

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

MAY - 2022

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 January, 2022 to May, 2022, 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 **12-05-2022** 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: 12th May, 2022

Place: Greater Noida

ACKNOWLEDGEMENT

We would like to express our gratitude to our supervisor/reviewer, Dr. Michael Raj TF (Professor), who guided us throughout this project. His enthusiasm, knowledge and exacting attention to detail have been an inspiration and kept our work on track. We would also like to thank our friends and family who supported us and offered deep insight into the study. The generosity and expertise of one and all have improved this study in innumerable ways and saved us from many errors; those that inevitably remain are entirely our own responsibility.

ABSTRACT

When we have a tendency to rummage around for one thing on the internet, we get heaps of search results associate degraded 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

	Page No.
Candidates Declaration	I
Acknowledgement	II
Abstract	III
Contents	IV
List of Table	V
List of Figures	VI
Acronyms	VII
Chapter 1 Introduction	1
Chapter 2 Literature Survey	2
Chapter 3 Problem Formulation	4
Chapter 4 Proposed Work	5
Chapter 5 Tools and Modules Used	6
Chapter 6 Project Architecture	9
Chapter 7 Project Lifecycle	10
Chapter 8 Result	11
Chapter 9 Conclusion and Future Scope	12
Reference	14

LIST OF FIGURES

S.No.	Title	Page No.
1.	AGGREGATION PROCESS	5
2.	ARCHITECTURAL DIAGRAM	9
3.	DFD DIGRAM	9
4.	PROJECT LIFE CYCLE DIAGRAM	10
5.	HOMEPAGE OF CONTENT AGGREGATOR	12
6.	CATEGORIES OF CONTENT AGGREGATOR	12
7.	ABOUT OF CONTENT AGGREGATOR	13

ACRONYMS

B.Tech.	Bachelor of Technology
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 use of the written and visible content has remittent among children. They point 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 by claiming that the most of the young people's info from sources is 70%, therefore hints at an occasional worth of young people trying content. children have numerous reasons that cause such as lack of time, preferences in media, and other content that doesn't seems interesting to the people. Additionally, to it Casero shows the dearth of connectedness is glean back to the lost connections to various experiences and people's interest. In 2015 the church bench conducted research, observant a decline in cable content compared to a rise in the year 2013. Also, they need to determine that the newspaper and other standard content has declined few years ago.

Social media as a replacement for standard content

As explicit by Tsagkias, Rijke, and Weerkamp an enormous parts of the sources mentioned on social media can be track easily. Instance 80% of standing improves on several sites have completely different content are related. additionally, to it, they all argue that social media reminiscent of weblog posts and tweets are coupled expressly or indirect to articles. expressly during thiscases proves a link to initial sources is provided, wherever it indirectly meaning to solely content of the knowledge while not linking. Additionally, that already discussed the church bench center has determined the bulk of social media users use the platform as means of gathering various information. The study by the church bench centre determined that websites blogs and social media posts and tweets are the foremost common digital ways of finding contents, supporting the thought of social media work standard. It's attainable mention that social media content and news are very much connected.

A social media isn't solely way to connect the folks however conjointly remain connect direct and indirect surroundings of users. Because of the decline of cable contents and social media and several websites became a vital supply. Thus, it is possible for the social media may be a change for standard content.

Finally, it's attainable to mention that a reduce in newspaper and broadcast content is observable. numerous factors which plays a vital role, nonetheless, tactic to translate the content for the youth people is that the biggest problem. 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 aggregation services is new era of beginning. There are various publications and content online. They publish their content on many different platforms. Now, imagine after you open 5-10 websites each day costs a lot of time to waste. But information knowledge is everything in today's society. It's attainable to mention that social media content and news sites are connected. Social media isn't solely a way to connected to the folks however conjointly to remain connected to the surroundings of people directly and indirectly. Because of the reduce of cable content increase in social media and news websites became a vital supply of content to the people. Thus, it is possible to say that social media content may be a replacement for standard content.

A Content human makes this work easy in very easy way. Content aggregator, you'll choose the online sites you wish to follow or reach. The content aggregations collect the articles, post, blogs and many more for you simply a one clicks to urge info from different websites. A content aggregator is combination of web crawlers and web applications. each of those technologies has its implementation in Python which produces the content from different places.

