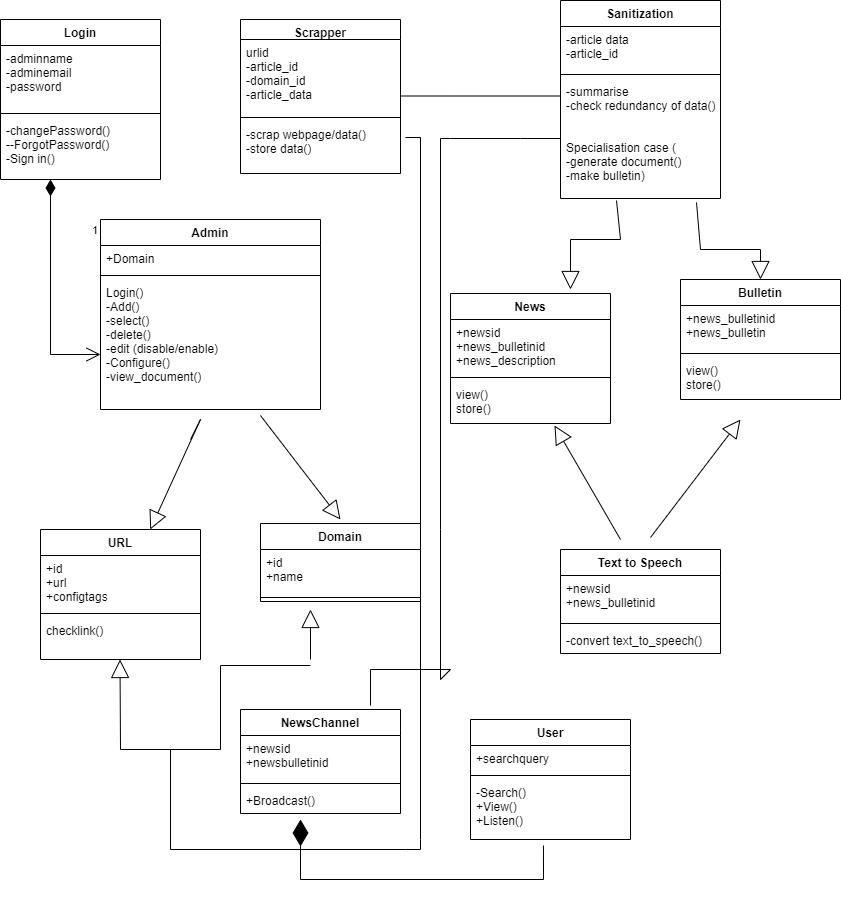
# 5. Software design and modeling:

