Automated News Reporting Agency

B.S. (CS) Final year project proposal

Yusra Masood (18B-093-CS)
Rushan Rafiq (18B-087-CS)
Yumna Asim (18B-047-CS)
Shiza Khan (18B-130-CS)

Batch: 2018
Date: 31st-August-2021

DEPARTMENT OF COMPUTER SCIENCE
USMAN INSTITUTE OF TECHNOLOGY

Usman Institute of Technology
Department Of Computer Science

FINAL PROJECT APPROVAL FORM

The Head of Department,
Computer Science Department,
Karachi.

Date: 31st-Ausust-2021
Batch: 2018

Subject: **Bachelor of Science in Computer Science/Software Engineering Final Year Project**

Respect Sir,

We, the below listed students of final Year BS (CS) class, desire to undertake work on the following project.

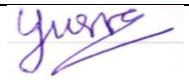
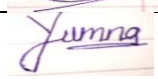
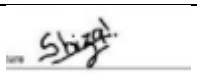
Automated News Reporting Agency

We request you to kindly grant approval for undertaking the work on the above-cited project. I abide by all terms and conditions mentioned below.

1. I have selected this project on my own.
2. I have no objection working under the supervision of male/female supervisor, or if my project work is evaluated by male/female externals.
3. I am sure I can complete this project till ___June 2022_____.
4. I am eager to work under the supervision of advisor assigned to this project.
5. I understand that FYP committee can modify the scope of the project as and when required.
6. I know that if do not appear in regular project progress presentations/milestones my project will be disqualified.
7. I know that if I do not appear in mid project presentation, whenever it is scheduled, I will not be eligible for final project viva
8. I fully understand that "***cheating***"* may lead to cancelation of my project.
9. I understand that the decision of the FYP evaluation committee, for all issues, would be final, and no objections will be accepted.
10. I have no objection presenting my project to external or internal examiner assigned by the Head of the Department.
11. Project and Product deliverables at the time of submission of final year project every group is responsible to submit complete running system along with printed reports, source code, hardware (if any) etc. to the project coordinator.
12. It would be the responsibility of Project coordinator to keep record of all projects in a system (in running form) as it would help to continue next project in continuation, depends upon the scope and application of project.
13. Proper dressing and way of presentation should be in English during proposal defend session, milestones and final presentations.
14. Marking of milestones and final presentation should be based on individual evaluation of each faculty members and marks would be granted during session.
15. When we go for proposal defends session a list of all previous projects with their brief introduction must be available during session for our reference. (Introduction, Scope of project, tools and technology and batch must be available).
16. I understand that it is my responsibility to update my advisor and FYP committee members with the status of my project and submit reports on time.
    - Copying code from any resources

- Using off the shelf components without prior permission
- Outsourcing your project
- Hire a resource for the completing the FYP code or any part of the project.

Yours sincerely,

S.No.	Roll No.	Name	Email	Signature
1.	18B-093-CS	Yusra Masood	ymasood@students.uit.edu	
2.	18B-087-CS	Rushan Rafiq	rushan@students.uit.edu	
3.	18B-047-CS	Yumna Asim	yasim@students.uit.edu	
4.	18B-130-CS	Shiza Khan	shkhan@students.uit.edu	

Contents

PROJECT EXECUTIVE SUMMARY **1**

PROBLEM STATEMENT 1
PROPOSED SOLUTION 1
VALUE 1

PROJECT OVERVIEW **2**

PROJECT OBJECTIVES **3**

PROJECT SCOPE **3**

IN SCOPE: 3
OUT OF SCOPE: 3
DELIVERABLES PRODUCED 4
PROJECT ESTIMATED EFFORT/COST/DURATION 4
ESTIMATED EFFORT HOURS: 5
ESTIMATED DURATION: 6

PROJECT ASSUMPTIONS **6**

PROJECT RISKS **7**

PRODUCT RISKS 7

PROJECT APPROACH **8**

TOOLS AND TECHNOLOGIES 8

OUTSTANDING ISSUES **8**

EXPECTED FINAL PRODUCT **8**

SIMILAR PRODUCTS AVAILABLE **10**

LIST DOWN SIMILAR PRODUCTS AVAILABLE ONLINE WITH URL **10**

PROJECT APPROVALS: **11**

Project executive summary

Problem Statement

Nowadays the world is constantly escalating in terms of technology and this has been true for most of the fields including journalism and news media. Production and consumption of news media have never been this easy but this comes at the expense of some disadvantages. With the increasing competition in the news media field, we're sometimes delivered with unauthentic, fake or incomplete news whose sole purpose is to increase viewership or money. To solve for this problem we need an authentic system which could deliver the correct and complete news that's worth of the user watching it.

