A Project Report

on

Content Aggregator with Mood Detection

Submitted in partial fulfilment of the

requirement for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of

Dr. Michael Raj TF

Professor

Department of Computer Science and Engineering

Submitted by

18SCSE1060004 - ANKIT RANJAN
18SCSE1010043 - SHIVAM SETHI

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA, INDIA

DECEMBER - 2021

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"CONTENT AGGREGATOR WITH MOOD DETECTION"** in partial fulfilment of the requirements for the award of the Bachelor of Technology in Computer Science and Engineering submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of September, 2021 to December, 2021, under the supervision of Dr Michael Raj, Professor, Department of Computer Science and Engineering, of School of Computing Science and Engineering , Galgotias University, Greater Noida.

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

18SCSE1060004 - Ankit Ranjan

18SCSE1010043 - Shivam Sethi

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Guide

(Dr. Michael Raj TF, Professor)

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **ANKIT RANJAN (18SCSE1060004)** and **SHIVAM SETHI (18SCSE1010043)**, has been held on **20-12-2021** and his work is recommended for the award of Bachelor of Technology in Computer Science and Engineering:

Signature of Examiner(s)   **Signature of Supervisor(s)**

Signature of Project Coordinator   **Signature of Dean**

Date: 20th December, 2021

Place: Greater Noida

ACKNOWLEDGEMENT

ABSTRACT

When we have a tendency to rummage around for one thing on the internet, we get heaps of search results associate degreed it's extremely wearing to travel through all of the content thus we produce an application that kinds and displays the specified content in line with our requirements. Content soul is an entity that pulls along net or media content, applications or each from on-line sources for employ or resale. It's a way of curating content. It's attainable to mixture content manually, however in today's data-driven world, machine-controlled content aggregation, with the assistance of package tools. Content aggregation involves aggregation materials in one place, such as:

1) Blogs

2) Newsletters

3) Newspaper And Magazine Articles

4) Social Media Posts

This application can permit user to seem for the content they require in line with their interest, searches, links, saves and plenty of more. Content aggregators realize the publication' information and place it in a very place that readers will access it. This ensures that buyers find the knowledge that they really need in one place instead of many moot results from a groundwork engine.

Victimisation Python modules and libraries to form a distinct perform within the application. victimisation the database to store data in application surroundings to piece the applying by the user. As these libraries facilitate and act as tool that gathers, collects and accumulates content items equivalent to articles, social media posts, videos, images, updates and press releases from differing kinds of media shops then displays all that information via links on one single page.

CONTENTS

LIST OF FIGURES

ACRONYMS

B.Tech.	Bachelor of Technology
M.Tech.	Master of Technology
BCA	Bachelor of Computer Applications
MCA	Master of Computer Applications
B.Sc. (CS)	Bachelor of Science in Computer Science
M.Sc. (CS)	Master of Science in Computer Science
SCSE	School of Computing Science and Engineering
HTML	Hyper Text Markup Language
CSS	Cascade Style Sheet
JS	JavaScript
PY	Python
OS	Operating System
RAM	Random Access Memory

CHAPTER-1
INTRODUCTION

All people have at some purpose run into we have a tendency to site an internet site a web site that gathers content from all corners of the online onto one page, from wherever we are able to opt for those we like, and head on over thereto various blog. Not only do such blogs facilitate us save time, however they conjointly provide a platform for content generators to showcase their work. Content individuals don't pay the publisher to induce rights to their work, as they furnish them due credit and publish the work below the initial creator's name. Hence, the sole real value of running an aggregator is no matter it takes to power the technology won't to do so. As we are going to see during this project, that technology isn't all that dearly-won or advanced.