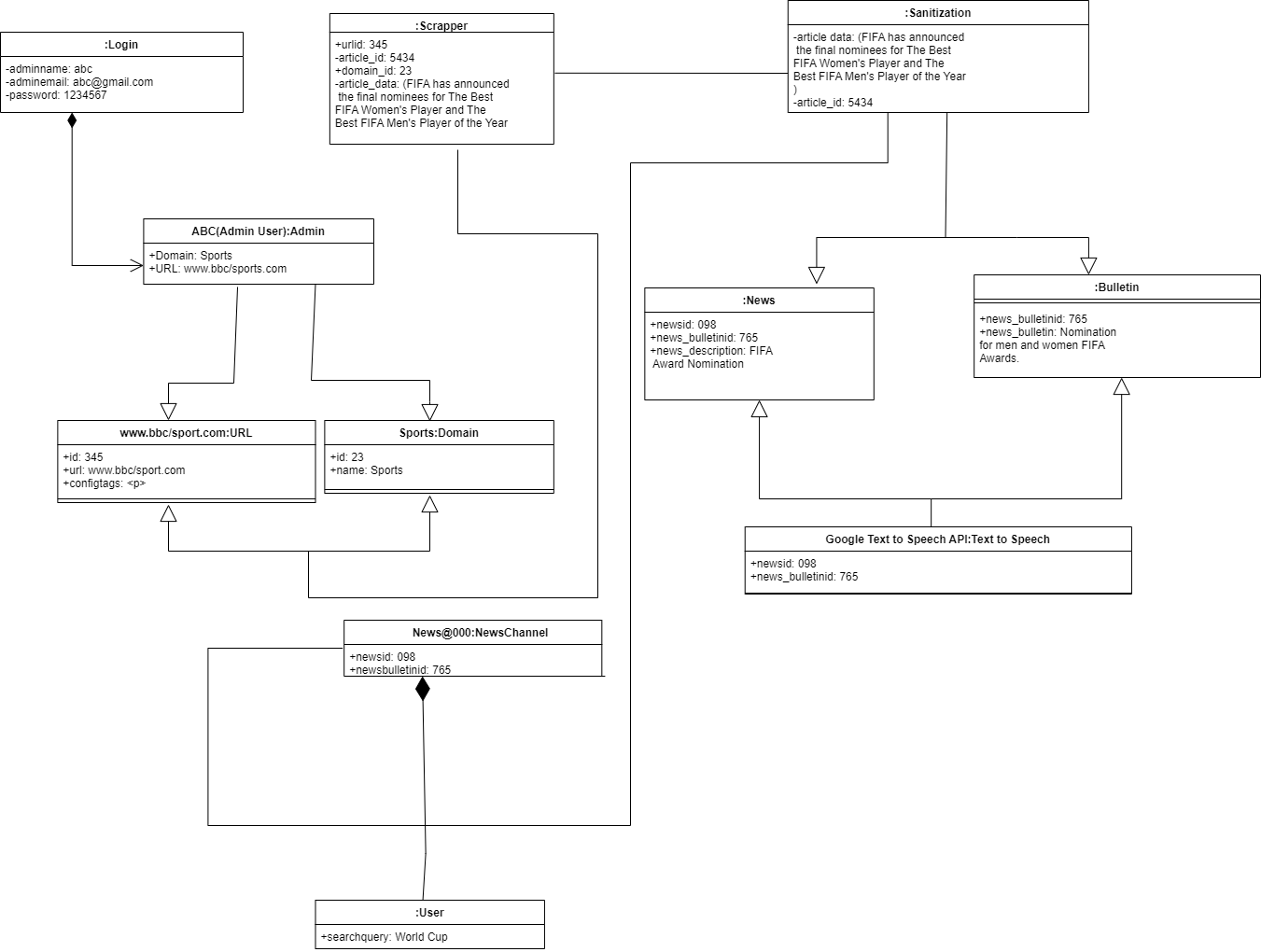
Software designing is basically the behavioral and structural overview of project which is used for better understanding of a project. For software designing we have used UML language in order to describe the software graphically.

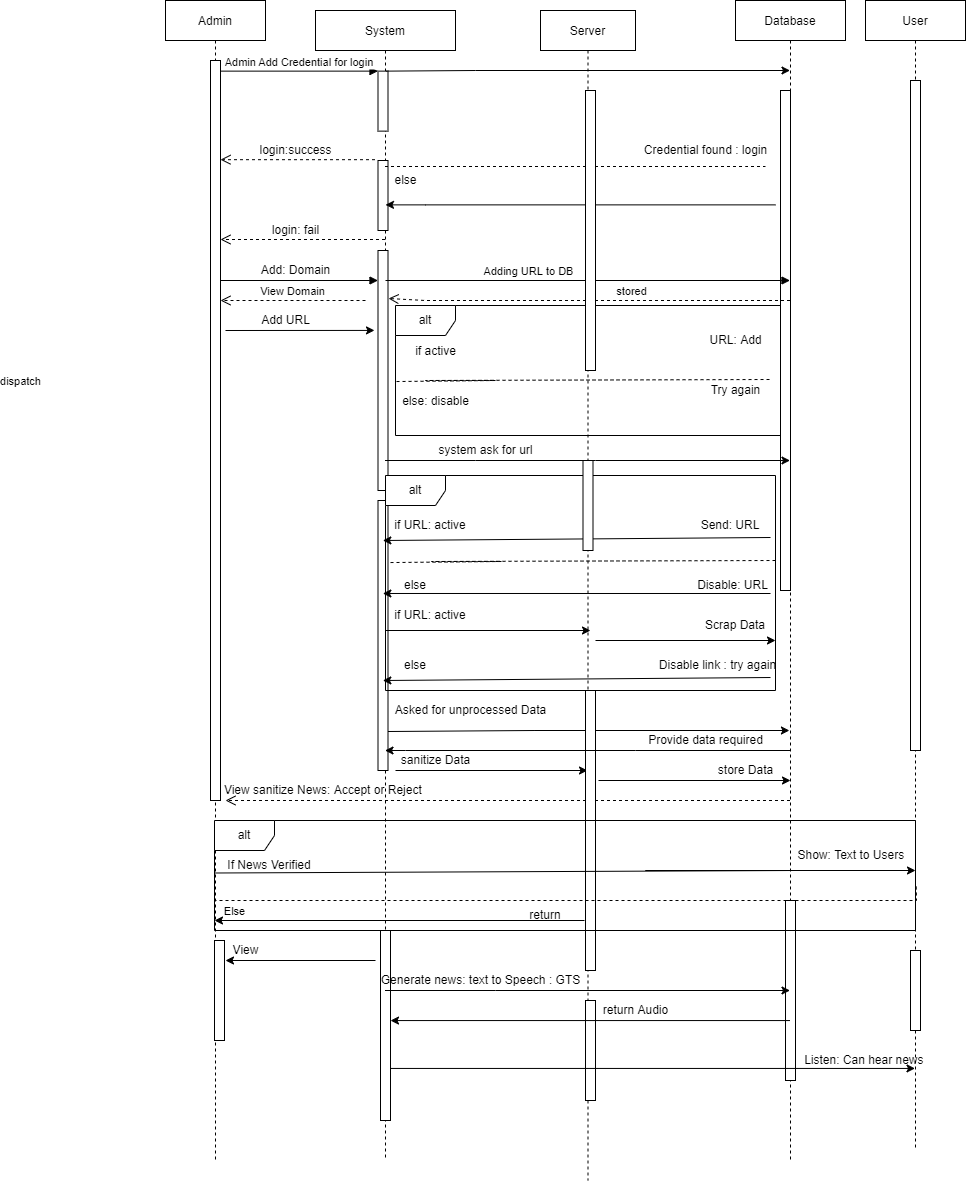
Software modeling is a vital part in software development process. In this project approach of requirement gathering was observed, according to the problem people faced, and what steps might be observed to counter the ones troubles after analyzing approaches prototypes had been developed and tested in main scenarios. In our project we use Kanban development process which is responsible for managing and improving the work flow.