Proposed Solution

We therefore plan on building automated news reporting agency system which can scrap data from different news reporting sources, filter those data for authentic content only, aggregate, summarize and synthesize it and broadcast the final accurate news (would be broadcasted by any animated character or so). Moreover our system would have the ability to view/search for any older/past news as well (a facility which is not provided by the traditional news reporting channels.)

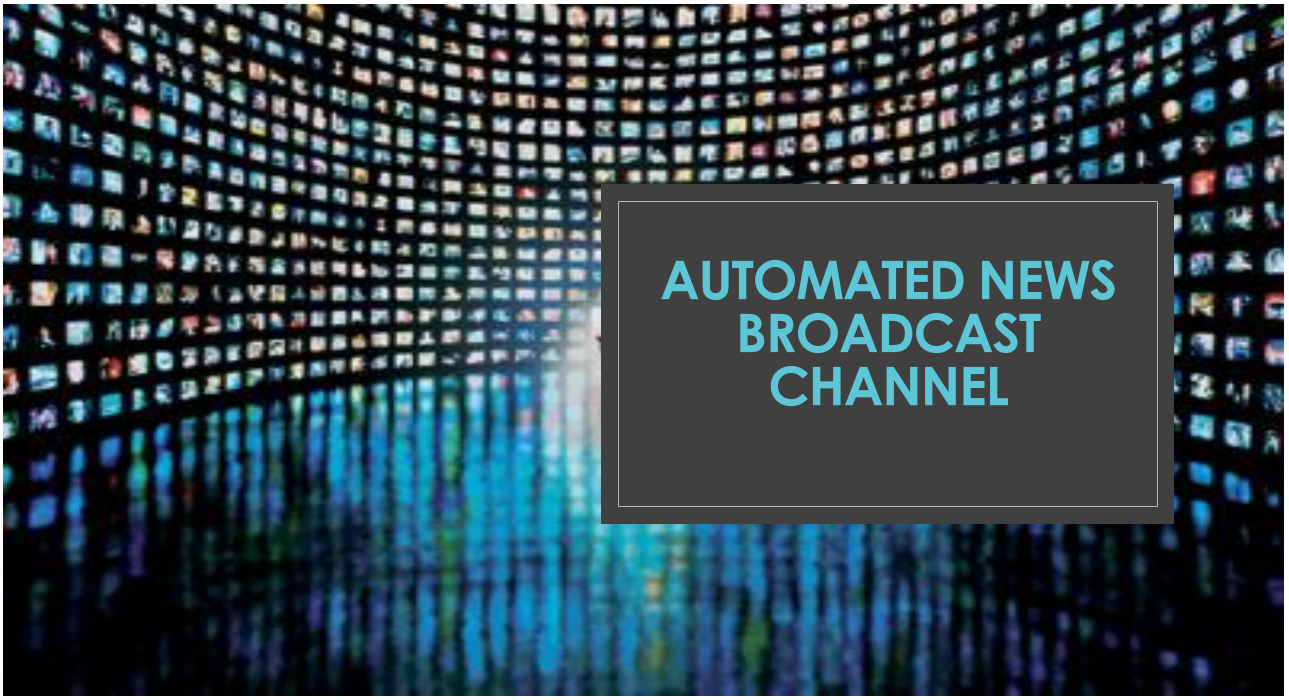
For the time being we would be narrowing down our domain so we would be working on only the weather related news. With the passage of time we would then further increase its domain to sports, entertainment and current affairs.

Value

The total hours that we would be putting in the project would be approximately 1600 with the total cost of approximately Rs 4500-5000. A time duration of approximately 8 months would be required to complete this project

This project would help reduce the use of unauthentic and fake news in the general public by providing them with the correct and authentic content. It would further help save the user's time since it would provide a faster access to the news content.

Below is the attached power point presentation for reference.



Project overview

The world is becoming increasingly dependent on technology. Gadgets/automated systems/ modern equipment are now becoming a part of our daily life. Almost all of the fields and disciplines have been affected by this revolution in technology and it wouldn't be wrong if we say that journalism is one such field which has been impacted by it the most.

