

# **CLASP**

**COVID-19 LITERATURE ANALYSIS AND  
SUMMARISATION PLATFORM**

**HARSH SAKHRANI**

**TANMAY PARDESHI**

**KAUSTUBH ODAK**

**SALONI PAREKH**

**SNIGDHA SINGH**

# TABLE OF CONTENTS

## 1. INTRODUCTION

- 1.1. Motivation
- 1.2. Solution
- 1.3. Overview
- 1.4. Scope
- 1.5. Features
- 1.6. References

## 2. USE CASES

- 2.1. User and Text Summarization
- 2.2. User and Information Retrieval
- 2.3. User and Q & A Model
- 2.4. User and News System
- 2.5. User and Anomaly Detection
- 2.6. Complete Use Case Diagram

## 3. DATA FLOW

- 3.1. Data Flow/Architecture diagram

## 4. USER INTERFACE

## 5. TECH STACK

# 1.INTRODUCTION

## 1.1 MOTIVATION

Since the advent of the pandemic, the number of research papers written about it, analysing its various aspects have shot up exponentially. As the world strives to create a vaccine, it is not possible to take into account every ounce of the data available, in its original, vast form. A lot of useful data remains unattended.

The wildfire-like spread of the coronavirus has raised millions of questions worldwide. Panic rises as the sources of data become more unreliable each passing day.

The fact that no technology in 2020 was able to accurately predict the global impact of this virus has led to its unprecedented spread and irrevocable damage to life and economy worldwide.

## 1.2 SOLUTION

A tool that summarises the sea of data available down to its most essential, key features is crucial. We can ensure that scientists have every fact to help develop the vaccine.

We provide a solution where the user can enter his/her query and have a research paper answer it, instead of a news tabloid or any other unverified source of data. This diffuses the surge in panic considerably.

We track the healthcare spend among other parameters and generate an alert in places where this becomes out of the ordinary. Our solution thus

alerts the scientists across the globe to any anomalous behaviour that may lead to the type of grave crisis we are facing today.

## **1.3 OVERVIEW**

The project is a web application that connects the user to 3 key functionalities - text summariser, Q & A model, a search engine specifically for research papers on COVID-19, and an anomaly detection map.

The text summariser summarises the file and the result is rendered on screen for the user.

The text summarisation makes use of unsupervised learning and summarises the document by identifying the key points using TextRank which generates context.

The user can upload the file (screenshot, .doc, etc. ) he wishes to summarise on our website.

To answer the user's query pertaining to the medical impacts of the coronavirus a Q&A model has been created. The user enters his/her query onto our website and we respond with the concise, appropriate answer to the query. Along with the answer, for further reference of the user the document that was referred to for the response is also rendered. This satisfies the need of the user when he wishes to look deeper into the response.

News about this pandemic is scattered all across the world wide web. Our platform provides a one stop solution for users across the globe to look at the medical and research updates on COVID-19. We compile the

most reliable sources of information from across the internet and display them one click away.

The anomaly detection lays out the status of the entire world in the form of a map. The locations showing abnormal behaviour with respect to parameters like GDP, money spent on healthcare, etc. are highlighted with a red spot. This alerts the users to the possible advent of incidents in those places.

## **1.4 SCOPE**

- Easy access to concise data which will catalyse the process of development of a vaccine.
- Reduction of panic in the users worldwide by answering their questions using biologically correct context.
- Prevention of future pandemics by alerting the appropriate personnel to anomalous changes.
- Enable one click access to research news scattered across the internet on COVID-19.

## **1.5 FEATURES**

- Simple, responsive and engaging user interface.
- Summarisation of lengthy research articles to their useful key points.
- Easy availability of correct answers for questions on research of COVID-19.
- One spot access to global research updates regarding the pandemic.
- Clear visualisation of anomalous behaviour using anomaly map.
- Application runs smoothly across all devices (Mobile phone, Desktop, Laptop etc.)

→ The login and register feature enables the user to view the previously searched queries, questions and generated summaries.

## 1.6 REFERENCES:

→ [Unsupervised Text Summarization Using Sentence Embeddings](#) :

This research paper explains the process of text summarization using unsupervised methods. It is done by clustering sentence embeddings trained to embed paraphrases near each other.

→ [Application and analysis of text summarization for biomedical domain content](#) :

In this research paper, the approach is to implement and analyse abstractive and extractive text summarization machine learning models for general language as well as biomedical domain-specific text. For abstractive text summarization, a sequence-to-sequence model that utilizes recurrent neural networks (RNNs) for biomedical text summarization. For extractive text summarization, a pretrained BERT model is used.

→ [Supervised Machine Learning for Extractive Query Based Summarisation of Biomedical Data](#) :

This paper explores the impact of several supervised machine learning approaches for extracting multi-document summaries for given queries. It compares classification and regression approaches for query-based extractive summarisation using data provided by the BioASQ Challenge.

→ [Information Retrieval as Statistical Translation](#) :

This paper proposes a new probabilistic approach to information retrieval based upon the ideas of statistical machine translation. The main approach is a statistical model on how a document can be translated into a query.

