A platform for managing research papers from multiple websites using micro services

A Project Proposal- Java LP Training

Premisha Premananthan

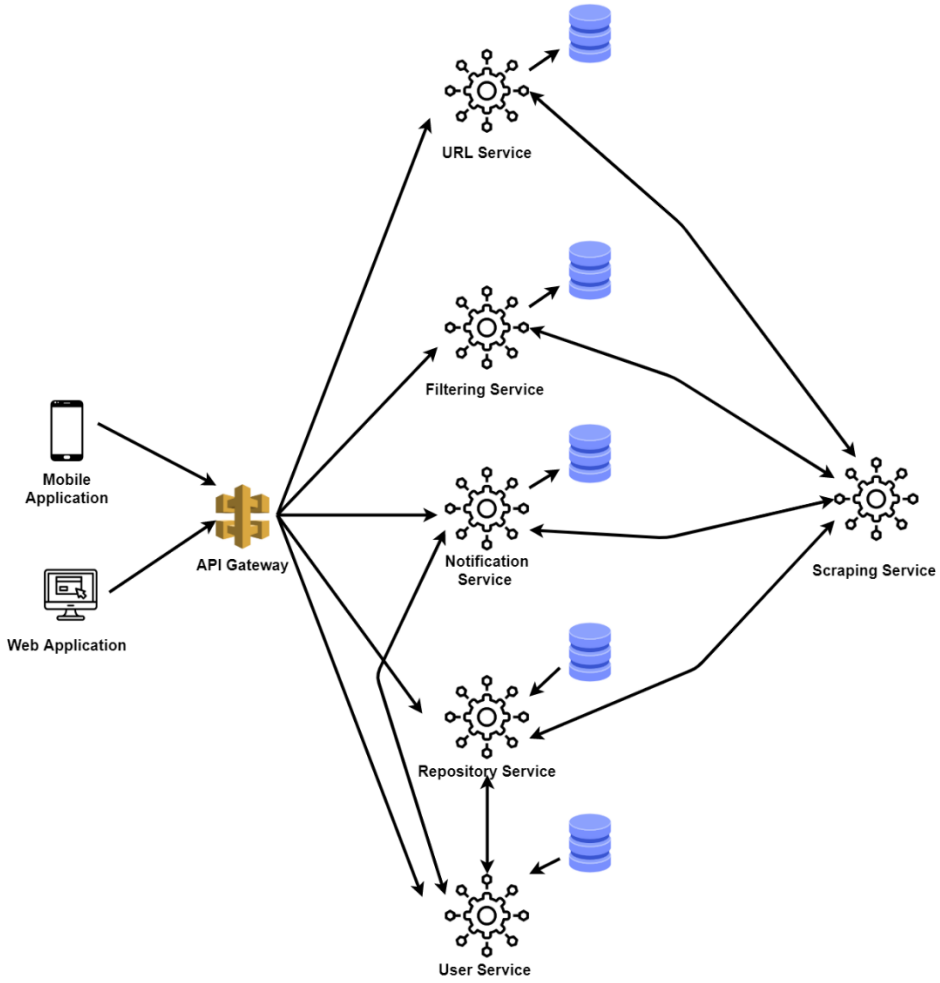
Premisha1995@gmail.com

Introduction

Researches helps to identify the solutions or problems in trending of each domain. But finding suitable gaps from larger levels of documents is the hardest part in research. Specially many researchers faced critical issues while working on finding related studies. As a researcher I personally felt how hard that works are. So in this project we are focusing to provide a platform which notify the related studies as well as it guides researchers to find related works easily.
When you took an actual scenario of literature review works, First researchers need to find related works in the selected domain as well as technological section. They initially worked with bunch of keywords. Most popular sites for research papers are IEEE explore, Science direct, Emerald, Research gateway, wiley online library. So mostly we go for each site and did the same keyword search and filter works. Mostly you will get duplicate works and each site has different trending reads mechanism so according to their users feedback you will get research papers. After that only you have to go for the read and analysis stuff. So a lot of energy and time spend repeatedly for each research paper's sites. Also if researcher wants recent updates again those previous works did in loop.
So now, you get how critical this issue and the immediate needs of this solution. So this proposal will provide a platform where users can get popular research paper sites' papers with preprocessed work like duplicate removals also users will get notification for future updates according to their user profile interests and recent searches.

Microservices are an architectural style that develops a single application as a set of small services. Each service runs in its own process. The services communicate with clients, and often each other, using lightweight protocols, often over messaging or HTTP. In this proposal, scraping is used for getting raw data from external websites. It's built on top of stdlib and has a simple abstraction around pulling in data from the raw HTML of sites it scrapes, as well as pulling in structured schema.org, and Open Graph metadata. Scraping public data is not illegal but you should make sure that you are permitted to scrape the content before scraping.

Abstract Architecture



Concerns of microservices

User Service

User can create account and through this user data only the results will be shown to the users. This service will be responsible for user data and will be used for authentication.

URL Service

User can CRUD web site URL records using this service. These web sites will be used to scrape the details according to user defined filters(genres) and search results.
As an initial step a list of web sites currently supported by the system will be presented and user can select the web sites from that list.

Filtering Service

This service will be handling a list of research paper objects and filter them according to the genres pre-defined by the user. This will maintain a database of genres which user prefer like year wise or domain wise.

Scraping Service
This service will use the URL service to get the web site URLs and then scrape the data of the home page. Then all the research papers found through given the websites will be sent to Repository Service and then to the Filtering Service. Filtered list of research papers will be sent to Notification Service.

Notification Service
This will get a list of research papers from scraping service and if there are new research papers sent them to the user as an email notification. This will maintain a database of the notified research papers.

Repository Service
This service will provide a list of all the research papers which found through respective websites, without any filtering. So that user can access information of multiple web sites at one place.

Technologies Will be Use

Backend
Spring Boot Microservices
MySQl
JSoup for web scraping: jsoup.org

Frontend
React Js

Duration : 6 weeks

Problems

Parallelly wants to learn web scraping also React js is not that much familiar.