Journalism and in particular news media has been hugely influenced by the modern technology. It has never been this easy to develop, discover, produce and consume the news as it is now thanks to this technology. But all of this now comes at a price. With the growing easiness in the news media field, a competition has now arisen; each news channel is competing to produce more and more news in order to gain a higher amount of viewership and in the process of producing more content some people have taken an unethical route. Some channels now deliver fake, unauthentic or incomplete news just so they can increase their ratings or make more bucks. Moreover these news channels are bombarded with tremendous amount of ads which further consumes the precious time of the users.

Thus in order to solve for these problems we aim to develop an automated news reporting agency whose sole purpose would be to deliver complete and authentic news content in fast amount of time. Consequently this further aims to save the precious time of user as he doesn't have to go on navigating different news related websites finding for any particular news or checking for whether it is authentic or not. Accordingly authenticity along with reliability and fastness are the major goals of this system. Moreover user would also have the freedom to search for any older news simultaneously on the same news broadcasting channel. (This feature is not given by any news broadcasting channel for the time being)

Project objectives

The automated news reporting agency project will meet the following objectives:

- By June 2022, our team aims to develop an automated news reporting agency website by putting in 1600 hours effort (8 hours per week for each member)
- Provides an easy to use, user friendly platform for customer/user which can not only be used to watch news but also search for any past news simultaneously.
- This will help in providing user with fast access to the news which can approximately safe 10% of their time that was wasted in browsing and surfing for genuine news.
- It will further help in reducing the consumption of fake/unauthentic news in the people by approximately 10-20%

Project scope

The scope of this project includes and excludes the following items.

In scope:

- The development of an automated news reporting agency to produce genuine news content
- For the time being our System would be based on weather related news only
- NLP based training of the system so that the system could produce the correct and accurate results
- Development of graphics for the website which could give user a more user friendly and responsive website feel
- Designing of the system's animated characters which would present the news
- Testing of the system before making it accessible for the general public

Out of scope:

- Further enhancement of this system in other news related fields (such as entertainment, current affairs)
- Enhancement of system such that present the news with relevant pictures.
- With the use of cookie can show the news related to user's preferences.
- Rather than using animated character, video can be generated using related images/news
- Development of mobile app for this system

Deliverables produced

The Project deliverables of our project are as follows:

- Project Deliverable 1: Analysis of the requirements of the system (Delivery of system's feature list and initial proposal)
- Project Deliverable 2: Delivery of the other system design related diagrams (The class, Use Case, Activity, Sequence Diagrams etc.)
- Project Deliverable 3: Delivery of the system's other detailed diagrams such as component, deployment and Package diagrams and the result of the test cases
- Project Deliverable 4: Delivery of the complete system report (consisting of all the details, working, conclusion, any user training or system maintenance related article & future enhancements)

The Product deliverables of our project are as follows:

- Product Deliverable 1: Delivery of System Diagram (a prototype model on which our system would be based)
- Product Deliverable 2: Delivery of the GUI of the website's front-end/back end
- Product Deliverable 3: Delivery of the system in operable form (i.e. its NLP based working i.e. the filtration, summarization and synthetization of the news related content)
- Product Deliverable 4: Complete finished website/working system for the news reporting agency