→ [1 Statistical Language Modeling For Information Retrieval Xiaoyong Liu and W. Bruce Croft Center for Intelligent Information Ret :](#)

This paper reviews research and applications in statistical language modelling for information retrieval (IR) that has emerged within the past several years as a new probabilistic framework for describing information retrieval processes.

→ [Unsupervised Question Answering by Cloze Translation :](#)

This research paper explores to what extent high quality training data is actually required for Extractive QA, and investigates the possibility of unsupervised Extractive QA. This problem is approached by first learning to generate context, question and answer triplets in an unsupervised manner, which we then use to synthesize Extractive QA training data automatically.

→ [\(PDF\) A review on anomaly detection in disease outbreak detection :](#)

Gives a brief description about detection of pandemic like anomalies using AI

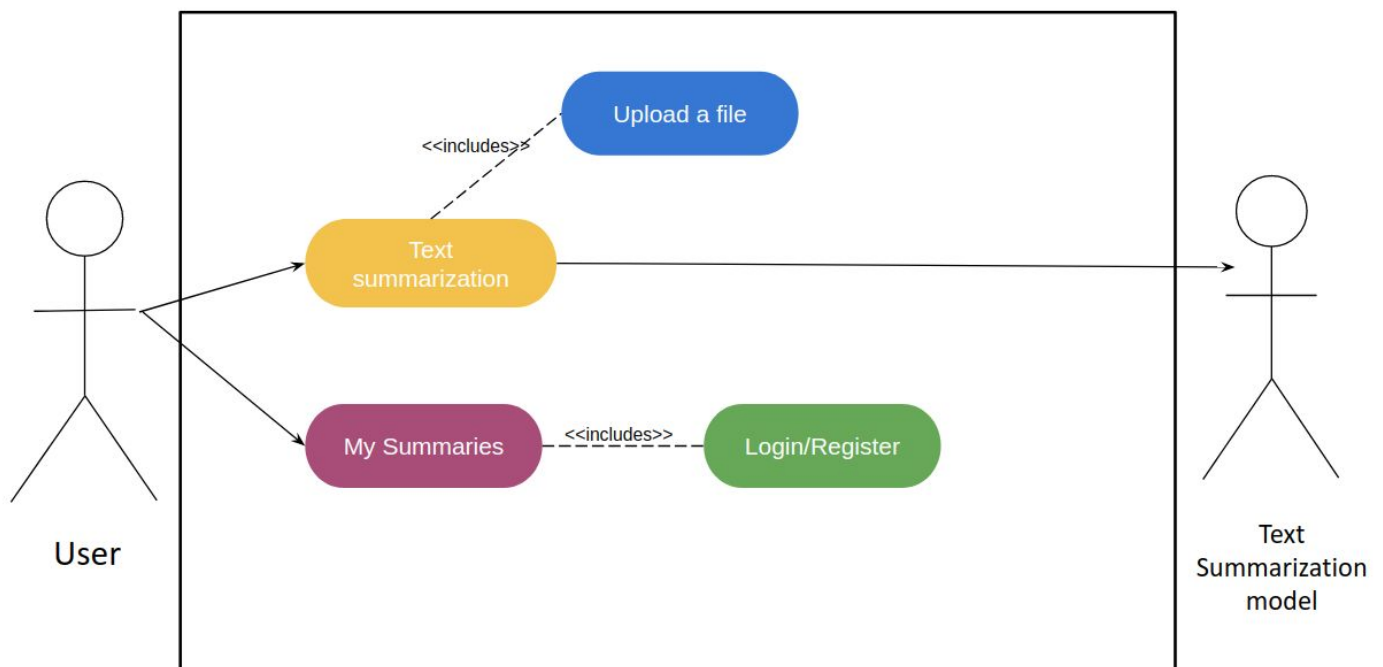
→ [\[1810.04805\] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding :](#)

The paper explores the architecture of the current State Of The Art Language Representation Model - BERT

