

Intelligent Content Aggregator And Knowledge Synthesizer

By

Manjith Eguturi(20WU0101005)

Hotragn Pettugani(20WU0101030)

Vissarapu Srinath(20WU0101031)

Mentor: Dr. Kiran Kumar Ravulakollu



AGENDA

Problem Statement

Background
and
Introduction

Literature
Review

Use Case

Architecture
Diagram

Methodology

Preliminary
Results

Expected
Outcomes

Gantt Chart

Problem Statement

The problem lies in the time-consuming and often unsatisfactory process of retrieving relevant information from websites due to the vast amount of available content



Background And Introduction

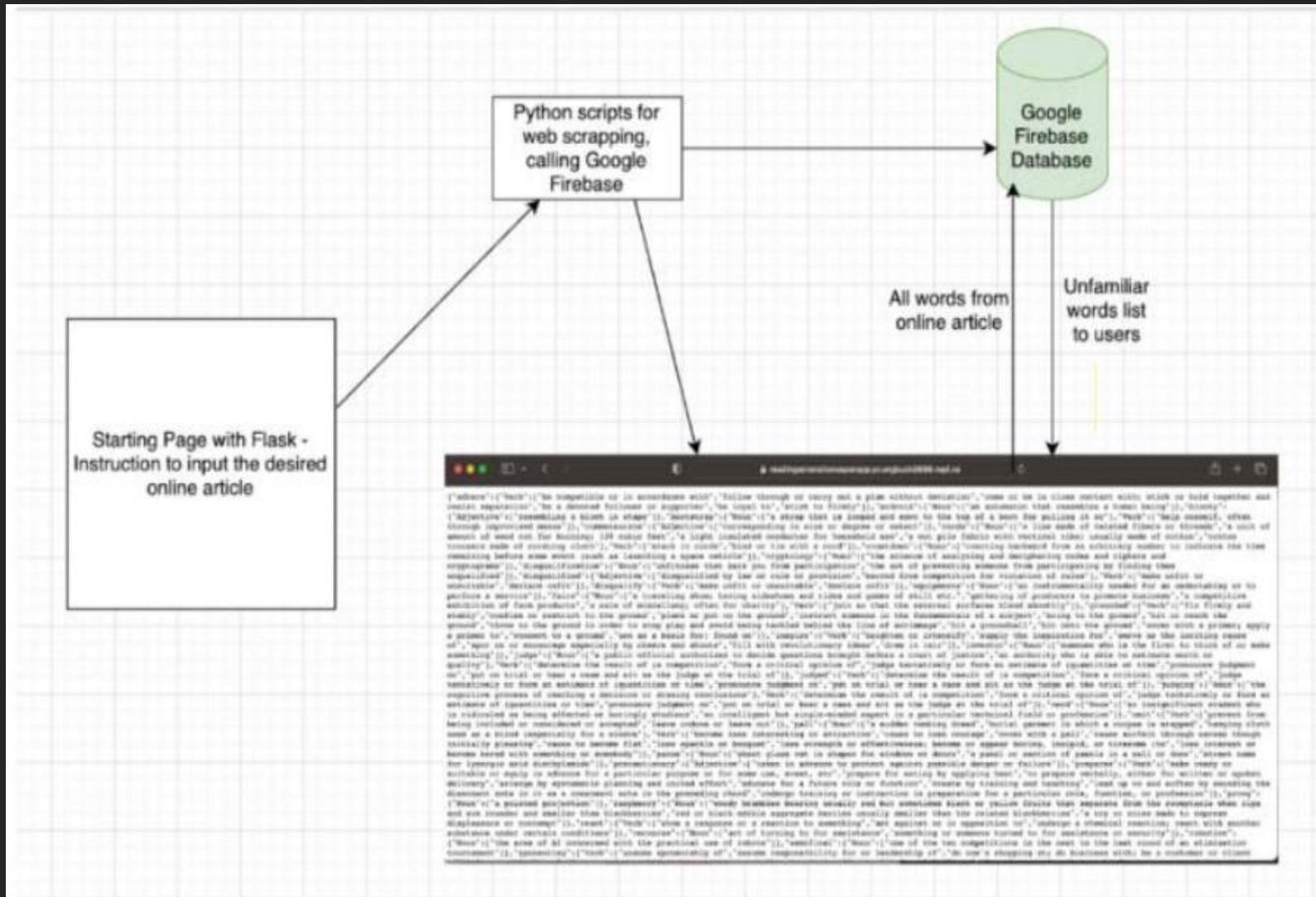
- Copy fish[4], MaxAI[2], and WebChatgpt[3] which can extract text or summarize w.r.t the feature of the extension
- Multiple Written styles from different websites
- Extension that can increase productivity and performance