A content individual is any tool that gathers, collects and accumulates content items adore articles, social media posts, videos, images, updates and press releases from differing kinds of media shops then displays all that information via links on one single page. They greatly increase the accessibility of such pieces of content and conjointly facilitate users have a one-stop destination for all the data they need blessings of Python Content individual

Advantages of Python Aggregator:

• Pulls relevant information from everywhere the online

• Automatic update of the data

• Timely delivery of recent information

• Users save time and energy

• Python is simple to find out and use

• Powerful filtering capabilities

• extremely customizable platform

CHAPTER-2
LITERATURE

As explicit by Brown, Jones, Patterson and Casero - Ripolles the usage of written and broadcasted news has remittent among children. Mindich points out that 80% of America's population underneath thirty don't retrieve news-type newspapers on a daily basis, 70% of individuals higher than 30 do not as well. Further, a lot of Mindich supports his statement by claiming that the median age of Yankee people's info from broadcasted sources is 60, therefore hinting at an occasional worth of young people trying content. children have numerous reasons that cause such a decline: lack of time, preferences in numerous media, or content that doesn't meet the people's interests. additionally to it Casero -RipollÃľ's claims the dearth of connectedness is derived back to the missing connection to experiences and interest. In 2015 the church bench center has conducted research, observant a decline in cable news compared to a rise within the year 2013. Moreover, they need to determine that the newspaper has declined from 2013 to 2014.

Social media as a substitute for standard content

As explicit by Tsagkias, Rijke, and Weerkamp an enormous part of the content mentioned on social media can be traced back. For, instance 85% of standing updates on several sites have completely different content are related. additionally, to it, Tsagkias, Rijke, and Weerkamp argue that social media reminiscent of weblog posts and tweets are typically coupled expressly or implicitly to articles. expressly during this case means a link to the initial article is provided, wherever implicitly means that solely content of the knowledge is mentioned while not linking to it. additionally, that as already discussed the church bench center has determined that the bulk of social media users use their chosen platform as a means of gathering news information. Another study by the church bench center determined that websites and social media are the foremost common digital ways of gathering contents, supporting the thought of social media work standard news. dead it's attainable to mention that social media and news are deeply connected.

Nowadays, social media isn't solely a way of staying connected to folks however conjointly to remain connected to the direct and indirect surroundings of people. Moreover, because of the decline of cable news and social media and news websites became a vital supply of news. Thus, it is possible to state that social media may be a substitute for standard news.

Finally, it's attainable to mention that a decline in newspaper and broadcast news is observable. numerous factors play here a vital} role, nonetheless, the tactic of translating news for the youth is that the biggest issue. children don't feel connected to the news from now on because of their presentation and language. Content human is an application that collects content from completely different sources, organizes them, and puts them in one place for consumption. Content aggregator service is a very important beginning of the day. There are various publications and content online. They publish their content on multiple platforms. Now, imagine after you open 10-20 sites each day. The time you waste to realize information. info gain is everything in today's world. dead it's attainable to mention that social media and news are deeply connected. Nowadays, social media isn't solely a way of staying connected to folks however conjointly to remain connected to the direct and indirect surroundings of people. Moreover, because of the decline of cable news and social media and news websites became a vital supply of news. Thus, it is possible to state that social media may be a substitute for standard content.

A Content human makes this task easier. in a very content aggregator, you'll choose the online sites you wish to follow. Then the content aggregator collects the articles for you. And, you're simply a click away to urge info from numerous websites. A content aggregator is a combination of web crawlers and web applications. each of those technologies has its implementation in Python. which produces it easier for us.

So, Content humans can add three steps:
• It scrapes the web for the articles.
• Then it stores the article's images, links, and title.
• The hold on objects within the info is served to the consumer. The client gets information in a very nice template.