## 2. USE CASE

### 2.1 USER AND TEXT SUMMARIZATION

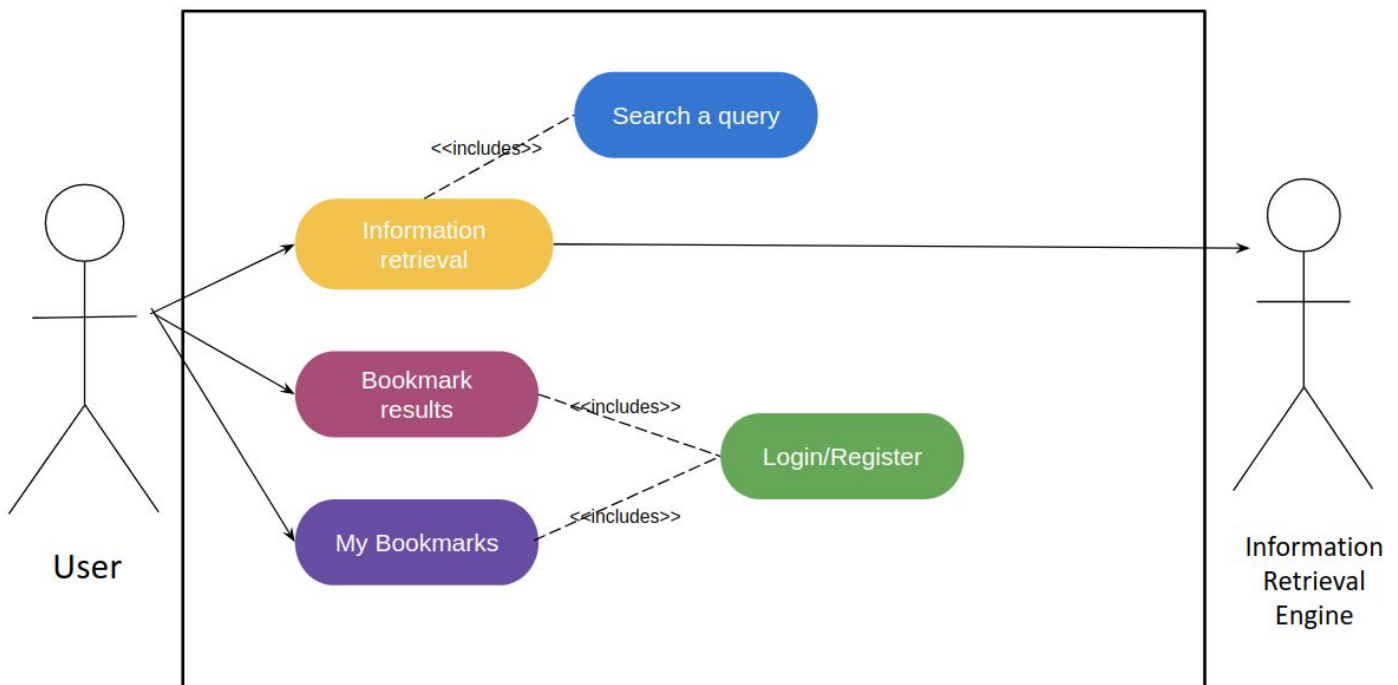
1. The user can upload any file of type .txt, .docx or an image file of the text to be summarized.
2. The summary of the paper and the link to navigate to it will be generated and displayed on screen. At this link, the user can view or download the summary in pdf format.
3. If the user has signed in, the summaries are stored automatically in the 'My summaries' section. The user can view or delete summaries in the future.





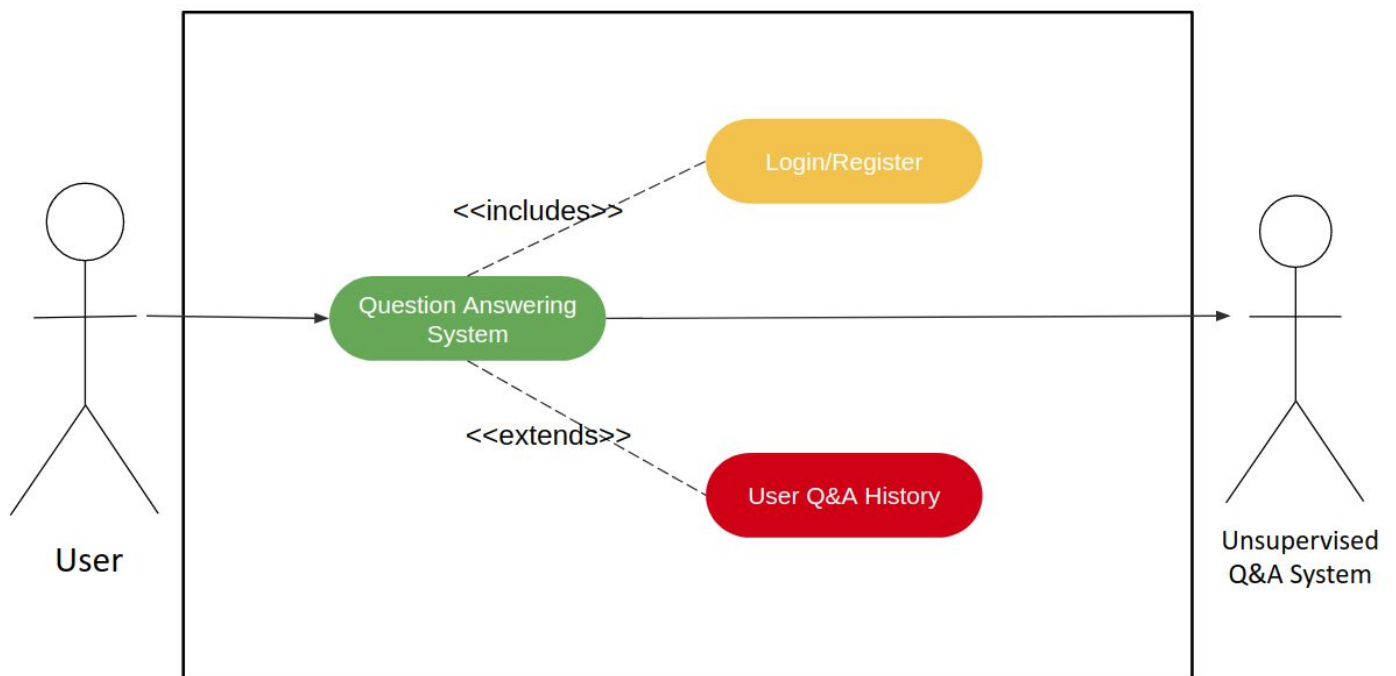
## 2.2 USER AND INFORMATION RETRIEVAL

1. The user can search for a research paper by its title or its author's name.
2. A list of 20 results is generated relevant to the user's query. The user can open any of these by clicking on the title or the body of the result card.
3. The user can also bookmark these results by signing up and logging in, and view or delete these bookmarks in the 'My Bookmarks' section.