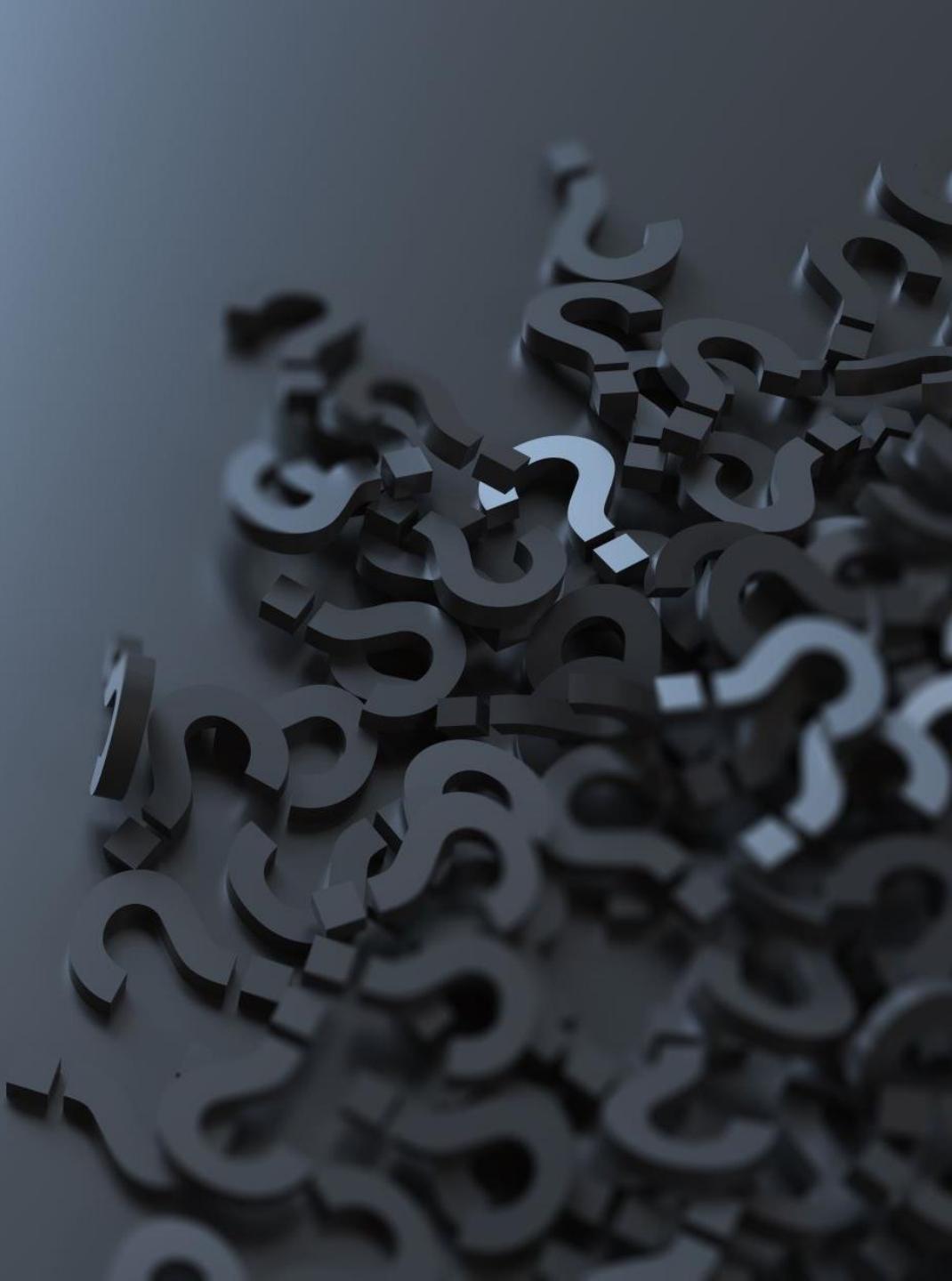


Literature Review [1]

- An extension that has been developed to help users with poor English using the Python dictionaries, Google Cloud Firebase, and Flask. This extension will generate pre-defined difficult words based on an article/website by parsing the content.
- With the information extracted, it will map the definition of words generated to a dictionary and check if the extension can accurately generate the words.
- In addition, this extension will have a track of the most difficult words for users and use this information for future users to have a seamless experience.
- In conclusion, this extension provides a simple understanding of difficult words to users in English.

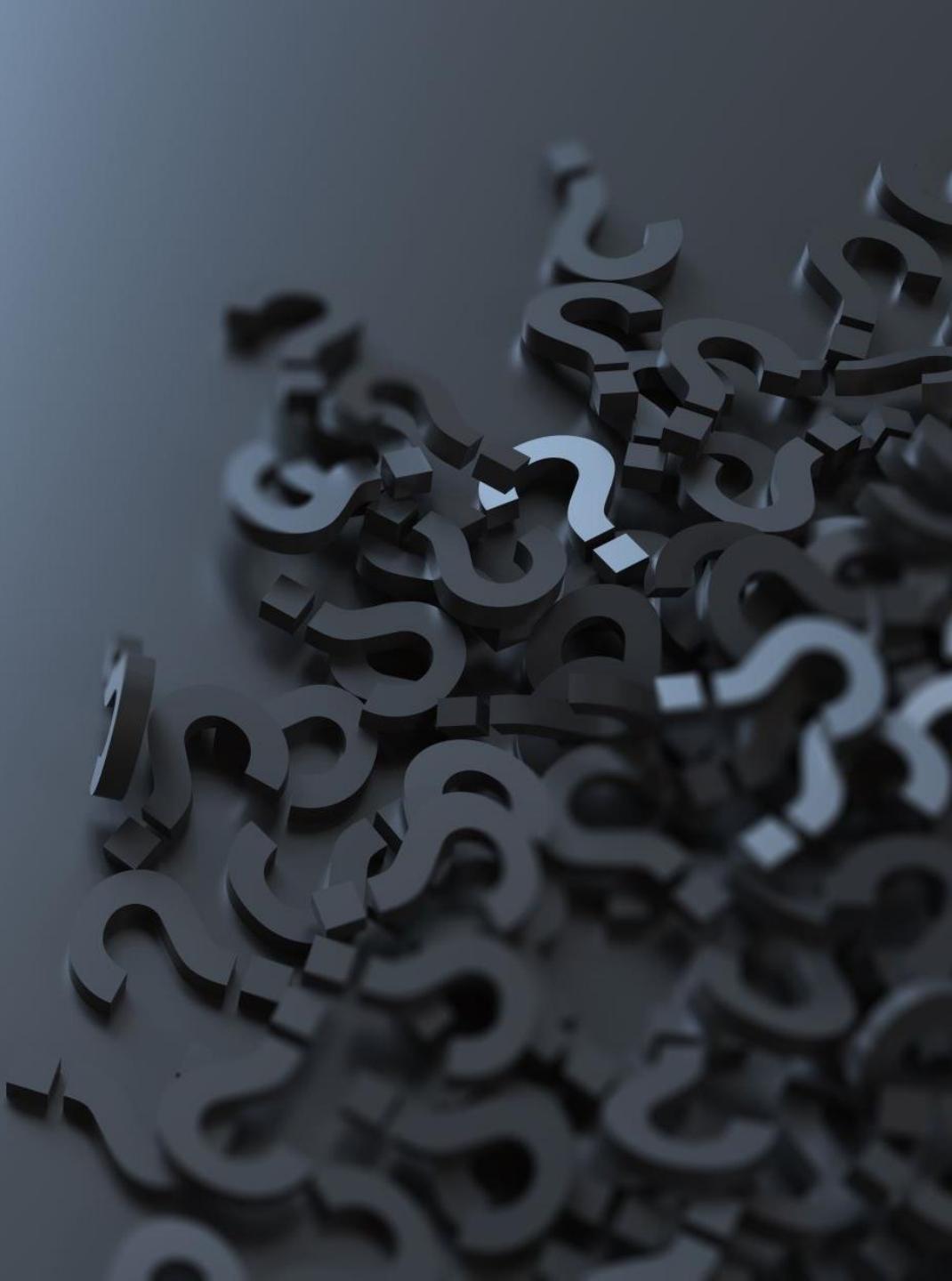
Inferences from the Work [1]