Content humans can add three steps:

- It scrapes the content from different websites.
- Then it stores the article's images, links, and title.
- The hold on objects within the info is served the consumer. The client and user gets information in enhanced 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 maximum critical step in developing a content material hyperlink internet site is understanding precisely what content material you'll convey and the way regularly you'll assessment it. There are numerous sorts of web sites that cowl content material that fall into distinct categories: information aggregates, assessment editors (Metacritic, Rotten Tomatoes), social media links (Hootsuite, Flock), survey links, and more. Only after identifying what kind of aggregator do you need to construct and what fee do you need to convey whilst you could determine which equipment to use. When you create a content material hyperlink, it's also critical to provide you with the proper call on your internet site. A area call is one of the first matters customers word and might have an effect on the user's first appearance. We propose which you hold it concise, unique, accurate, and appropriate for search engine optimization and make it clean to spell and remember. An aggregator is a internet site with a number of statistics. Users have to accept a strong clear out out for you to get the facts they want withinside the shortest feasible time. Most a hit compilers additionally content material via way of means of classifying statistics into categories.

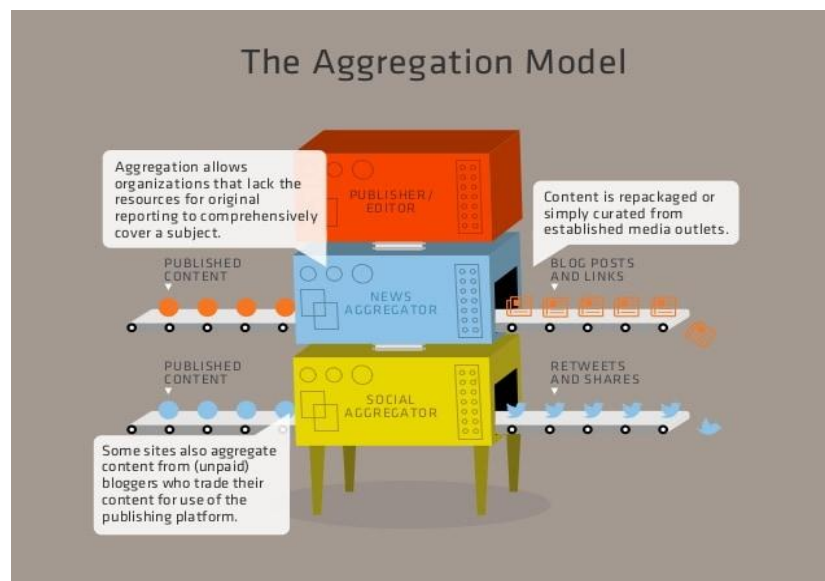


Fig.1. Aggregation Process

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. plan|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 state of affairs inside that an entrepreneur begins offevolved a enterprise with little capital, looking ahead to cash except for out of doors investments. a non-public is said to be bootstrapping after they conceive to located and construct a company from non-public price range or the operative sales of the brand new enterprise.

Manage - Custom instructions make use of the manage.py document to run your code. When you run manage.py, Django registers any module that's inside a management/instructions/ listing as a usable command.

Code:

```
#!/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(
            "Couldn't import Django. Are you sure it's installed and "
            "available on your PYTHONPATH environment variable? Did you "
            "forget to activate a virtual environment?"
        ) from exc
    execute_from_command_line(sys.argv)

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

manage.py in Content Aggregator

Feed Parser - To recover RSS channels from webcasts to your product and inspect them, you'll find ways of utilizing the channel parser library. in a fairly parsed feed with feed-parser, you could furthermore have get section to a novel trademark known as Entries. this allows in it to emphasize through everything about added substances of the energy supply. Whenever you have incorporated your records into the webcast episodes, you are consistent with go. Sections to see help for each digital recording episode at the feed and investigate whether or not or presently no longer it exists to your information base.