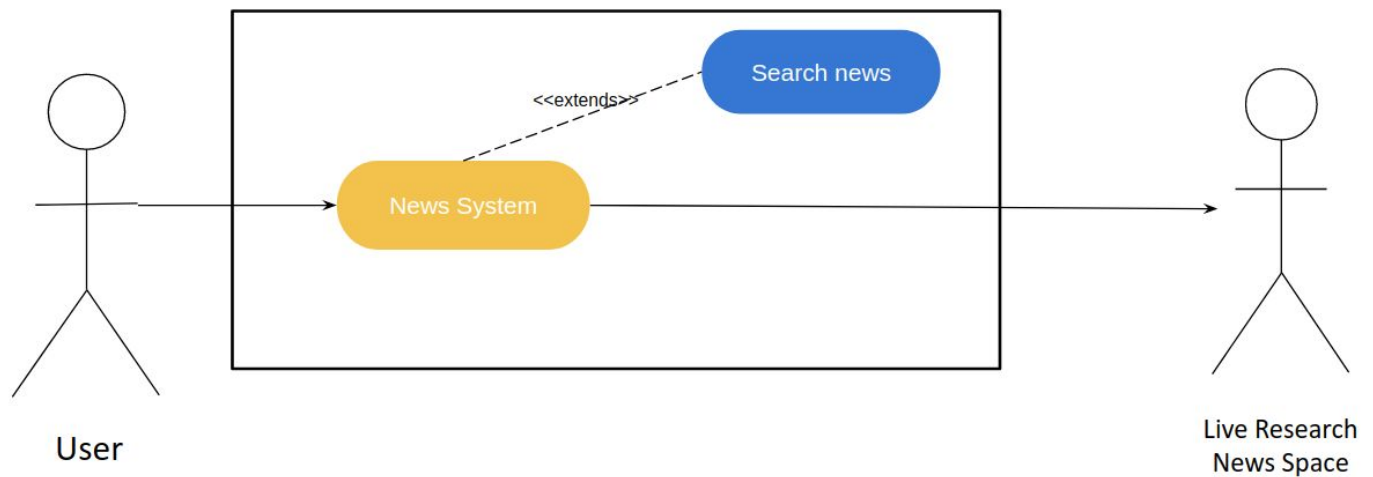
## 2.3 USER AND Q & A MODEL

1. Users have to login or sign up to use this feature.
2. Users can input a question and an answer will be given along with the paragraph from where the answer was extracted as well as the title of the research paper consisting of that paragraph.
3. All the data is stored in the database for future reference and can be accessed in the My Questions section of the website.



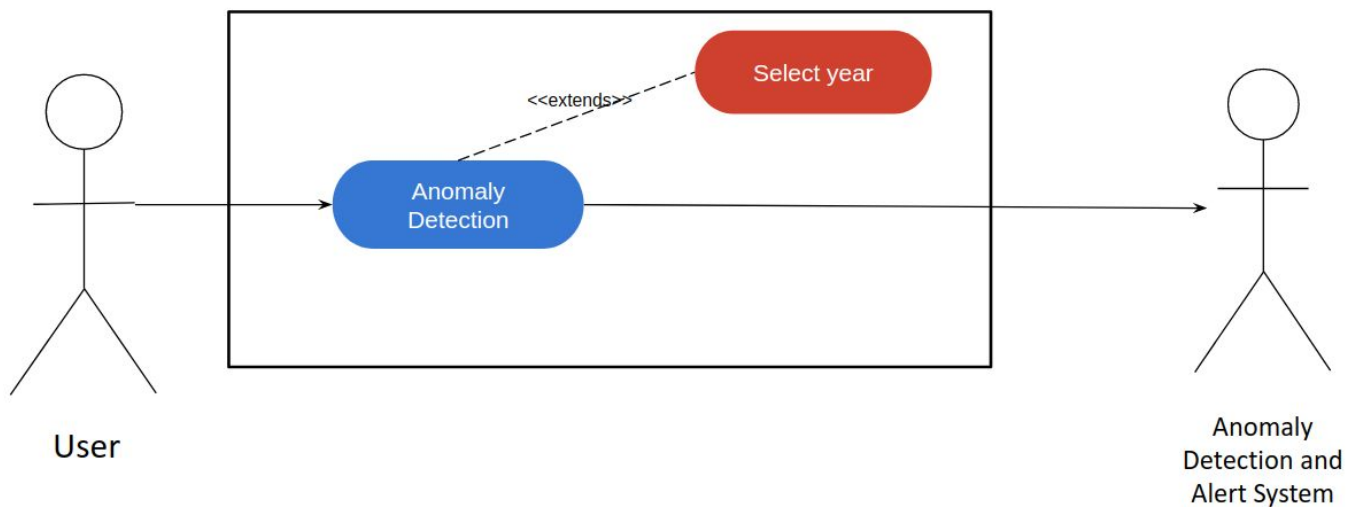
## 2.4 USER AND NEWS SYSTEM

1. This is the homepage of the website where real time news headlines are displayed.
2. Users can click on the headline to access the entire news articles.
3. An option to search a query to get specific news articles is also provided.