Existing Extension [2]

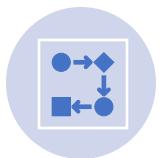
- The MaxAI extension has a variety of features that use most versions of ChatGPT and provide results for the user.
- It helps in summarizing a text from a single page, Generating the content for email, and even as an assistant chatbot.
- It works as an assistant and uses the ChatGPT 4 to produce results only with the premium plan whereas the free plan only provides limited ChatGPT usage per day and a mini menu for a text selected.



Existing Extension [3]

- Given any prompt with web access, the extension can give a response based on the latest news in ChatGPT. It can even scrape data from a URL by restricting responses to a specific URL and giving responses respectively.
- It can perform a one-click prompts wrt required language, Tone, Writing style and Topic for a specific selected article/website.
- Many extensions can replicate the same usage, but the mentioned extensions are the popular ones with most users.

Objectives



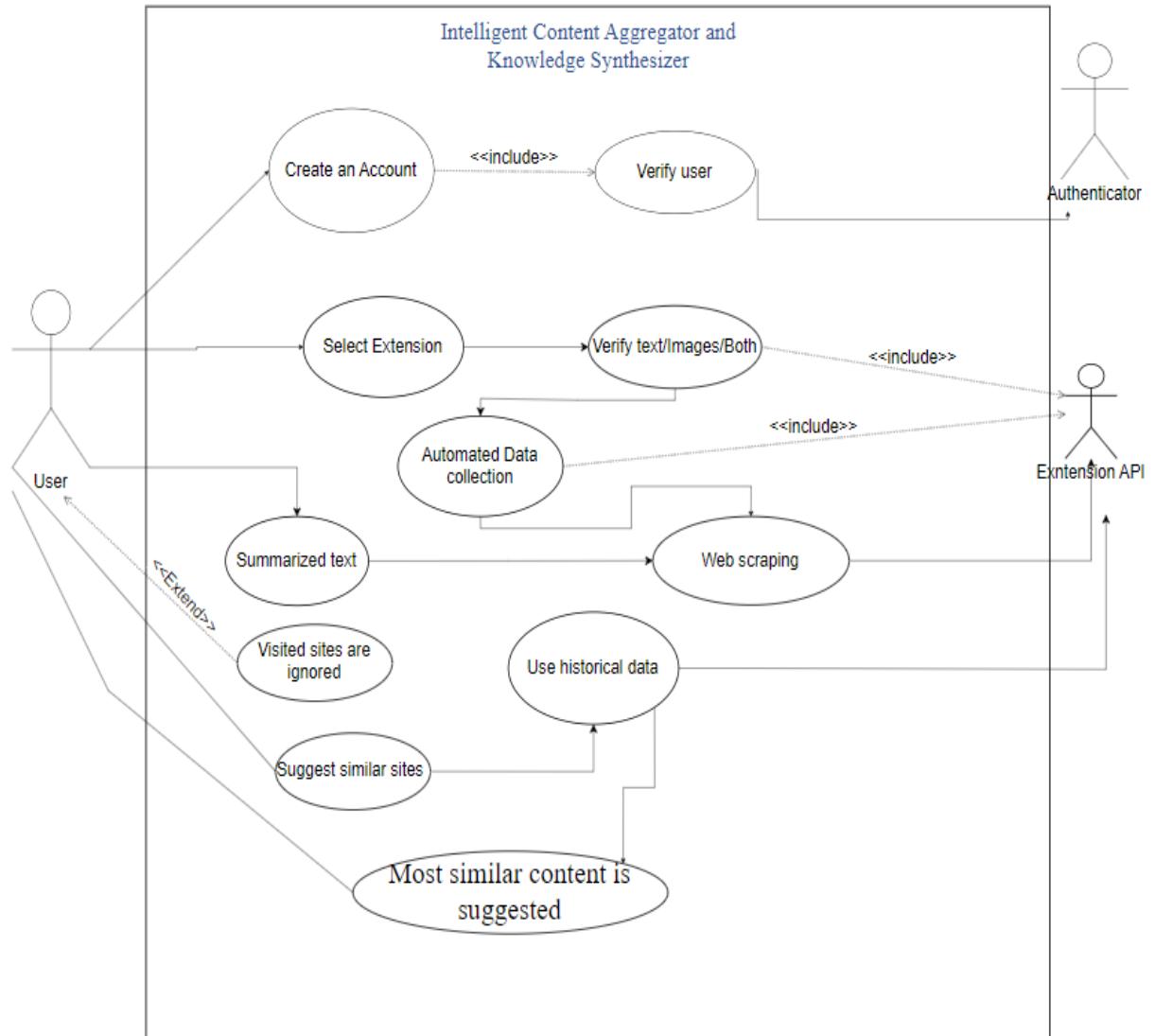
Help users to gather and summarize information automatically from visited websites.