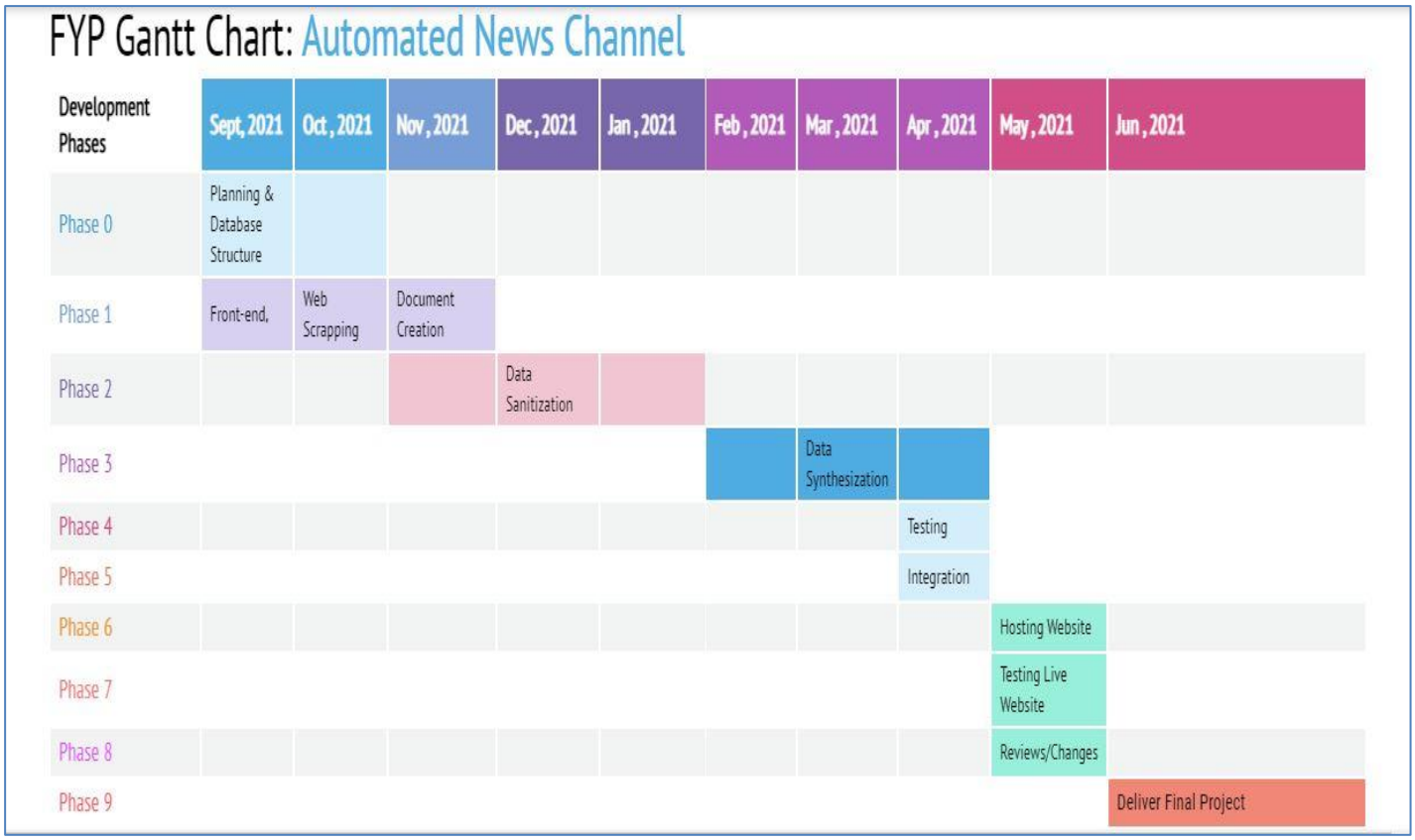
Each of the Project and Product would be delivered in each milestone so for example milestone 1 would consist of the delivery of the Project Deliverable 2 and Product Deliverable 2.

Project estimated effort/cost/duration

This is an 8 month long project and as a team of 4 members we would be putting in 1600 hours in it (2 hours per day for each group member)

The total cost can be approximately around 5000 which would be needed for domain hosting

GANTT Chart:



Estimated effort hours:

The total working hours that would be put into the system would be approximately 1600. This is based on 2 hours per day for each group and since we're a 4 member team so a total of 8 hours per day.

A total of 5 working days (i.e. Monday to Friday) which means 40 hours per week

And a total number of 40 working weeks (approximately, excluding the exam and holiday weeks), so a total of 1600 hours.

Estimated duration:

Milestone	*Date completed*	*Deliverable(s) completed*
Project planning	*06/06/2021*	Project idea: Develop an automated news reporting agency which would be able to deliver complete and authentic news content in fast amount of time. Researching on this idea through different research papers and other related ideas on internet.
Milestone 1	*06/09/2021*	• Initial Proposal idea • System Diagram • Requirement Analysis
Milestone 2	*11/09/2021*	• GUI based front end and back end
Milestone 3	*03/31/2022*	• An executable system with correct NLP based working i.e. the filtration, summarization and synthetization of the news related content
Milestone 4	*05/31/2022*	• Complete finished website/working system for the news reporting agency
Project conclusion	*06/30/2022*	The working of the system along with its complete system report, details, future enhancements and maintainability.

Project assumptions

In order to identify and estimate the required tasks and timing for the project, certain assumptions and premises need to be made. Based on the current knowledge today, the project assumptions are listed below. If an assumption is invalidated at a later date, then the activities and estimates in the project plan should be adjusted accordingly.

- The maximum cost we would be encountering to make this project would be approximately Rs. 4000 - 5000 that is for domain hosting, other than that we don't need to buy any hardware or software component
- There would be no notable changes in the programming languages we would be using (i.e. python) which affects the development of our project.