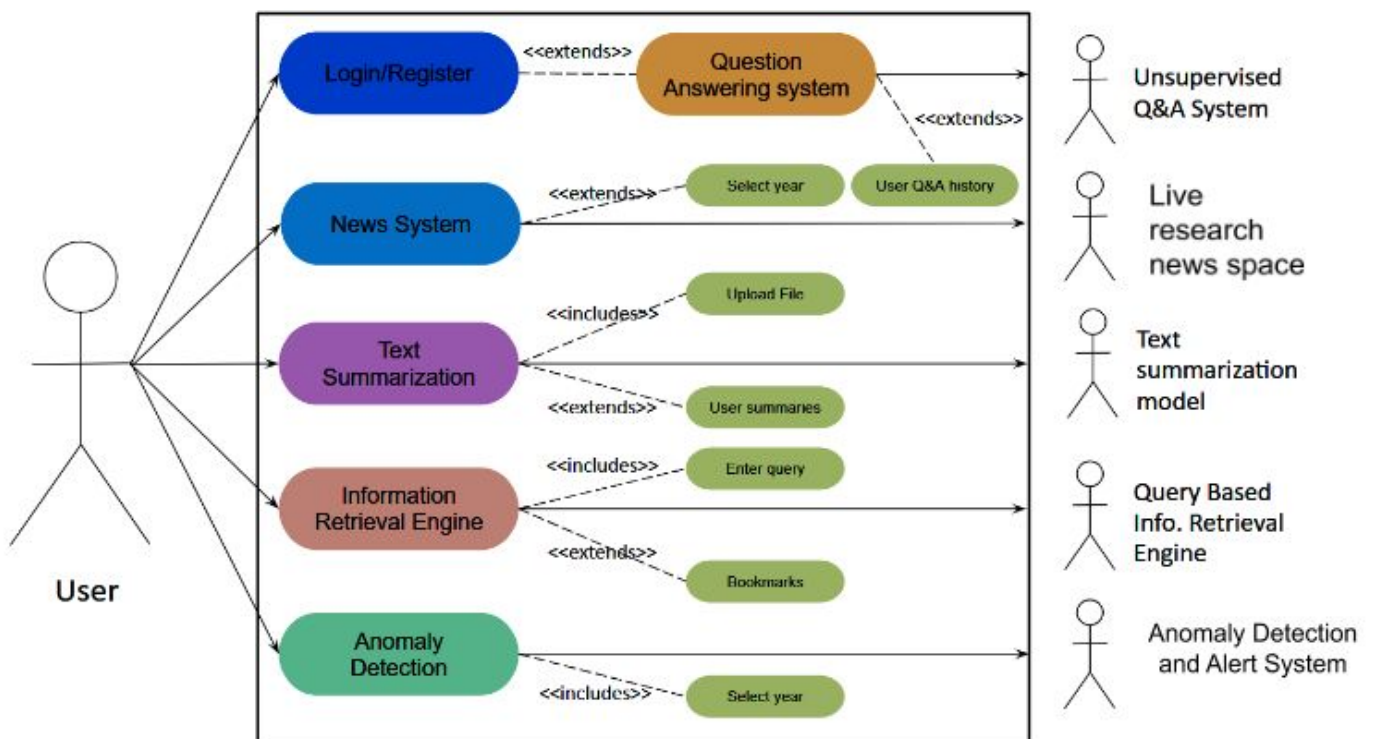


## 2.5 USER AND ANOMALY DETECTION

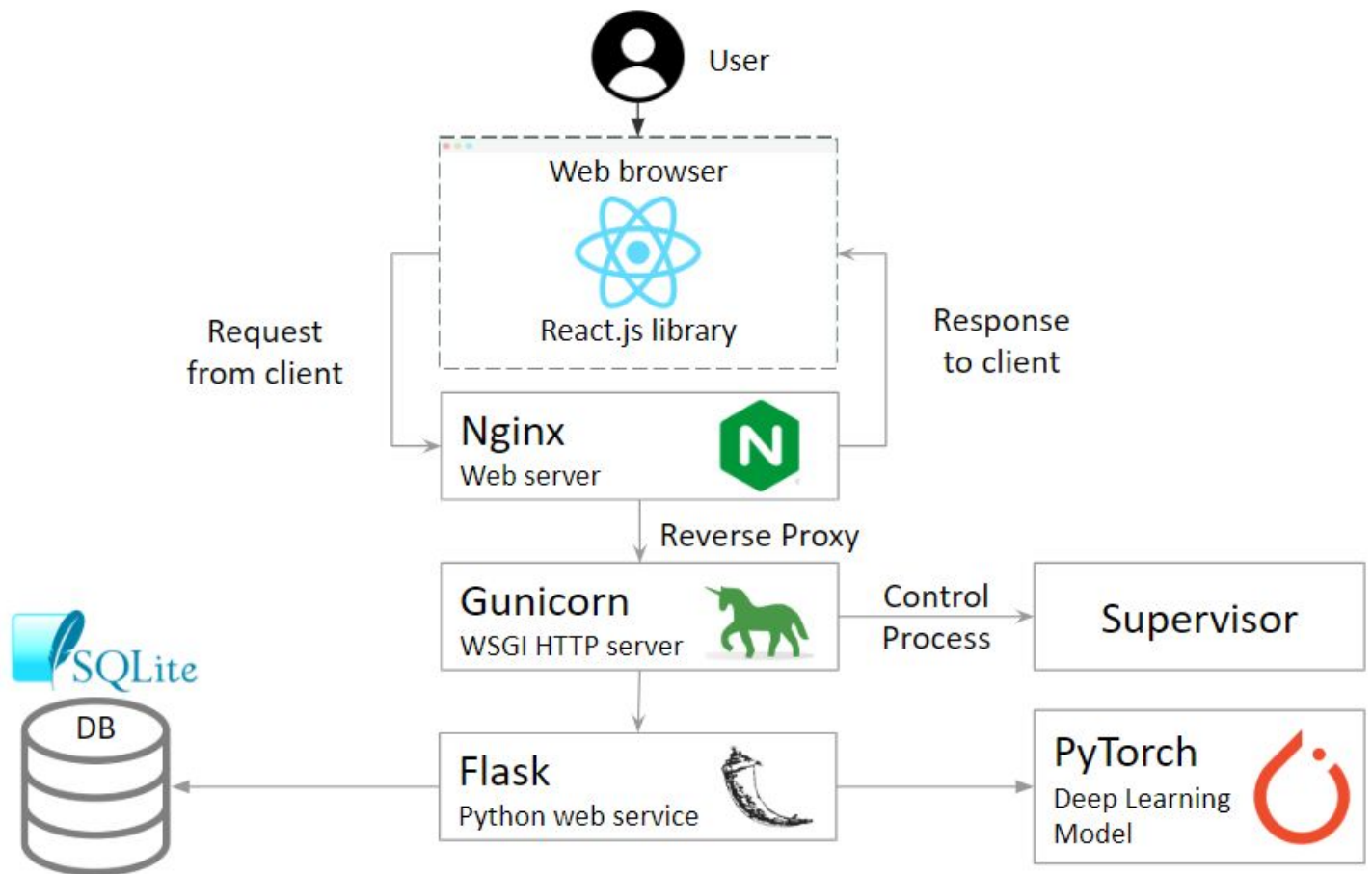
1. Anomaly detection makes use of statistical trend analysis for more than 180 countries all around the world to detect abnormal behaviour with respect to parameters like GDP and Total Amount of Health Care Spend.
2. Users can get a visual representation on the world map for this feature.
3. Additionally, users can select a year in which they want to analyse anomalies