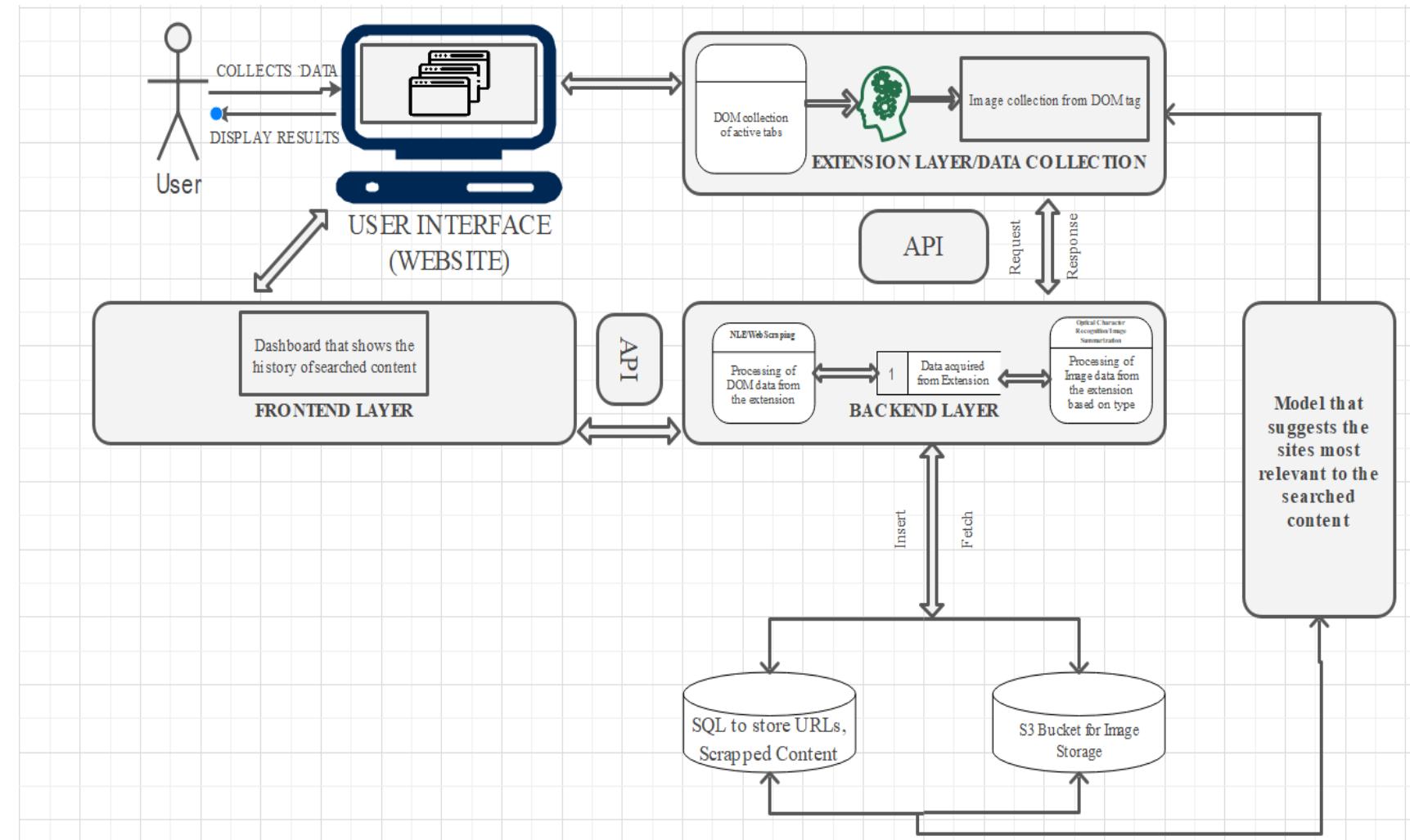
Overcome Challenges:
Redundancy.



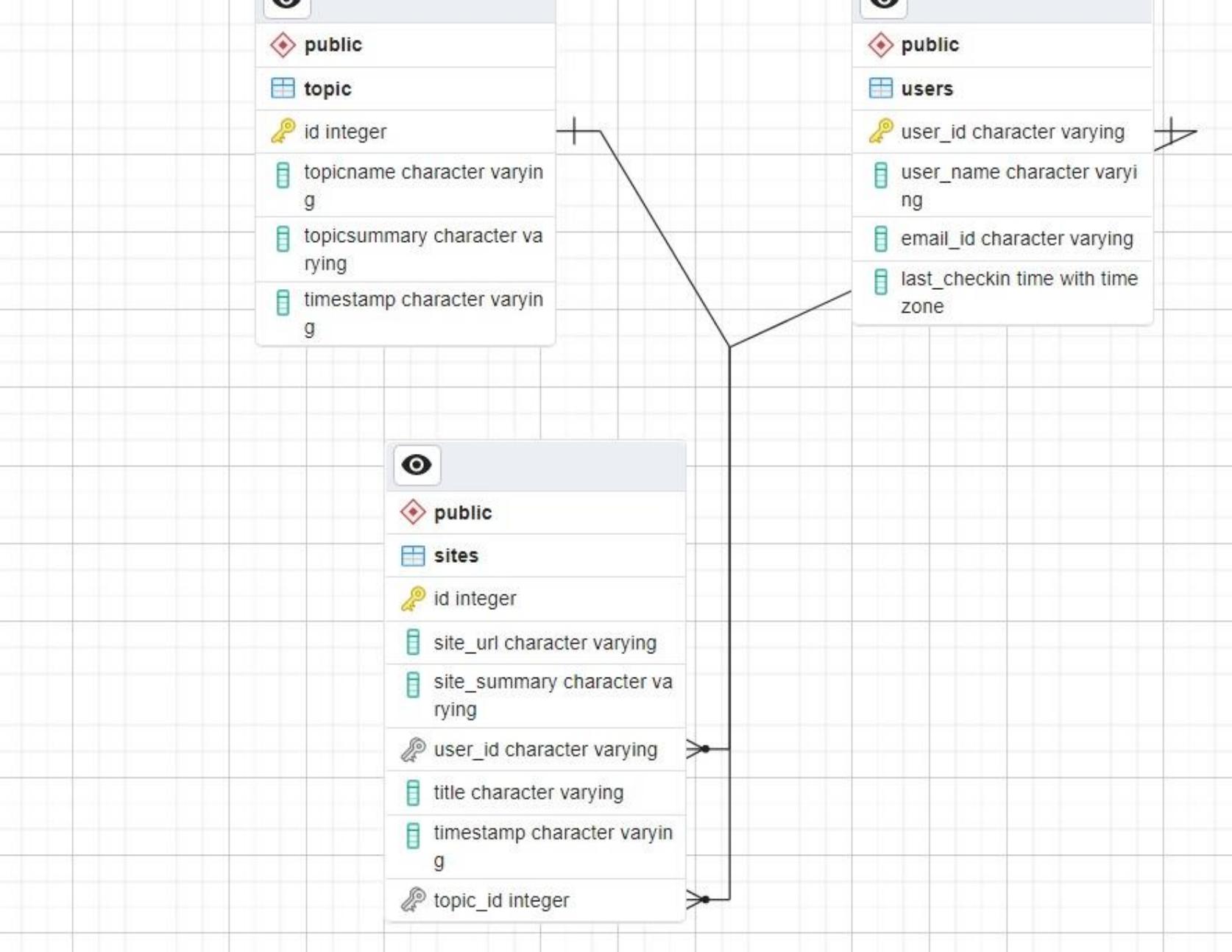
Minimal time to Research
and site suggestions.