CHAPTER-3
PROBLEM FORMULATION

For this, the elemental elements steps are to make a group of rules and arithmetic logic fundamentals, python programming, and info control. produce a net-crawler-kind software package in Python. minister an inventory of net websites from that you just got to acquire and gather statistics. begin with some web websites that you suppose may have nice content material. Expand this listing as before long as you'll establish for further garage or whereas your database expands to preserve extra statistics. Use libraries like hypertext transfer protocol request sender to scrape the required content material off those web websites. Background software controls the utilization of Pyqt5. Use MySQL info to buy information needed with the help of victimization the net-crawler for indexing. Send the statistics over from the database to the WP and your content material collector is right to go. Also, we have a tendency to use mil options right here to line the temper and hobby and plenty of various factors withinside the software package to form it further pleasant for the patron to interact. Content aggregation isn't something but compilation statistics on any subject material for one or extra associated key phrases and business enterprise them in an exceedingly single platform, net site, or blog. It is machine-controlled (RSS feeds) and collects statistics utterly wholly} totally on key phrases. we would be scraping the net website for getting articles. we have a tendency to-Scraping technique to extract statistics from the websites. We extract important statistics from the websites. during this case, we are able to be extracting the articles from the internet site. we are able to use the stunning soup and requests module. These libraries are the bs4 and requests and modules are used for net crawling. once quoting the distinctive text, transfer your insights as well, i.e., does one compromise or trouble their view, or become any statistics neglected out.

CHAPTER-4
PROPOSED WORK

The most important step in creating a content link website is knowing exactly what content you will bring and how often you will review it. There are several types of websites that cover content that fall into different categories: news aggregates, review editors (Metacritic, Rotten Tomatoes), social media links (Hootsuite, Flock), survey links, and more. Only after deciding what type of aggregator do you want to build and what value do you want to bring when you can decide which tools to use.

When you create a content link, it is also important to come up with the right name for your website. A domain name is one of the first things users notice and can affect the user's first appearance. We recommend that you keep it concise, unique, accurate, and suitable for SEO and make it easy to spell and remember. An aggregator is a website with a lot of data. Users should be given a solid filter so they can get the information they need in the shortest possible time. Most successful compilers also content by classifying data into categories.
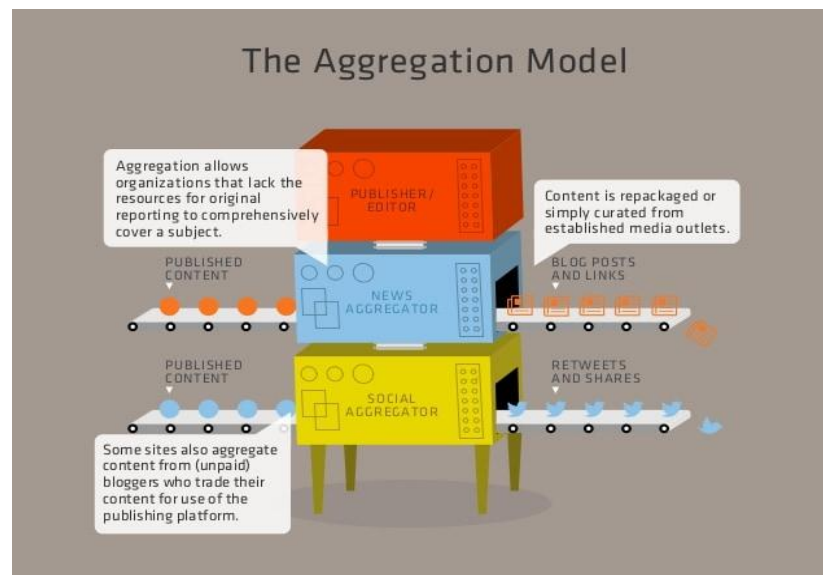

Fig.1. Aggregation Process

CHAPTER-5
TOOLS AND MODULES USED

1. OS- Windows

2. Python libraries

3. Databases and Storage

4. PyCharm (IDE)

5. GIT

6. HTML/CSS

7. Django framework or Bootstrap

Hardware:

1. Desktop Device:
    a.) 1GB Memory
    b.) 4GB RAM or more
    c.) Windows 8 or above

2. Active internet connection

Implementation and Description of Modules

Home-Page - This module used to renders all the objects of Python Content, Programming Content, Hiring Content, and totally different categories. To render the objects of the model we've an inclination to are pattern the all() technique that returns all the objects of the model. once that, we tend to are passing all this information. mistreatment HTML, CSS & JS for the homepage and to connect pages.

