 2016, ISBN 1519051077, 9781519051073