System Architecture



Data Base Design





Methodology

1

DOM Elements from the visited pages will be collected through an extension which will be built using JavaScript. Once the DOM content is accessed it will be sent to the backend using an API. image tags will be extracted separately and will be sent to the backend in the form of URL or file format.

2

Model Based Extraction- As all websites do not follow the same structure, we try eliminating most of the unused content in the first step. We will try and build a machine learning model which should be able to separate tags based on their content.

3

Development of an algorithm that can do the below:
Based on the title of the DOM. Once all documents are segregated, the summarization of each title will be performed. Further, there will be a summary of all similar titles together and a separate summary considering all content.

4

A website will be developed to visually track the stats of a user and similar content using the data repository constructed & updated based on extension usage

Preliminary Results

1

The screenshot shows a browser window with a sidebar extension. The main content area displays the text of a Java OOPs Concepts article. A red box highlights the sidebar, which contains tabs for 'Programs', 'OOPs', 'String', 'Exception', 'Multithreading', and 'Collections'. Below these tabs, there are two buttons: 'Get all Text from Paragraph' and 'Send to API'. The sidebar also lists some sections: 'Object', 'Advantages', 'Disadvantages', 'Difference between Object and Class', and 'Object-Oriented Design'. The main content area has a header 'Java OOPs Concepts' and several paragraphs of text about OOPs concepts like inheritance, data binding, polymorphism, etc.

Extension for extracting text from active tab

The screenshot shows a browser window displaying a FastAPI documentation page. The title is 'FastAPI 0.1.0 OAS 3.1 /openapi.json'. Below it, under the 'default' section, is a 'POST /addDOMdata Adddomdata' endpoint. The 'Parameters' section indicates 'No parameters'. The 'Request body' section is marked as 'required' and has a 'application/json' type. It contains a JSON schema with a single field 'domContent' containing the extracted text from the Java OOPs Concepts article. The 'Servers' section at the bottom states: 'These operation-level options override the global server options.'

API for receiving extracted content to the backend

Preliminary Results

2

The screenshot shows a web browser window with the URL javatpoint.com/java-oops-concepts. The page title is "Java OOPs Concepts". On the left, there's a sidebar with a navigation menu for Java topics. The main content area discusses Object-Oriented Programming (OOPS) concepts like inheritance, data binding, polymorphism, etc. A red box highlights a "Get Summary" callout box in the top right corner, which contains a brief summary of OOPS. Below the main content, there's an advertisement for "Stayfree Nights" sanitary pads.

Get Summary

Object-oriented programming (OOPS) is a programming paradigm that uses the concept of "objects" to represent real-world entities, along with their attributes and behaviors. These objects can interact with each other to form complex systems. Some key concepts in OOPS include inheritance, data binding, polymorphism, abstraction, encapsulation, coupling, cohesion, association, aggregation, and composition. In OOPS, inheritance allows one object to acquire all the properties and behaviors of a parent object, promoting code reusability and runtime polymorphism. Polymorphism enables one task to be performed in different ways, while abstraction hides internal details and shows only the necessary functionality. Encapsulation binds code and data together into a single unit, enhancing security and reducing complexity. Objects can be associated with each other through relationships such as association, aggregation, and composition. Association represents a relationship between two objects, while aggregation and composition represent stronger forms of association where one object contains other objects as a part of its state. OOPS offers several advantages over procedural programming, including easier development and maintenance, data hiding, and improved simulation of real-world events. Popular object-oriented programming languages include Java, C#, PHP, Python, and C++. It's worth noting that object-based programming languages follow most of the features of OOPS, but they lack inheritance. Examples of object-based programming languages include JavaScript and VBScript. Additionally, while constructors in Java do not explicitly return a value, they play a crucial role in creating and initializing objects.

Java OOPs Concepts

In this page, we will learn about the basics of OOPS. Object-Oriented Programming is a paradigm that provides many concepts, such as **inheritance**, **data binding**, **polymorphism**, etc.

Simula is considered the first object-oriented programming language. The programming paradigm where everything is represented as an object is known as a truly object-oriented programming language.

Smalltalk is considered the first truly object-oriented programming language.

The popular object-oriented languages are Java, C#, PHP, Python, C++, etc.

The main aim of object-oriented programming is to implement real-world entities, for example, object, classes, abstraction, inheritance, polymorphism, etc.