```
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

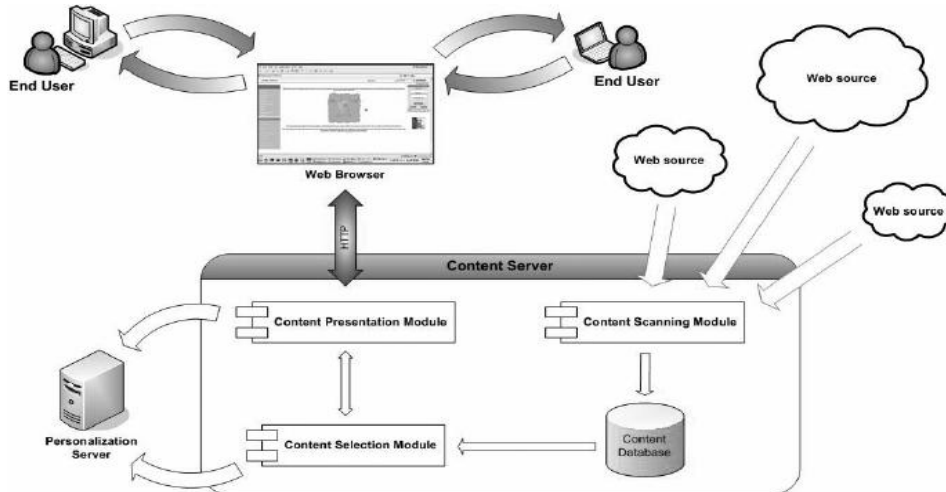


Fig. 2. Architectural Diagram

2. DFD of Content Aggregator

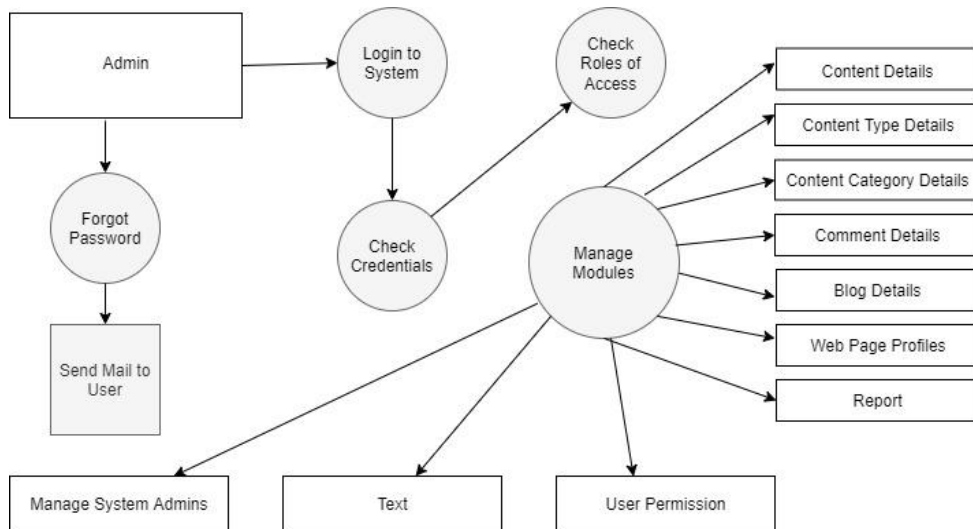


Fig. 3. DFD Diagram

CHAPTER-7 PROJECT LIFE CYCLE

Project Life Cycle-

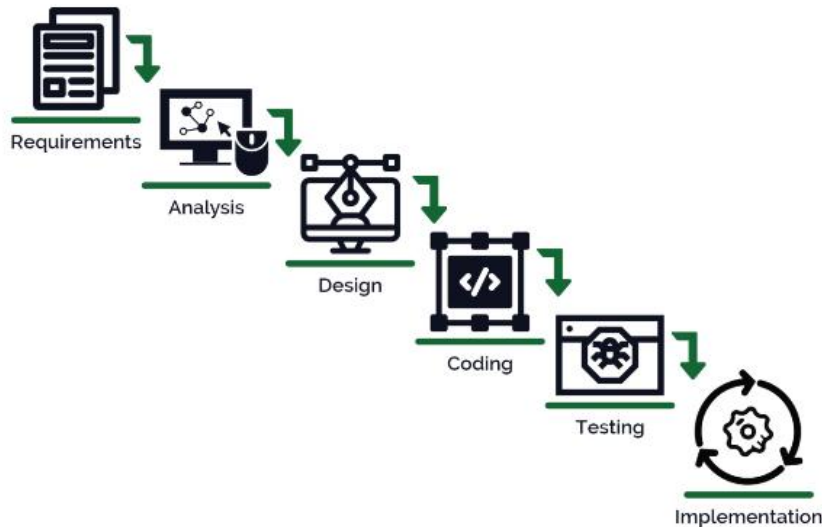


Fig. 4. Project Life cycle Diagram

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 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 deal can save your jobs at periods the database, and it will prepare save bushed and unsuccessful process runs. Having this records is right for you due to the developer or internet site admin due to you may reveal whether or not or now no longer or now no longer any duties fail. The posts that you simply} simply have scraped and assign applicable filters at the content.