Django - Django settings for content somebody project.

Generated by 'django-admin' mistreatment Django 3.1.2.Django - apscheduler is also a pleasant choice for quickly and easily adding basic designing choices to your Django applications with stripped dependencies ANd very little or no additional configuration. planl|the perfect|the best} use case all told likelihood involves running a couple of of tasks on a group execution schedule.

albeit history are hold on among the database, if any errors occur, it's an honest plan to log them for debugging. you'll add some basic work settings to your application.

Bootstrap - Bootstrapping describes a situation within that an entrepreneur starts a company with little capital, wanting forward to money excluding outside investments. a private is claimed to be bootstrapping once they conceive to found and build a corporation from personal finances or the operative revenues of the new company.

Manage - Custom commands utilize the manage.py file to run your code. When you run manage.py, Django registers any module that's within a management/commands/ directory as a usable command.

Code:

```python
#!/usr/bin/env python
"""Django's command-line utility for administrative tasks."""
import os
import sys


def main():
    """Run administrative tasks."""
    os.environ.setdefault(
        "DJANGO_SETTINGS_MODULE", "aggregator.settings"
    )
    try:
        from django.core.management import execute_from_command_line
    except ImportError as exc:
        raise ImportError(

        ) from exc
    execute_from_command_line(sys.argv)


if __name__ == "__main__":
    main()
```

manage.py in Content Aggregator

Feed Parser - To retrieve RSS feeds from podcasts in your application and analyze them, you will learn how to use the feed-parser library. in a highly parsed feed with feed-parser, you may also have access to a special attribute called Entries. this allows it to iterate through each of the components of the power supply. Once you've incorporated your information into the podcast episodes, you're good to go. Entries to view help for each podcast episode on the feed and check whether or not it exists in your database.

```python
import os
import sys
import apscheduler
import django
from django import get_wsgi_application
import feedparser
import bs4
import request
```

Libraries in content aggregator
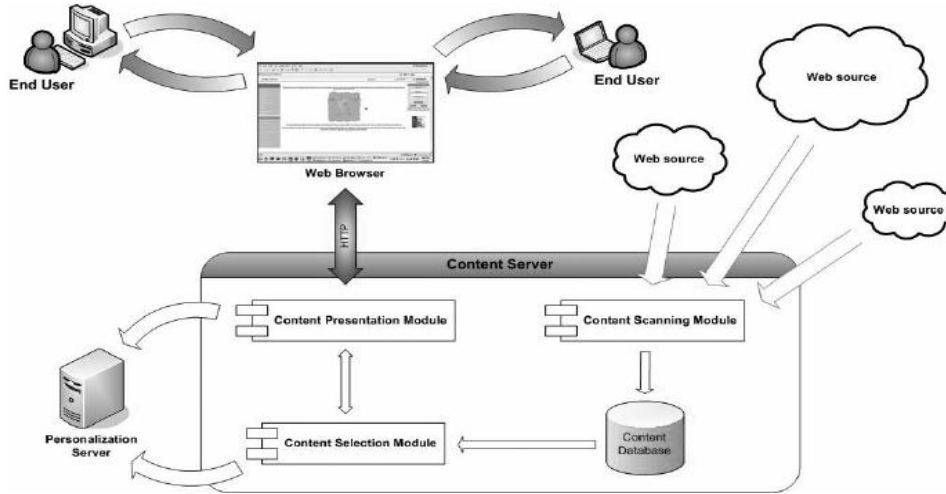
CHAPTER-6
PROJECT ARCHITECTURE