OOPS (Object-Oriented Programming System)

Object means a real-world entity such as a pen, chair, table, computer, watch, etc. **Object-Oriented Programming** is a methodology or paradigm to design a program using classes and objects. It simplifies software development and maintenance by providing some concepts:

- Object
- Class
- Inheritance

AD

SLEEP WORRY FREE on period nights

Get upto 100% leakage protection with Stayfree Nights

Cottony soft cover

longer lasting

Buy Now

Based on technical and consumer tests conducted by ITC in 2009. Up to 200 days of usage for 8 days. When used correctly it is safe and there is no risk of infection. The product does not contain preservatives.

Summarized content for the active tab

Results

The screenshot shows a web browser window with the URL <https://www.geeksforgeeks.org/java-oops-concepts>. The page title is "Java OOPs Concepts". The content discusses the basics of Object-Oriented Programming (OOPS), mentioning inheritance, data binding, and polymorphism. It also highlights Simula and Smalltalk as early OOPS languages. A sidebar on the right contains links for "Object-Oriented Programming", "Advantage of OOPS over Procedure-oriented programming language", and "Difference between Object-oriented and Object-based programming language". The browser interface includes a navigation bar with links like "Home", "OOPS", "String", "Exception", "Multithreading", "Collections", "JavaFX", "JSP", "Spring", and "Java". There are also buttons for "Get Summary", "Get All Tabs", "Login", and "Sign Up". An advertisement for "Free ChatGPT Extension" is visible on the right.

In this page, we will learn about the basics of OOPS. Object-Oriented Programming is a paradigm that provides many concepts, such as **inheritance**, **data binding**, **polymorphism**, etc.

Simula is considered the first object-oriented programming language. The programming paradigm where everything is represented as an object is known as a truly object-oriented programming language.

Smalltalk is considered the first truly object-oriented programming language.

The popular object-oriented languages are Java, C#, PHP, Python, C++, etc.

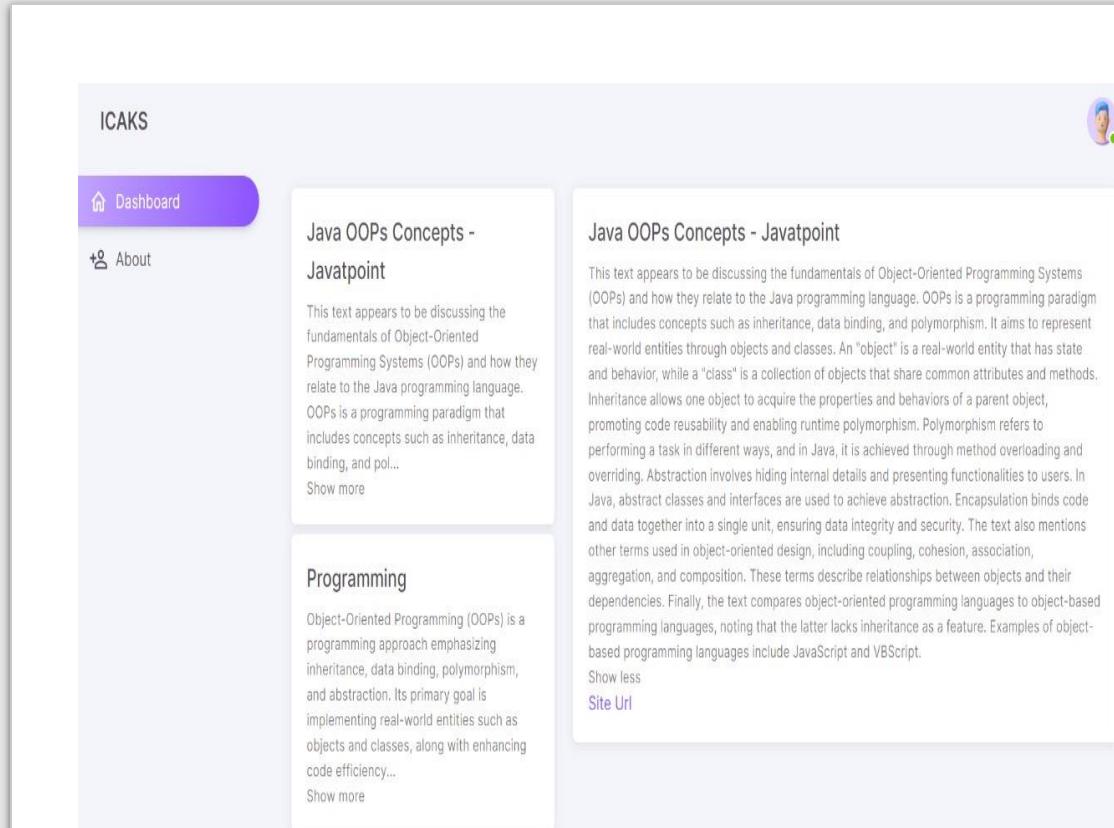
The main aim of object-oriented programming is to implement real-world entities, for example, object, classes, abstraction, inheritance, polymorphism, etc.

OOPS (Object-Oriented Programming System)

Object means a real-world entity such as a pen, chair, table, computer, watch, etc. **Object-Oriented Programming** is a methodology or paradigm to design a program using classes and objects. It simplifies software development and maintenance by providing some concepts:

- o Object
- o Class

Summarization of content from different websites



The screenshot shows a web application interface with a sidebar on the left and two main content areas on the right.

Left Sidebar:

- ICAKS
- [Dashboard](#) (highlighted in purple)
- [About](#)

Content Area 1 (Left):

Java OOPs Concepts - Javatpoint

This text appears to be discussing the fundamentals of Object-Oriented Programming Systems (OOPS) and how they relate to the Java programming language. OOPS is a programming paradigm that includes concepts such as inheritance, data binding, and polymorphism. It aims to represent real-world entities through objects and classes. An "object" is a real-world entity that has state and behavior, while a "class" is a collection of objects that share common attributes and methods. Inheritance allows one object to acquire the properties and behaviors of a parent object, promoting code reusability and enabling runtime polymorphism. Polymorphism refers to performing a task in different ways, and in Java, it is achieved through method overloading and overriding. Abstraction involves hiding internal details and presenting functionalities to users. In Java, abstract classes and interfaces are used to achieve abstraction. Encapsulation binds code and data together into a single unit, ensuring data integrity and security. The text also mentions other terms used in object-oriented design, including coupling, cohesion, association, aggregation, and composition. These terms describe relationships between objects and their dependencies. Finally, the text compares object-oriented programming languages to object-based programming languages, noting that the latter lacks inheritance as a feature. Examples of object-based programming languages include JavaScript and VBScript.