CHAPTER-8

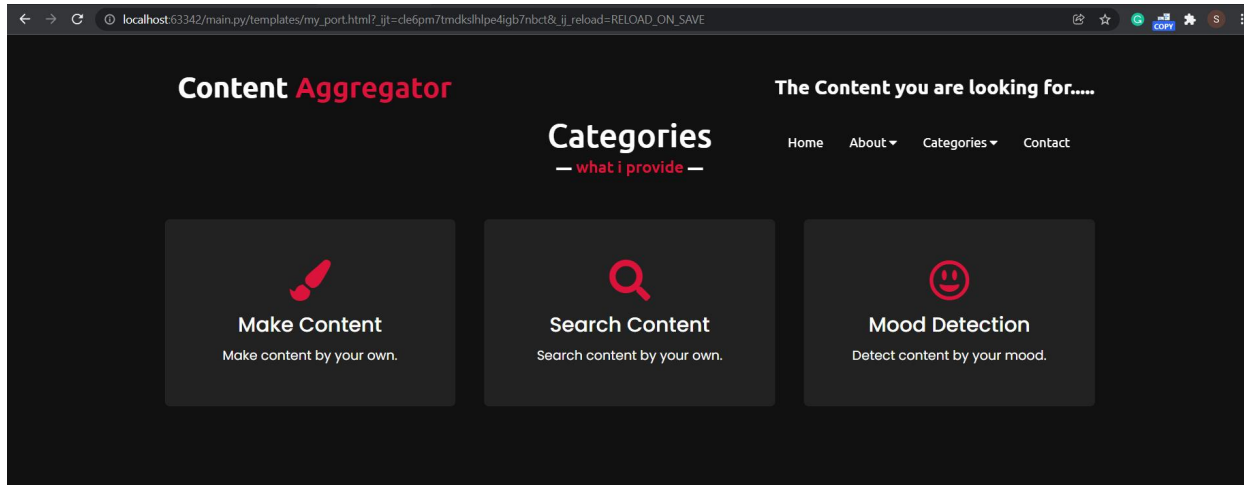
PROJECT ARCHITECTURE

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 part of the file is used so on define that element may be modified, enforced or stepped forward as a consequence as manufacture Associate in Machine stepped forward model of the information aggregation and validation gadget. This segment will manage the accomplishable upgrades for every example of the gadget the information technology algorithm, the corpus and conjointly the internet website online itself. Especially, type and dislike carry out is likewise regarded as a further technique of substantiating information. Here modified model, that incorporates faux and actual labels as opposed to like and dislike labels, which permits customers to vote manually whether or not or now no longer or now no longer an article is actual or now no longer. alongside facet the already used algorithm, such an technique would possibly motive better and additional straightforward consequences for the user. This, is attributable to the real reality that the parsed articles from the faux information web sites incorporate political information, wherein the parsed articles from actual information web sites furnished various classes of articles adore recreation and television shows. this could want had Associate in Nursing impact at the training activity of the classifier and if corrected in all likelihood motive and stepped forward outcome.