1. Architectural

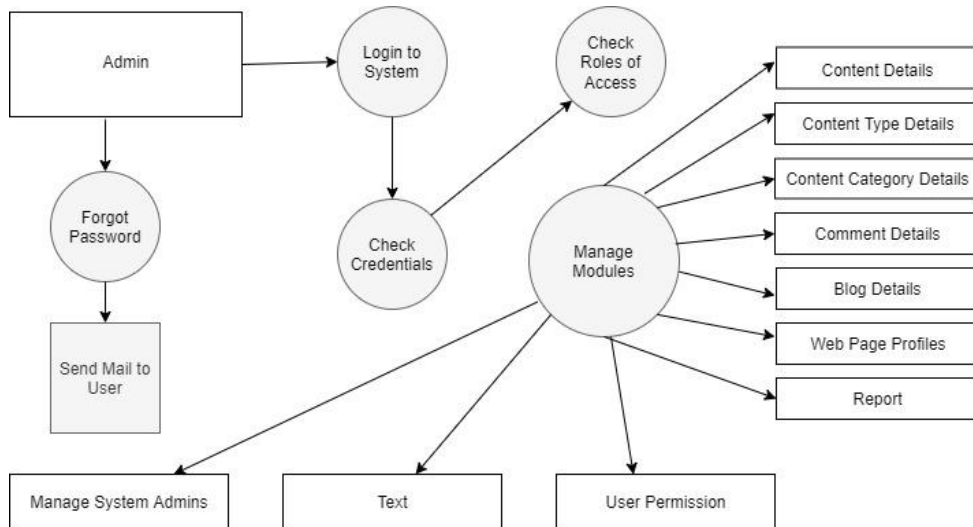Fig. 2. Architectural Diagram

2. DFD of Content Aggregator

Fig. 3. DFD Diagram

CHAPTER-7
PROJECT LIFE CYCLE
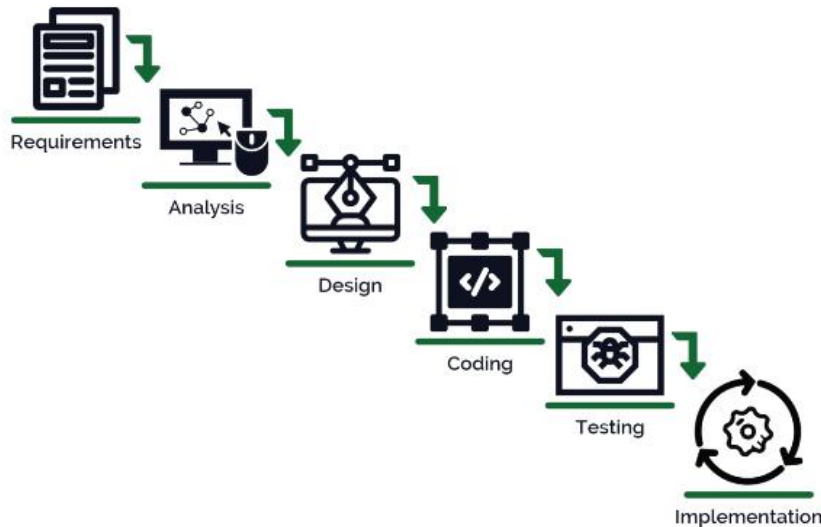
Project Life Cycle-

Fig. 4. Project Life cycle Diagram

Database Design in the Project

Start with several sites that you just assume will have the foremost effective content. Expand this list once you'll organize for tons of storage or once your information expands to hold additional information. Also, check whether or not or not the net sites you've got listed support API as this allows the online crawler to scrape information. Record the parsed data into a information on the MongoDB platform or Django platform as per your convenience and experience. Use a MySQL database to store information required by the web-crawler for indexing.