## 5.1 Class Diagram

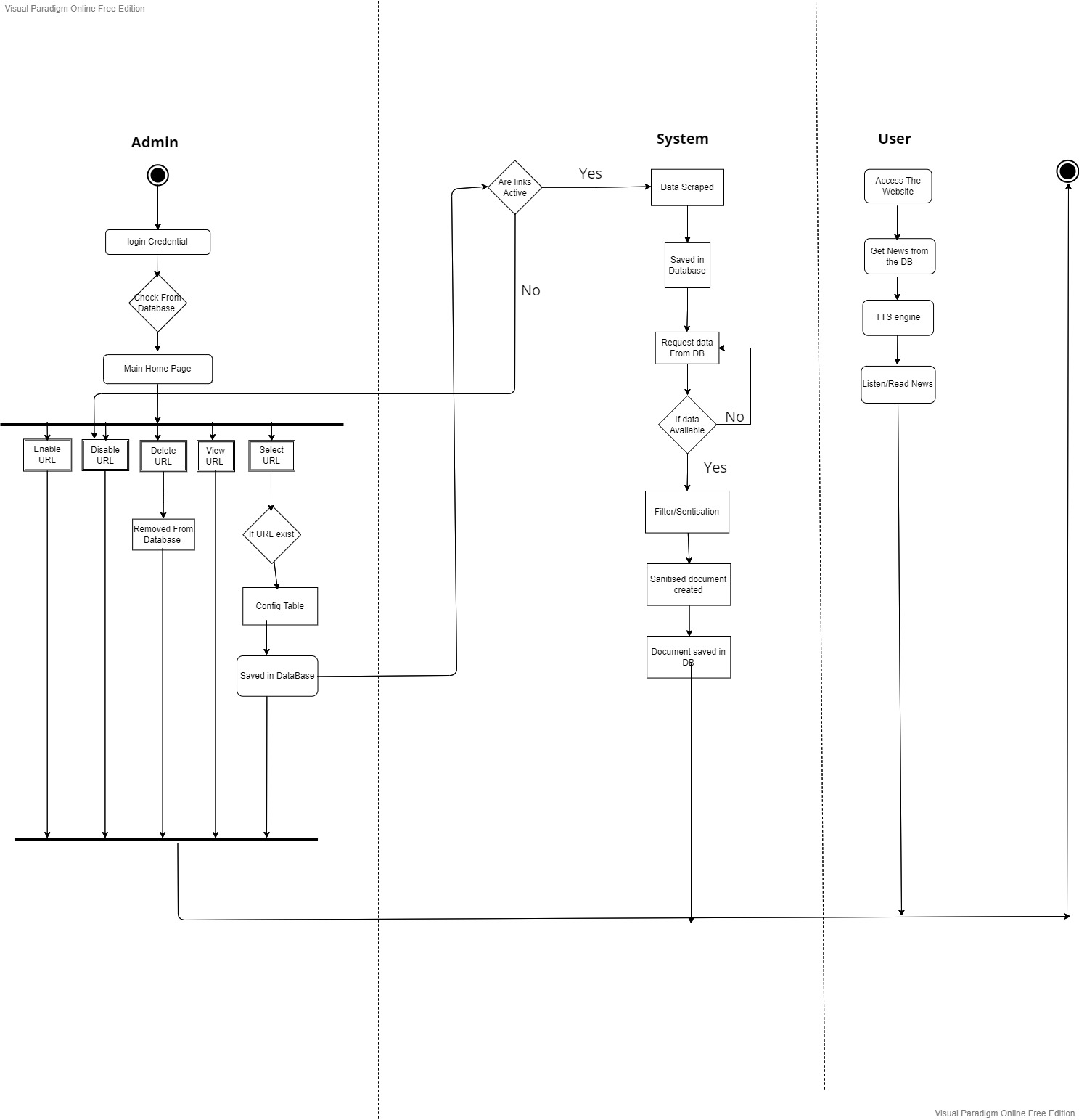


## 5.2 Object Diagram:



5.3 Sequence Diagram:

## 5. 4 Activity diagram:



# 6. Algorithm analysis and complexity:

## 6.1 Scrapper Algorithms:

There are many algorithms for web scrapping but we are using following algorithms which are best suited. Here we are using combination of Beautiful Soup 4 and Selenium to reduce the data scrapping time.

### 6.1.1 BeautifulSoup:

It is a Python library that is use for creating parse tree of parsed pages which is used to pull out the data from HTML and XML files, which is beneficial for web scrapping.

### 6.1.2 Selenium:

Selenium is also a python library which is use for web scrapping. It is a tool for testing web applications. This can be done in a variety of ways. For example, allow tapping of the button, enter content in the structure, Skim the website to see if everything "ok" and so on.

### 6.1.3 Working of algorithm:

BeautifulSoup unable to scrap the data from those web pages where JavaScript is disable so in contrast with this we use Selenium which scrap the data in its own chrome browser by calling its module i.e. webdriver which opens the web page in Selenium browser and enable the JavaScript of the web page, but this process takes more time for web scraping as compare to BeautifulSoup. Therefore, we are using combination of Beautiful Soup 4 and Selenium to reduce the data scrapping time. Where Selenium is use to open the web page in its own browser in order to enable the JavaScript and then we use BeautifulSoup to scrap the data from it.

## 6.2 Summarization and Filtration Algorithms:

After getting the raw data by applying scrapping techniques, now the data will get summarized and filtered. For this purpose, we are using transformers and tune them for the training and testing of data. Whereas, transformers are used for changing one sequence into another with the aid of encoder and decoder. Here we are using BERT as an encoder and GPT-3 as a decoder.

### 6.2.1 BERT:

BERT is a machine learning framework for natural language processing that is open source (NLP). BERT is a programme that uses surrounding text to help computers grasp the meaning of ambiguous words in text. The BERT framework was trained using Wikipedia text and can be fine-tuned using question and answer datasets.

BERT (Bidirectional Encoder Representations from Transformers) is based on Transformers, a deep learning model in which each output element is connected to each input element, and the weightings between them are dynamically calculated based on their connection.

### 6.2.2 GPT-3:

GPT-3 stands for Generative Pre-Trained Transformer, it is a unidirectional transformer which takes few demonstrations to comprehend tasks and carry out them. It is built with 175 billion parameters which is 100 times more than GPT-2 transformer. It has abilities like writing articles which are difficult to recognize that whether the article is written by a human or by a computer.

## 6.3 Text-to-Speech Algorithm:

Text to Speech (TTS in short) is used to convert a desired text into a speech form. With more advancement in technology TTS engines became a necessary component to take inputs and give outputs in speech form as opposed to the previous computers which used only the written input from keyboards or typewriters.

### 6.3.1: GTTS (Google Text-To-Speech):

The Google Cloud Text-to-Speech API (Beta) allows developers to embed naturally-sounding synthetic human voice into their applications as playable audio. The Text-to-Speech API converts text or SSML (Speech Synthesis Markup Language) input to audio data such as MP3 and LINEAR16 (the encoding used in WAV files).

URL Links or Article names:

* <https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04> (for transformer)
* <https://www.dataquest.io/blog/web-scraping-python-using-beautiful-soup/> (for BS4)
* <https://huggingface.co/docs/transformers/model_doc/gpt> (for gpt)
* <https://medium.com/walmartglobaltech/the-journey-of-open-ai-gpt-models-32d95b7b7fb2> (for gpt-3)
* <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model> (for bert)