

PROPOSAL FORM
FINAL YEAR PROJECT IN COMPUTERSCIENCE/SOFTWARE
ENGINEERING
Usman Institute of technology

**TITLE IN CAPITAL LETTERS, FONT SIZE 14,
TIMES NEWS ROMAN, CENTRED, BOLD**

YUSRA MASOOD, 18B-093-CS, A
RUSHAN RAFIQ, 18B-087-CS, A
YUMNA ASIM, 18B-047-CS, A
SHIZA KHAN, 18B-130-CS, A

SUPERVISED BY:

Department of Computer Science

May 2021

PROPOSAL FORM
STUDENT INFORMATION FOR FINAL YEAR PROJECT

Fill in the information below:

Roll No	Enrollment No	Name	Email	Contact No
18B-093-CS	UIT/434/2018-19	YUSRA MASOOD	ymasood@students.uit.edu	03022825869
18B-087-CS	UIT/317/2018-19	RUSHAN RAFIQ	rushan@students.uit.edu	03072171344
18B-047-CS	UIT/431/2018-19	YUMNA ASIM	yasim@students.uit.edu	03212897252
18B-130-CS	UIT/345/2018-19	SHIZA KHAN	shkhan@students.uit.edu	03363077057

PROPOSAL FORM

**TITLE IN CAPITAL LETTERS, FONT SIZE 12,
TIMES NEWS ROMAN, CENTRED, BOLD**

Background	<p>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.</p> <p>Thus to solve for this problem our team decided to make an automated news reporting agency. This would be an AI based project which could be used to help reduce the unauthentic consumption of the news by the general public.</p>
Project Scope	<p>Currently there is no such news channel which produces this kind of news content (i.e. filters out the data for authenticity). Although some research has been done in this field in the latest years but there hasn't been any practical implementation of it .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. Our website would be able to deliver both broadcasting news as well as the past news in written form if any user wants to search for it. Furthermore the main purpose of this project would be to provide authentic news which would be formed using the whole process of Natural Language Processing. Where NLP would be used in filtering out the authentic news and generating correct news from it.</p> <p>The scope of this project is to provide a platform that automatically generate news by scraping data from different URLs/News channel websites and provide a user friendly interface so that user can get updated news also it will provide a summarized document that provide a synthesized and sanitized data which comes from different URLs of news channels and remove the redundancy of news. (The domain should be restricted for any one topic)</p> <p>For the time being our focus would be on the weather domain i.e. the news related to weather only. Once our goals for the project are achieved in this domain we would further focus on the other domains (such as sports, entertainment, current affairs etc.)</p>

Project Description	<p>The main purpose behind this project is to automate the process of news reporting channel by creating a web app which would aggregate the news from different news reporting websites, filter the authentic data, summarize that data and then represent it. This is done to reduce the constant use of fake, unauthentic news by different news media channels who now only aims to earn viewership & money rather than delivering authentic news to users/viewers.</p> <p>The main features of this project are synthesizing and sanitizing the scraping data that were extracted from different URLs, summarizing that data and representing/broadcasting it on the website.</p>
Expected Outcome	<p>For the Front end part: A user can watch the news broadcasted to them by the news reporting website. A user can also search for any past news (this feature not present in any news broadcasting channel until now)</p> <p>For the back end part: There would be a login/Sign up Page for the management team of the agency. The particular user 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).</p>

Method/Approach	<p>The working of our system would be as under</p> <ul style="list-style-type: none"> • There would be an admin at the backend who can login to the system. • The admin would then be able to select the list of URLs/News channel websites, feed them into the database. • On those websites web scraping technique would be applied to gather the required data related to news content. This technique would be applied after every set time interval period to refresh the new content. • Once the data is scrapped it would be filtered, summarized, synthesize and sanitize, checked for redundancy and a bulletin is made out of it. • The data would also be compared with the previous data present in database for further authentication. • The bulletin would be converted from written to oral from using text to speech. • It would then be presented by any animated broadcaster from our website. • Furthermore since the previous data would be saved in the database so the user would also have the facility to search for any old news simultaneously <p>For the implementation we would make use of the following libraries (The programming language used here would be python) :</p> <ol style="list-style-type: none"> 1. Urllib module and urllib.request library: A python library which would be used to fetch and open URLs 2. BS4: This Python library would be used to scrape data from the web pages. 3. Use of NLP (Natural Language Processing) Technique to synthesize data/NLP libraries 4. gTTS :- Python library to convert text to speech
Relevant references	<p><i>Books, journals, conference papers, and (not many) some internet links</i></p> <ul style="list-style-type: none"> ◦ 1) https://www.bbc.com/news/technology-40517420 ◦ 2) https://en.wikipedia.org/wiki/Automated_journalism ◦ 3) https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html ◦ 4) https://www.theguardian.com/commentisfree/2019/feb/01/why-i-created-a-robot-to-write-news-stories ◦ 5) Hayashi, Masaki, Bachelder, Steven, Tsuruta, Naoya. <i>Automatic Generation of News Contents from Blog Posts</i>. Received September 3rd, 2020; Accepted February 10th, 2021. ◦ 6) Aakash Rastogi, Anurag Yadav, Arpit Singhal, Pushpendra Tyagi. <i>REPNEWS-COTENT AGGREGATOR FOR NEWS AND TECHNOLOGY</i> . Issue: 05/May-2021. ◦ 7) Chukwugoziem, I., & Nwamouh, U. C. (n.d.). Development of an intelligent web based dynamic news aggregator integrating infospider and incremental web crawling technology. <i>International Journal of Scientific & Engineering Research</i> Volume 10, Issue 7, July-2019 370 ISSN 2229-5518.

PROJECT DETAIL

	TECHNOLOGY DOMAIN (TICK ONE OR MORE)
<input type="checkbox"/>	Desktop Application
<input checked="" type="checkbox"/>	Web Application
<input type="checkbox"/>	Mobile Application
<input type="checkbox"/>	Cloud Base Application
<input type="checkbox"/>	H/W Based Application

	Project Streams (TICK ONE OR MORE)
<input type="checkbox"/>	Block Chain
<input type="checkbox"/>	Image Processing
<input checked="" type="checkbox"/>	Artificial Intelligence
<input type="checkbox"/>	Computer Networks/Wireless Networks
<input type="checkbox"/>	Business Application / MIS Application
<input type="checkbox"/>	CASE Tools
<input type="checkbox"/>	System Software
<input type="checkbox"/>	Software Automation

For Office use only (To be filled by the Final Year Project Coordinator)

Sr. No	Evaluator Name	Remarks	Accepted/ Rejected

Idea Accepted	Idea Rejected
	Reason(s): Next Action Plan:

FYP COORDINATOR