Set-up a web-page on Word-press and elegance its layout sources, keywords and build SEO name for that web site. The Django -apscheduler package can store your jobs at intervals the database, and it'll put together store bushed and unsuccessful job runs. Having this history is good for you as a result of the developer or website admin as a result of you'll monitor whether or not or not any tasks fail. The posts that you {just} just have scraped and assign relevant filters on the content.
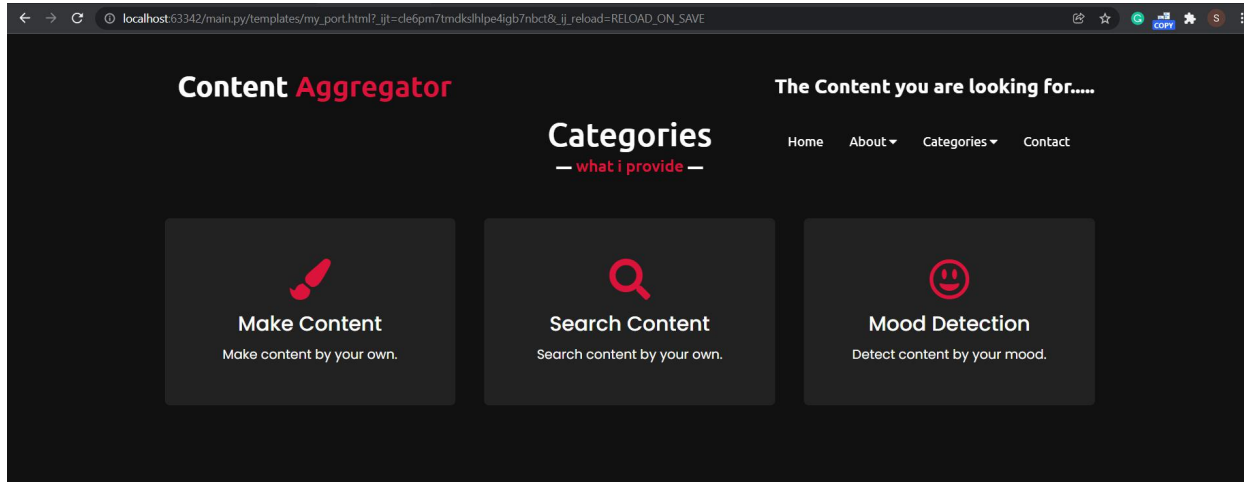
CHAPTER-8
PROJECT ARCHITECTURE

The most necessary a vicinity of the challenge was to form a system that meets the strict hardware limitations. These constraints, additionally as short contest duration, forced us of America to accept machine learning algorithms (such as fast Text) instead of refined models. we've an inclination to didn't have enough time to conduct all the experiments we have a tendency to needed or to shine the final word thus there' still area for improvement, though we've implemented some modifications once the highest of the contest.

This a part of the report is used so on dialogue that aspect could be modified, enforced or improved thus as manufacture Associate in Nursing improved version of the news aggregation and validation system. This section will discuss accomplishable enhancements for each instance of the system the news protagonist algorithm, the corpus and conjointly the net web site itself. Especially, kind and dislike perform is also viewed as an additional approach of substantiating news. Here changed version, that contains fake and real labels instead of like and dislike labels, which allows users to vote manually whether or not or not an editorial is real or not. along side the already used algorithm, such an approach might cause higher and extra trustworthy results for the user. This, is owing to the actual fact that the parsed articles from the fake news websites contain political information, where the parsed articles from real news websites provided varied categories of articles adore sport and tv shows. this might need had Associate in Nursing influence on the coaching job of the classifier and if corrected likely cause and improved outcome.
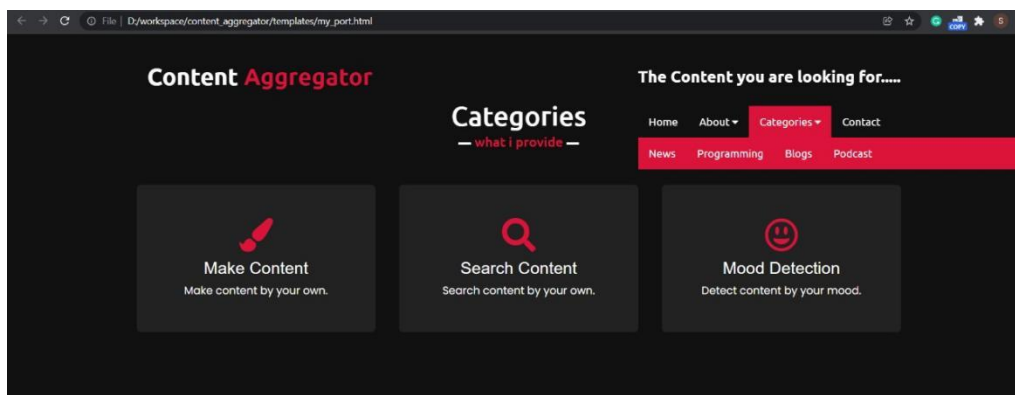
CHAPTER-9
RESULT