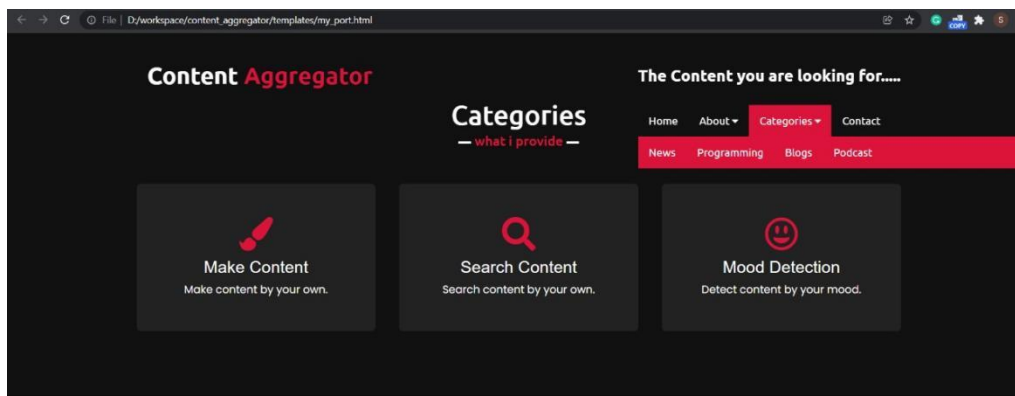
CHAPTER-9

RESULT



Our Reach
— what i know —

Fig. 5. Homepage of Content Aggregator



Our Reach
— what i know —

Fig. 6 Categories of Content Aggregator

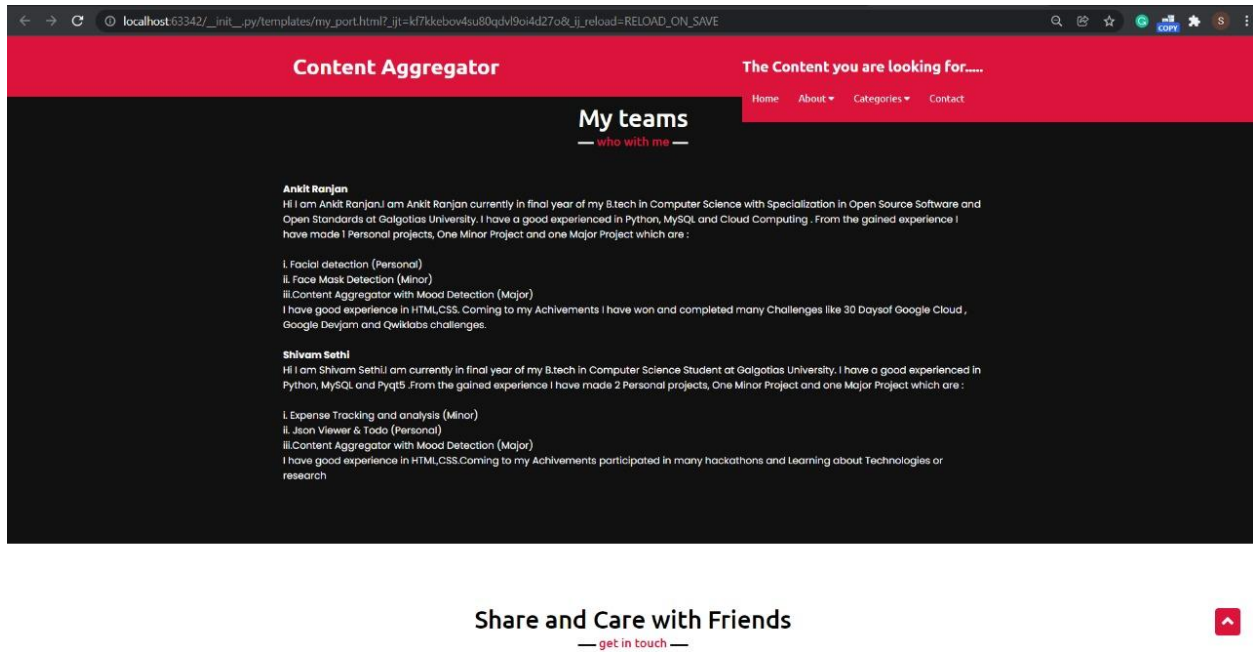
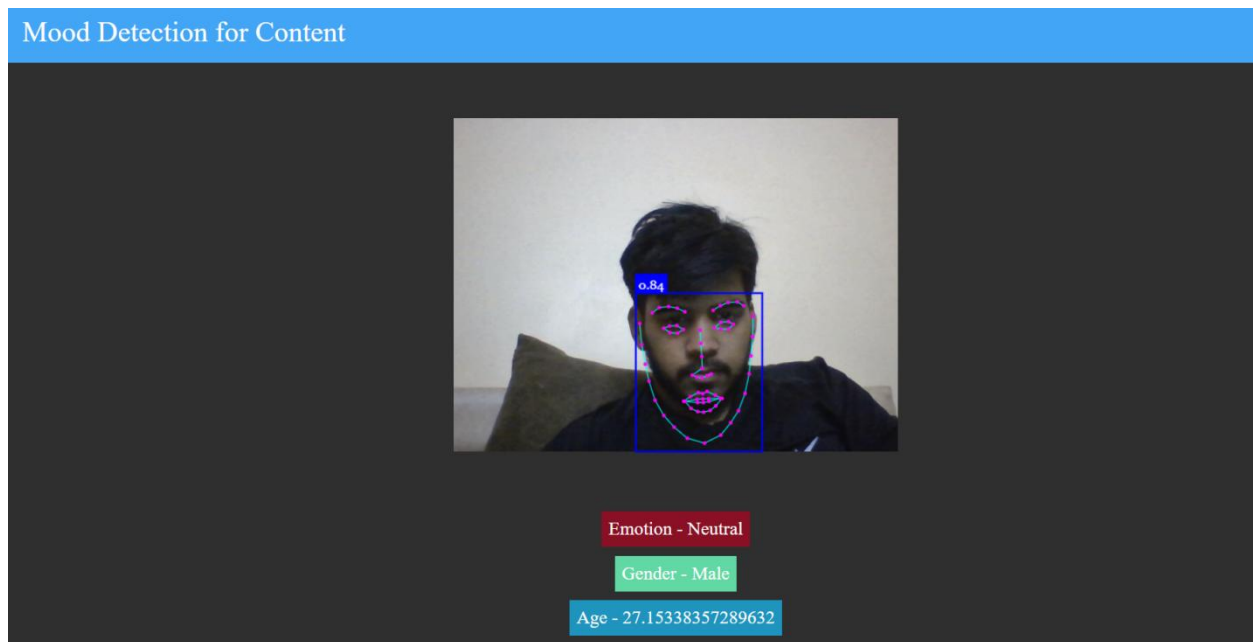