Show less

[Site Url](#)

Content Area 2 (Right):

FastAPI - Swagger UI

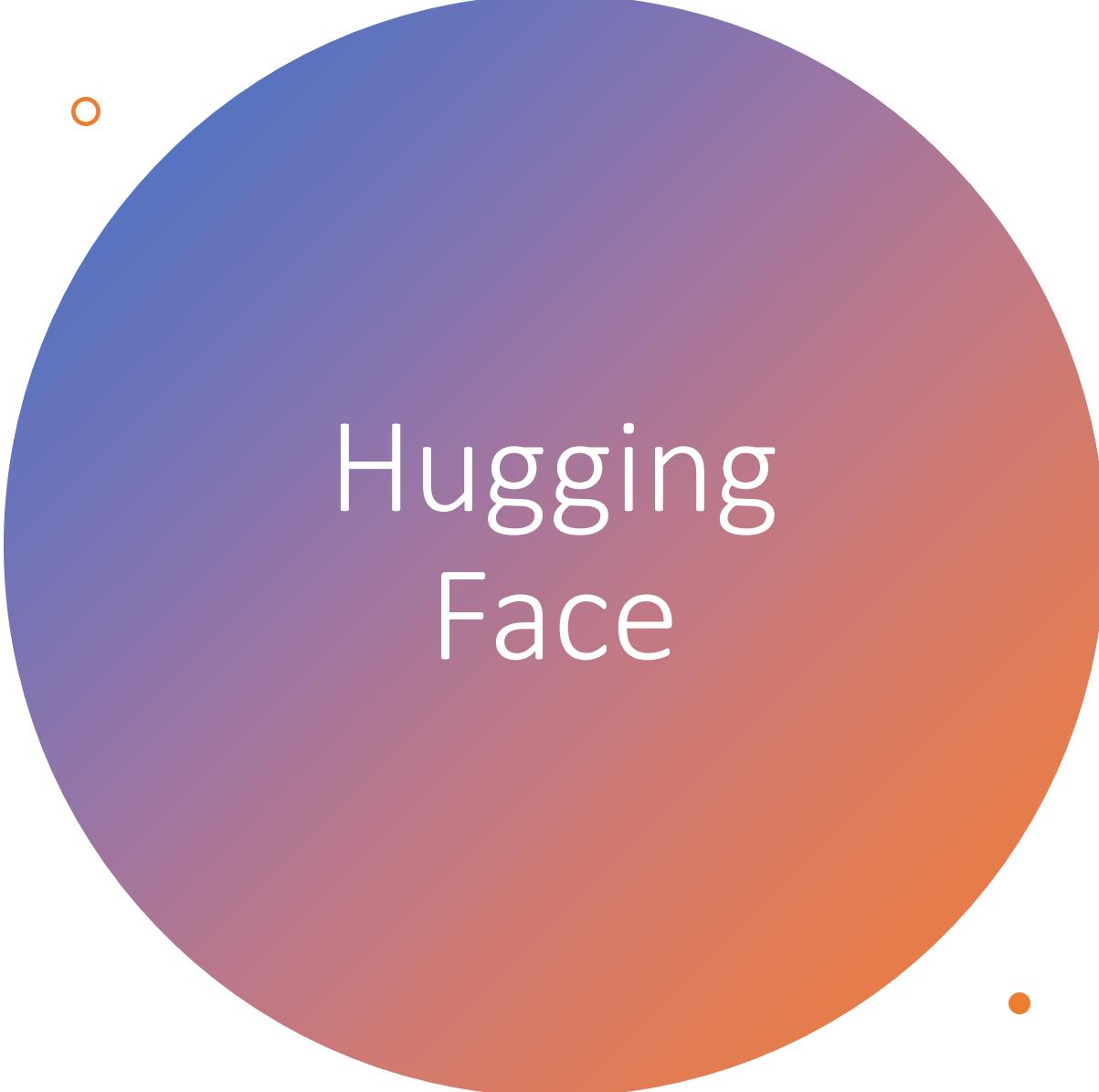
The text describes three topics obtained from GET/getalltopics API endpoint in FastAPI 0.1.0 with OpenAPI 3.1 specification. The first topic is about Object-Oriented Programming System (OOPS) concepts in Java. It explains fundamental concepts of OOPS like inheritance, data binding, polymorphism, encapsulation, and abstraction. It also covers other related terms such as coupling, cohesion, association, aggregation, and composition. The second topic introduces Programming, specifically focusing on Object-Oriented Programming (OOP) and Python. OOP is a programming approach that utilizes inheritance, data binding, polymorphism, and abstraction to enhance code efficiency, maintainability, and scalability. Keywords and constructs in Java, such as abstract class, interface, methods, constructors, static keyword, this keyword, instance initializer blocks, and final keyword, are discussed. Python is described as a simple, readable, and widely-used programming language with extensive community assistance and powerful libraries. The third topic pertains to Beverages but contains only code snippets, styling rules, and configurations for building a webpage, making it unsuitable for summarization. However, the code relates to a template script defining a div structure for menu items with server-side rendering, JavaScript functions handling postbacks and page request management, style declarations for specific DOM elements, interaction with third-party services or plugins, and configuration and initialization of analytics tracking. This response also provides information about response headers and controls accept header with example value schema string and no links schemas. Lastly, there is an embedded Swagger UI bundle code block showing the client-side representation of the API documentation generated using OpenAPI Specification.