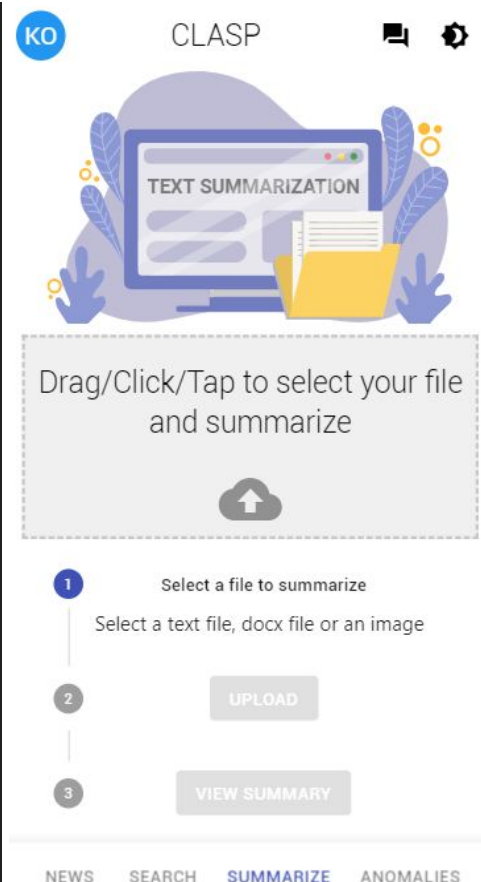
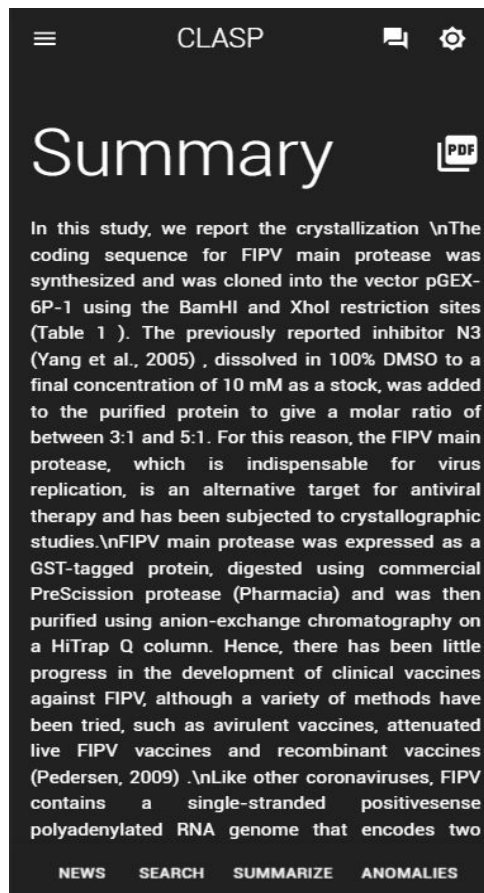
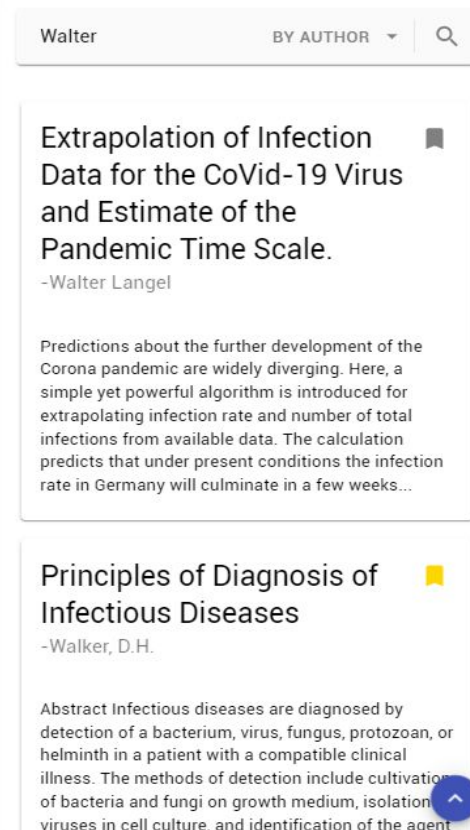
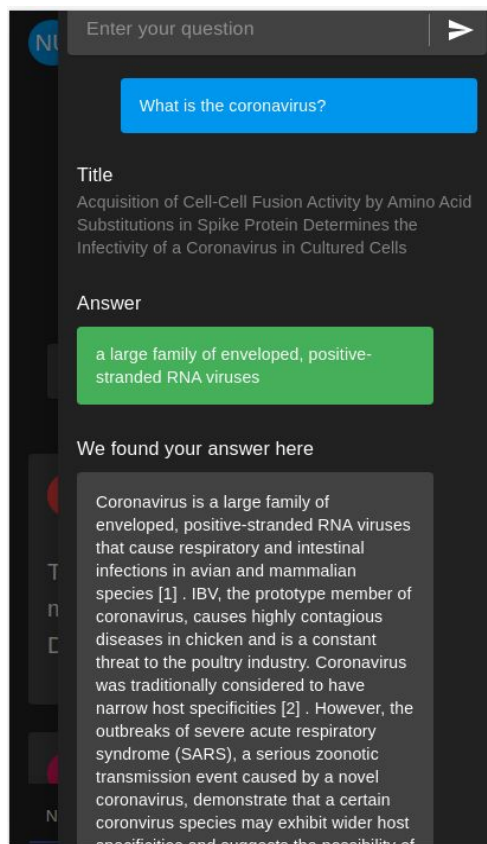
## 2.6 COMPLETE USE CASE



### 3. DATA FLOW




















## 4. USER INTERFACE



## My Bookmarks


Q Search ×




Title	Author	Actions	
> The past and present of malaria diagnosis	Walker, John		
> Discoveries in Molecular Genetics with the Adenovirus 12 System: Integration of Viral DNA and Epigenetic Consequences	Doerfler, Walter		
> Human Adenovirus Type 12: Crossing Species Barriers to Immortalize the Viral Genome	Doerfler, Walter		
> 7 Lung Infections	McCullough, Ann E.; Leslie, Kevin O.		
> Gene Modulation for Treating Liver Fibrosis	Cheng, Kun; Mahato, Ram L.		
> Gene therapy for cystic fibrosis lung disease	Sumner-Jones, Stephanie G.; Gill, Deborah R.; Hyde, Stephen C.		
> Virology	Boland, G. J.		
> Viral Infections of the Lung	Zaki, Sherif R.; Paddock, Christopher D.		