Fig. 7. About Section of Content Aggregator



Mood Detection in Content Aggregator

REFERENCES

1. Milles, Alishha (2009). "RIP RSS: Reviving Innovative Programs via 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. Hammerssley, Benn (2005). *Developing Feeds with RSS and Atom*. Sebastopol: O'Reilly Media, Inc.
4. Google Reader is useless however the race to update the RSS feed may be very alive". *Digital Trends*. July 2013. Retrieved 2015-12-21.
5. RSS Usage Statistics - trends.builtwith.com. Retrieved 21 December 2015.
6. Chowdhury, Sudatta; Landoni, Monica (2006). "News person services: user expectations and experience". *on-line info Review*. thirty (2): 100– 115
7. www.Awasu.com. Retrieved twenty seven Oct 2017.
8. Editorsweblog.org. World Editors Forum. Retrieved thirty one March 2009.
9. D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D Spyropoulos. net usage mining as a tool for personalization: a survey. *User Modeling and User-AdaptedInteraction*, 13(4):311–372, 2003
10. Hansell, Saul (24 Sept 2002). . *The ny Times*. Retrieved twenty Gregorian

PUBLICATION

2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N-22)



Ankit Ranjan <ankitranjann@gmail.com>

Notification 4th IEEE ICAC3N-22 & Registration: Paper ID 170

1 message

Microsoft CMT <email@mssr-cmt.org>
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Ankit Ranjan <ANKITRANJANN@gmail.com>

Sat, May 7, 2022 at 7:40 AM

Dear Author,

Greetings from Galgotias College of Engineering and Technology!!!

On behalf of the 4th ICAC3N-22 Program Committee, we are delighted to inform you that the submission of "Paper ID- 170 " titled " Content Aggregator with Mood Detection " has been accepted for presentation at the ICAC3N- 22 and will be sent for the submission in the conference proceedings to be published by the IEEE.

Please complete your registration by clicking on the following Link: <https://forms.gle/8acy23i3UbtwLkFXA> on or before 20 May 2022.

Note:

1. All figures and equations in the paper must be clear.
2. Final camera ready copy must be strictly in IEEE format.
3. Minimum paper length should be 5 pages.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences.
5. Violation of any of the above point may lead to rejection of your paper at any stage of publication.
6. Registration fee once paid will be non refundable.

If you have any query regarding registration process or face any problem in making online payment, you can Contact @ 8168268768 (Call) / 9467482983 (Whatsapp/UPI) or write us at icac3n.22@gmail.com.

Regards:
Organizing committee
ICAC3N - 22

Download the CMT app to access submissions and reviews on the move and receive notifications:
<https://apps.apple.com/us/app/conference-management-toolkit/id1532488001>
<https://play.google.com/store/apps/details?id=com.microsoft.research.cmt>

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our [Privacy Statement](#).

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Content Aggregator with Mood Detection

ORIGINALITY REPORT

11 %	11 %	3 %	6 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.slideshare.net Internet Source	2 %
2	en.wikipedia.org Internet Source	2 %
3	www.coursehero.com Internet Source	2 %
4	annalsofrscb.ro Internet Source	1 %
5	www.sportademics.com Internet Source	1 %
6	www.jetir.org Internet Source	1 %
7	www.e-ijaet.org Internet Source	<1 %
8	www.ijert.org Internet Source	<1 %
9	Submitted to Malaviya National Institute of Technology Student Paper	<1 %

10

VNUA
Publication

<1 %

11

Submitted to Anglia Ruskin University
Student Paper

<1 %

12

github.com
Internet Source

<1 %

13

digital.library.unt.edu
Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off