Extension Interface



Ablation Study

| Models | Bleu | Rouge | METEO R | BERT | Word Mover's Distance | Self-Bleu | Perplexity vp8=-/ |
|--|------|-------|---------|------|-----------------------|-----------|-------------------|
| MPT(MosaicML_84kTokens) | 0.38 | 0.36 | 0.22 | 0.51 | 0.46 | 0.30 | 80 |
| FLAN-T5 | 0.43 | 0.38 | 0.25 | 0.55 | 0.50 | 0.26 | 76 |
| philschmid/bart-large-cnn-samsum | 0.42 | 0.35 | 0.26 | 0.49 | 0.48 | 0.29 | 79 |
| Hugchat | 0.46 | 0.41 | 0.31 | 0.60 | 0.52 | 0.34 | 68 |
| Falconsai/text_summarization | 0.39 | 0.32 | 0.25 | 0.52 | 0.44 | 0.29 | 84 |
| ARTeLab/it5-summarization-mlsum | 0.40 | 0.37 | 0.29 | 0.54 | 0.47 | 0.28 | 77 |



Hugging Face

- Hugging Face's includes varied pre-trained models such as BART and T5, preparing to different summarization needs across various content types found in active tabs.
- Models on Hugging Face are finely-tuned and consistently achieve top scores across evaluation metrics like BLEU, ROUGE, and METEOR, ensuring the generation of accurate and informative summaries for users.
- Our extension can tailor the summarization process to suit different types of content, ensuring that summaries extracted from active tabs are relevant and comprehensive.

Expected Outcomes

Copy right
registration

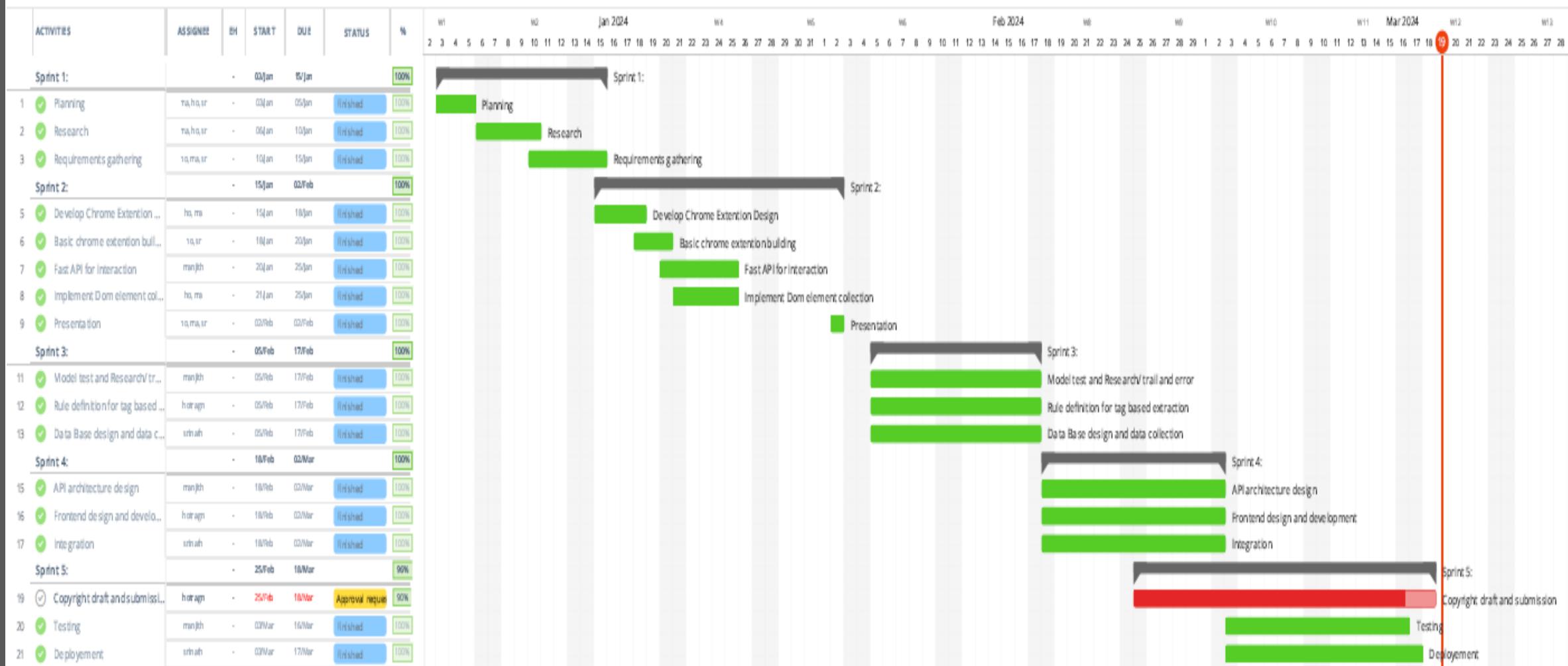
Chrome
Extension

Platform for
management of
researched
content

Gantt Chart

Intelligent content aggregator and knowledge...

Read-only view, generated on 19 Mar 2024



References

- [1] Zhao, Tongde, and Khoa Tran. "A Powerful CHROME EXTENSION: TRANSLATION PROGRAM USING PYTHON, WEBSITE ANALYSIS AND GOOGLE FIREBASE SERVICES." *CS & IT Conference Proceedings*. Vol. 13. No. 7. CS & IT Conference Proceedings, 2023.
- [2] <https://maxai.me/>
- [3] <https://chromewebstore.google.com/detail/webchatgpt-chatgpt-with-i/lpfemeiodjbpkiemkklgpmhlnfgcn>
- [4] <https://chrome.google.com/webstore/detail/copyfish-%F0%9F%90%9F-free-ocr-sof/eenjdnjldapjajjfmlgmkjaienebbj>
- [5] <https://huggingface.co/mosaicml/mpt-7b>
- [6] <https://huggingface.co/philschmid/bart-large-cnn-samsum>
- [7] <https://huggingface.co/chat/>
- [8] https://huggingface.co/FalconSai/text_summarization
- [9] <https://huggingface.co/ARTeLab/it5-summarization-mlsum>

Thank You