Fig. 5. Homepage of Content Aggregator

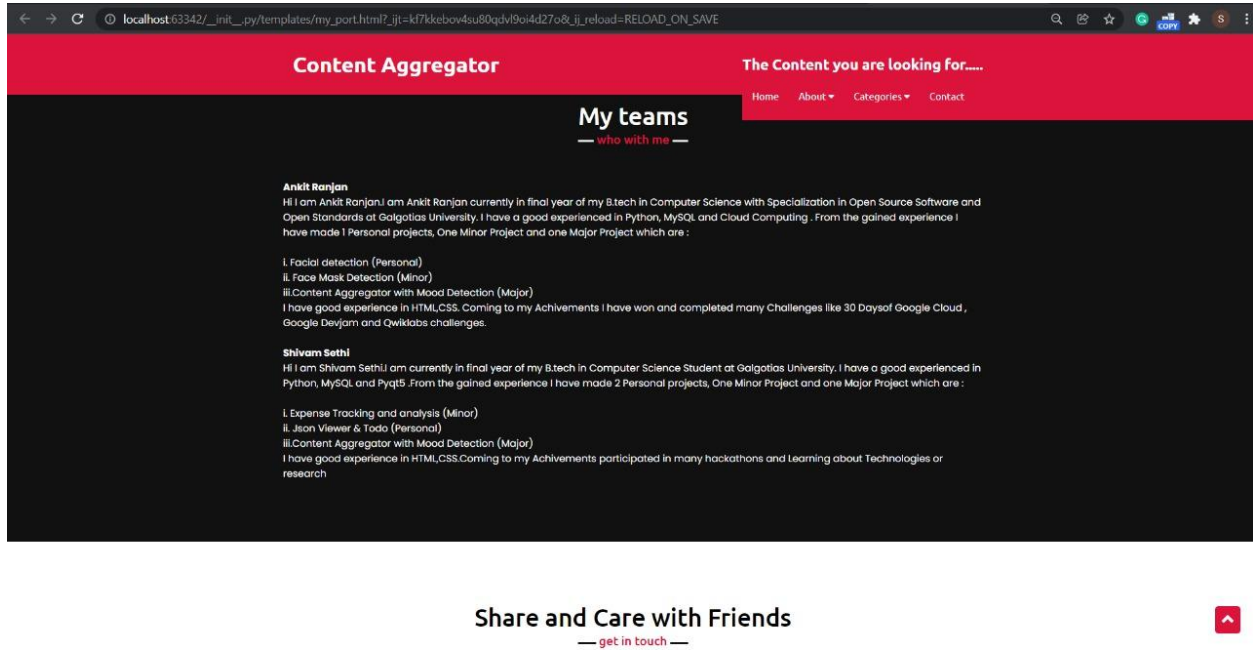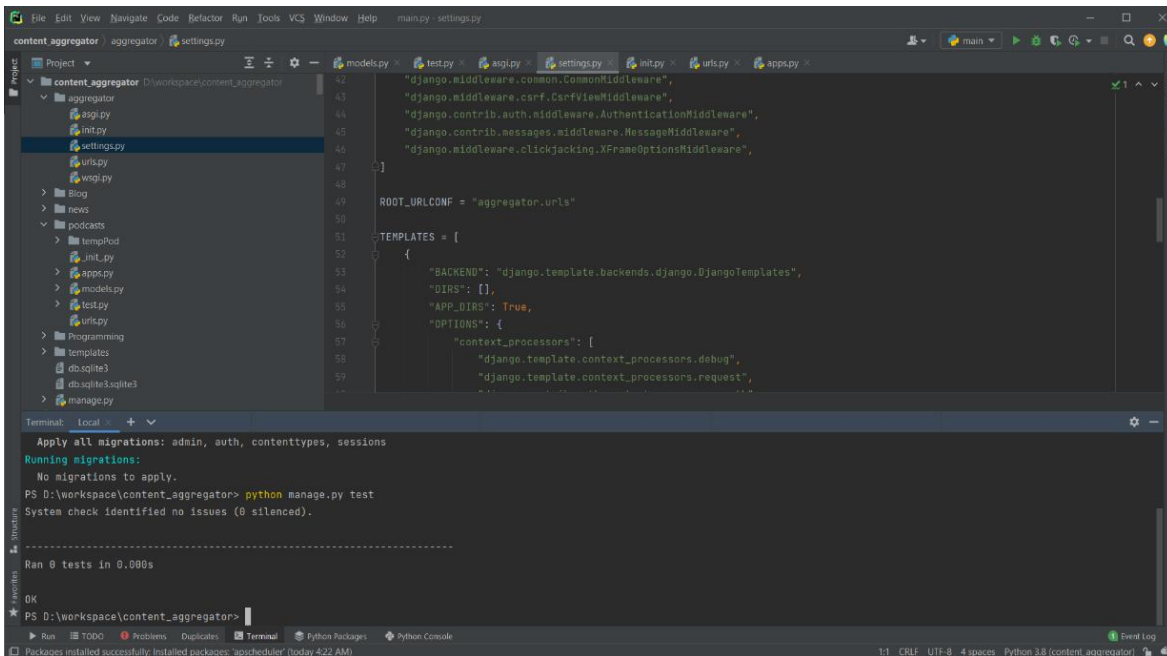Fig. 6 Categories of Content Aggregator