- The availability of the faculty members would be there in case we get stuck at some point in our project.
- No changes in the university (i.e. UIT/NED's) policy which affects the development of our project.
- We're able to achieve the NLP based working of the system (for synthetization and summarization of the news content) because this is what is the basis of our automated news reporting agency

Project risks

Project risks are characteristics, circumstances, or features of the project environment that may have an adverse effect on the project or the quality of its deliverables. Known risks identified with this project have been included below. A plan will be put into place to minimize or eliminate the impact of each risk to the project.

Risk Area	Level (H/M/L)	Risk Plan
1. The system isn't able to correctly synthesize and filter out the correct news content	High	In case the system isn't able to perform this task we need to go for a change in overall algorithm and design of the system. We might need to see for different algorithms or use some other libraries to achieve the tasks
2. There occur some changes in python language which makes it impossible for us to complete our project	Medium	Shifting to another language (such as Node js or C++)

Product risks

Risk Area	Level (H/M/L)	Risk Plan
1. If the URL/News Channels becomes down, from where we scrap data then this will generate error for our system	High	The admin should monitor the system and add in some other URLs which would be working at that time.
2. Issue of Data security i.e. data must be secured so third party doesn't make any changes in data/database	Medium	Steps should be taken to enhance security such as audit trailing
3. Hacking of news websites/the URLs	Medium	Again steps would be taken to enhance its security as well as the admin/administration would be trained about system maintainability

Project approach

The Project methodology/approach that we would be using would be the Kanban approach. Kanban is based on a continuous approach instead of working in already defined sprint cycles. This the main reason why we're choosing this approach i.e. it is more based on how user/project manager wants to develop the project and make amendments with time in the project rather than on a pre-defined methodology

Tools and technologies

- Use of Python as the programming language
    - Following libraries would be used in python :
        i. Urllib module and urllib.request library: A python library which would be used to fetch and open URLS
        ii. BS4: This Python library would be used to scrape data from the web pages.
        iii. gTTS :- Python library to convert text to speech
        iv. NLP Libraries
- SQLite/Oracle
- Pycharm/VS code
- Flask/Django
- Node.Js/HTML5/CSS
- For coordination purpose we would be using Trello (and Git)

Outstanding Issues

-

Expected Final product

For the Front end part:
A user can watch the news broadcasted to them by the news reporting website

Along with giving it the ability to search for any news simultaneously

For the back end part: There would be a login/Sign up Page for the management team of the agency. The particular admin would login after which he/she would be directed to the Main (Home) page from where they can then select & apply the desired changes (such as selecting website links for data scrapping).

Similar products available

Currently there is no such news channel which produces this kind of news content (i.e. filters out the data for authenticity). For the time being most of the news channel that we see broadcast news manually either by a news caster or we have websites which present the latest news in written form (for example the news channel of BBC and the website of BBC)

However some research has been done in this field in the latest years but there hasn't been any practical implementation on it. i.e. only the research papers have been written on it.

Other than this some steps have also been taken for automated news reporting & publishing such as:

1. AP, Fox and Yahoo uses AI to write news related to sports, financial summaries etc.
2. Dante A Pakistani robo-journal which was introduced back in 2019

List down similar products available online with URL

The links of the research paper are as follows:

i. Hayashi, Masaki, Bachelder, Steven, Tsuruta, Naoya. *Automatic Generation of News Contents from Blog Posts*. Received September 3rd, 2020; Accepted February 10th, 2021.

ii. Aakash Rastogi, Anurag Yadav, Arpit Singhal, Pushpendra Tyagi**.** *REPNEWS- COTENT AGGREGATOR FOR NEWS AND TECHNOLOGY* . Issue: 05/May-2021.

iii. Chukwugoziem, I., & Nwamouh, U. C. (n.d.). Development of an intelligent web based dynamic news aggregator integrating infospider and incremental web crawling technology. International Journal of Scientific & Engineering Research Volume 10, Issue 7, July-2019 370 ISSN 2229-5518.

The links for other news reporting and publishing AI robots are as follows:

i. https://pakwired.com/robot-journalist-dante-an-ai-content-creator/ (Dante)

Our project is basically similar to what has been defined in the research papers although for the time being we're working on limited scope (i.e. the weather domain only)

Project approvals:

Add any signatures that are important for the approval of the project. (Remove this comment section from final document.)

Name: Project Supervisor	Signature
Member, FYP Committee, UIT	Signature