Search for news

 AS  
3 minutes ago


Coronavirus India live updates: cases, deaths and news, today

 Times of India  
10 minutes ago

SOPs revised for flights to bring Indians home

 Times of India  
10 minutes ago

After wreaking havoc in Rajasthan and MP, locusts attack UP

 Times of India  
38 minutes ago

Domestic flights from today: Only asymptomatic passengers can fly, govt says

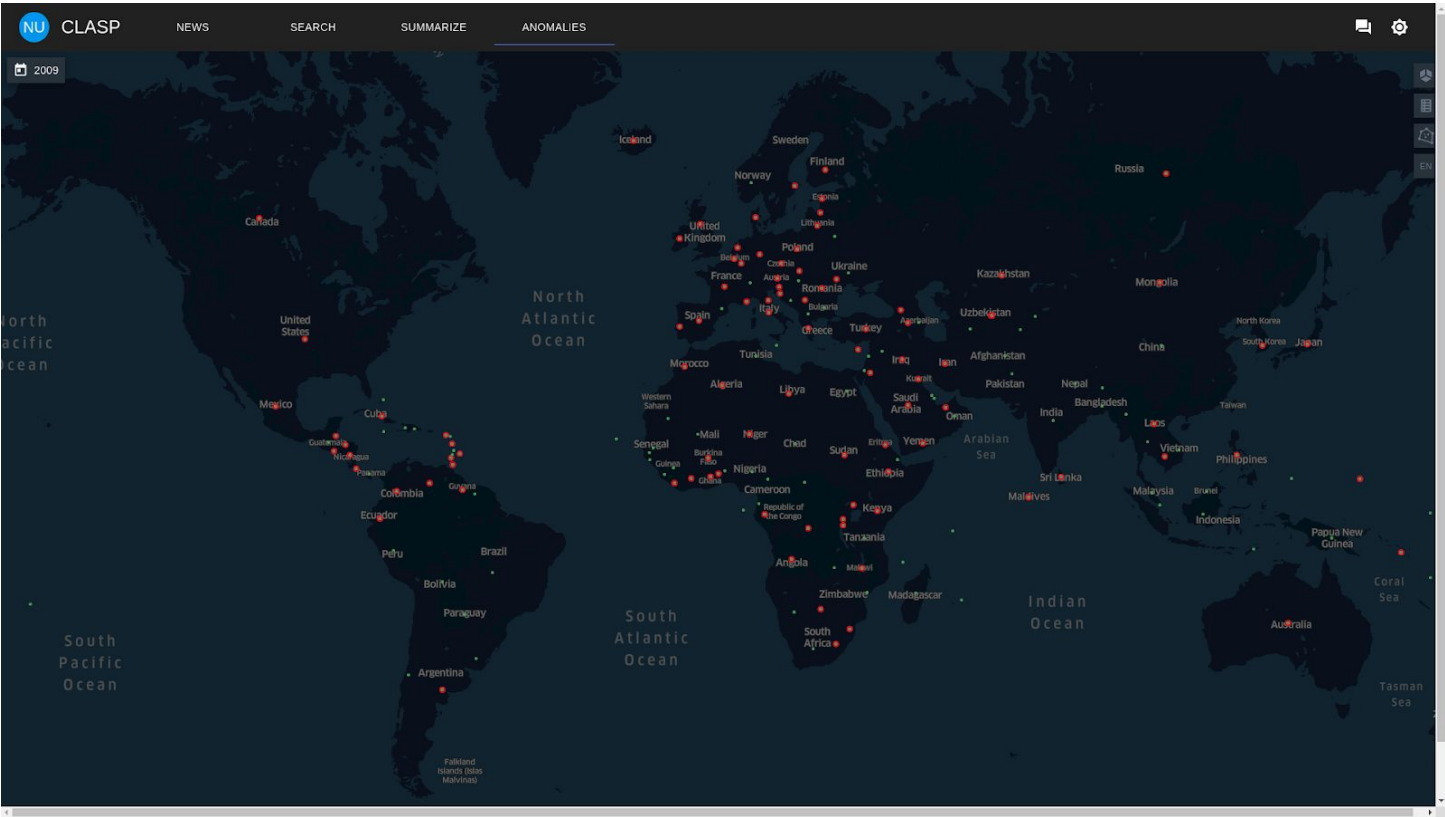
 Al Jazeera English  
55 minutes ago

Chile's hospitals under strain as cases near 70,000: Live updates

 India Today  
1 hour ago

Coronavirus LIVE Updates: States release guidelines for resumption of domestic flights





NU CLASP

### QnA history

Search			
Title	Question	Answers	Actions
<div>▼</div> <div>a large family of enveloped, positive-stranded RNA viruses</div>	What is the coronavirus?	Acquisition of Cell-Cell Fusion Activity by Amino Acid Substitutions in Spike Protein Determines the Infectivity of a Coronavirus in Cultured Cells	<div>🗑️</div>
<p>Coronavirus is a large family of enveloped, positive-stranded RNA viruses that cause respiratory and intestinal infections in avian and mammalian species [1]. IBV, the prototype member of coronavirus, causes highly contagious diseases in chicken and is a constant threat to the poultry industry. Coronavirus was traditionally considered to have narrow host specificities [2]. However, the outbreaks of severe acute respiratory syndrome (SARS), a serious zoonotic transmission event caused by a novel coronavirus, demonstrate that a certain coronavirus species may exhibit wider host specificities and suggests the possibility of crossspecies transmission of animal coronaviruses to human [3, 4]. Cross-species transmission was also observed in coronavirus transmissible gastroenteritis virus (TGEV) and human coronavirus OC43 [5] [6] [7]. These events highlight the importance of understanding the mechanisms of interspecies adaptation and transmission of coronavirus.</p>			
5 rows < 1-1 of 1 >			

## 5. TECH STACK

 PyTorch

 TensorFlow

 NLTK

 spaCy

 NGINX  
Part of FS

 gunicorn

 kepler.gl

 SQLite

 Flask

 React

 Redux

Tesseract.js

 MATERIAL-UI

 Azure