Fig. 7. About Section of Content Aggregator

Modules used in Content Aggregator

REFERENCES

1. Miles, Alisha (2009). "RIP RSS: Reviving Innovative Programs through Really Savvy Services". Journal of Hospital Librarianship.

2.   Jim (2007-01-01). "RSS: A Brief Introduction". The Journal of Manual & Manipulative Therapy. **15** (1): 57–58.

3. Hammersley, Ben (2005). Developing Feeds with RSS and Atom. Sebastopol: O'Reilly Media, Inc.

4. "Google Reader is dead but the race to replace the RSS feed is very alive". Digital Trends. July 2013. Retrieved 2015-12-21.

5. ^ "RSS Usage Statistics". trends.builtwith.com. Retrieved 21 December 2015.

6.   Chowdhury, Monica (2006). "News aggregator services: user expectations and experience". Online Information Review. **30** (2): 100–115. doi

7. www.Awasu.com. Retrieved 27 October 2017.

8. ^ "Google News and newspaper publishers: allies or enemies?". Editorsweblog.org. World Editors Forum. Retrieved 31 March 2009.

9. ^ Luscombe, Belinda (19 March 2009). "Arianna Huffington: The Web's New Oracle". Time. Time Inc. Archived from the original on March 21, 2009. Retrieved 30 March 2009. The Huffington Post was to have three basic functions: blog, news aggregator with an attitude and place for pre-moderated comments.

10. ^ Hansell, Saul (24 September 2002). "All the news Google algorithms say is fit to print". The New York Times. Retrieved 20 January 2014.