

Project Proposal

Web Summarizer and Shortener

COSC 4P02
January 9th, 2024
Naser Ezzati-Jivan

Tanvir Hasan - <i>Product Owner</i>	(6599328)	th18ai@brocku.ca
Muhammed Bilal - <i>Scrum Master</i>	(6695738)	bb18hb@brocku.ca
Hamza Sidat - <i>Developer</i>	(6599591)	hs18so@brocku.ca
Marium Nur - <i>Developer</i>	(7327182)	mn21xu@brocku.ca
Anjali Sabu - <i>Developer</i>	(7337033)	as21qj@brocku.ca
Abdul-Maalik Siddiqui - <i>Developer</i>	(6785828)	as19fa@brocku.ca
Amani Anderson - <i>Developer</i>	(6617344)	aa18ex@brocku.ca

Introduction

Significance:

This project aims to address the challenges users face in accessing relevant information efficiently in the digital age. Navigating through the vast amount of information available online without efficient technologies that provide concise summaries makes information retrieval a daunting task. To simplify access to information, the proposed tool aims to generate shorter URLs for easy sharing and provide succinct and insightful content summaries while employing LLMs or NLP, thus hoping to contribute to a more streamlined and user-friendly digital experience.

Problem:

When looking for quick access to relevant material in the age of technology, people may find it overwhelming due to the huge amount of information available online. Long URLs are difficult to navigate, and the lack of concise summaries may make it more difficult to understand information properly. It is obvious that to improve user experience and make information retrieval simpler, a tool that can create concise summaries and shorten URLs will be required.

Objective:

The objective is to create a program where the input would be a YouTube video or a webpage URL and the output would include a shorter URL for simple sharing. Additionally, the tool will make use of Large Language Models (LLMs) or Natural Language Processing (NLP) to provide short and insightful summaries of web content. Both free and premium aspects of the application would be available, such as user dashboard, social media integration, API access, and custom summarization levels.

Project Description

The Web Summarizer and Shortener project is a comprehensive approach created to tackle the difficulties presented by the vast amount of content available online. The project adds a number of premium features that go beyond basic functionality in an effort to offer a more personalized and enhanced user experience. All things considered, the Web Summarizer and Shortener project has the potential to completely change the way people engage with online content. It makes information sharing and consumption easier by integrating summarizing and URL shortening features.

Below is a brief understanding of what we have to do

Web summarizer and shortener

- *Can take either webpage or youtube video*
- *Shortens URL*

- Provides summary

Features:

- Short link creation - easy, should be able to be done in Python with a few lines
- Summarization:
 - If the link is not a youtube video, then feed the webpage text into the LLM.
 - If is a youtube video, see if has captions
 - If has captions - feed captions into LLM for summarization
 - If no captions - feed video into an autocaption API, then feed transcript to LLM for summarization
 - APIs to consider (AI):
 - <https://replicate.com/fictions-ai/autocaption?prediction=bfhbg5dbuextyddpxrv5ieqmau>
 - <https://kome.ai/tools/youtube-transcript-generator>
 - The user must be able to copy the summary and the shorten URL through the click of a button
 - Pro-feature may allow for downloads
- Custom Summarization Levels (pro-feature):
 - Need to ask the professor what this means for clarification
 - Perhaps the amount of words, how detailed it should be, the language used to summarize (formal/informal)
- Social media integration (pro-feature):
 - Need to ask what this means as well
 - Perhaps to be able to log in with your Google/Facebook account?
 - Or to be able to take Google and facebook links to summarize?
- API Access (pro-feature):
 - Need to ask what this means as well
 - Create an API for our project?
 - Access to the APIs that we used?
- User dashboard (pro-feature):
 - Self-explanatory, we save the links and summarization history, save their settings, and have a user interface that's functional.

Software engineering process and technical tools

We will be using the Scrum approach for this project. Because the Scrum method is iterative and allows for frequent product increments, it is advantageous to use it when developing the tool. The framework promotes team member collaboration, responsiveness to changes, and customer input. The success of the tool is influenced by releasing the product incrementally, adopting technology with flexibility, and prioritizing additions according to their significance.

Each sprint would be 2 weeks. Based on the product backlog created, a sprint planning meeting would help create a sprint backlog. Sprint Retrospective and Review meetings would be held at

the end of every sprint to discuss what went well/wrong during the last sprint and how the next sprint can be improved. The choice of technical tools and the planned two-week sprints demonstrate a commitment to flexibility, adaptability, and continuous improvement throughout the development lifecycle.

GitHub page:

Meeting Schedule

Weekly meetings every Tuesday from 3:00 PM to 4:30 PM

Week Of	Goals for completion
Jan 8	<ul style="list-style-type: none"> Discuss deliverables, the project topic and meeting schedules Create a GitHub page, discuss free and pro features, and timetable, review the doc and update as needed.
Jan 15	<ul style="list-style-type: none"> Upload all docs to Brightspace and GitHub. Project Proposal due on Jan 16
Jan 22	<ul style="list-style-type: none"> Write down user stories and create a product backlog Work on a Sprint backlog Release Planning document due on Jan 26
Jan 29	<ul style="list-style-type: none"> Begin Development Sprint Planning Meeting on Jan 30
Feb 5	<ul style="list-style-type: none"> Sprint Retrospective meeting on Feb 9 Have an improved plan for the next sprint
Feb 12	<ul style="list-style-type: none"> Begin the second sprint development
Feb 19	Reading Week
Feb 26	<ul style="list-style-type: none"> Sprint Retrospective meeting on Feb 27 Sprint Review meeting (with TA) Test the first version of the web summarizer tool Release the first version Progress Report 1 due on Feb 27
Mar 4	<ul style="list-style-type: none"> Being the third sprint development
Mar 11	<ul style="list-style-type: none"> Sprint Retrospective meeting on Mar 15
Mar 18	<ul style="list-style-type: none"> Being the fourth sprint development Sprint Review meeting (with TA) Progress Report 2 due on Mar 22
Mar 25	<ul style="list-style-type: none"> Test the second version of the web summarizer tool

	<ul style="list-style-type: none"> • Release the second version • Sprint Retrospective meeting on Mar 29
Apr 1	<ul style="list-style-type: none"> • Begin fifth sprint development • Testing the final version • Release the product
Apr 8	<ul style="list-style-type: none"> • Prepare a how-to-use guide along with the presentation
Apr 15	<ul style="list-style-type: none"> • Final Report due between 15th and 28th • Final Demonstration/Presentation between 17th to 26th