Testing Documentation

Project Title	SUM-UP
Module	CA400 - Final Year Project
Student 1 (Name & ID)	Zubair Asif (18314236)
Student 2 (Name & ID)	Alif Hossain (17314941)
Project Supervisor	Gareth Jones (gareth.jones@dcu.ie)
Date of Submission	06/05/2023

Heuristics Testing:

The process of heuristic testing involves specialists using general guidelines to assess the usability of user interfaces during independent walkthroughs and indicate problems. Evaluators make use of well-known heuristics and provide insights that might aid design teams in improving the usability of products from the very beginning.

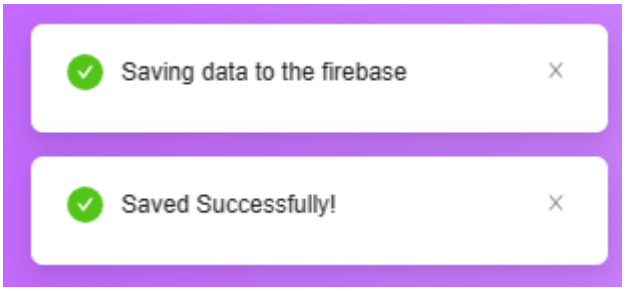
We undertook Heuristics testing/evaluation in order to optimise usability and at the same time eliminate design deficiencies.

Visibility of System Status

By generating messages that reassure the user of their activities, we ensured that our design constantly keeps the user informed of what is happening. Below are some instances.

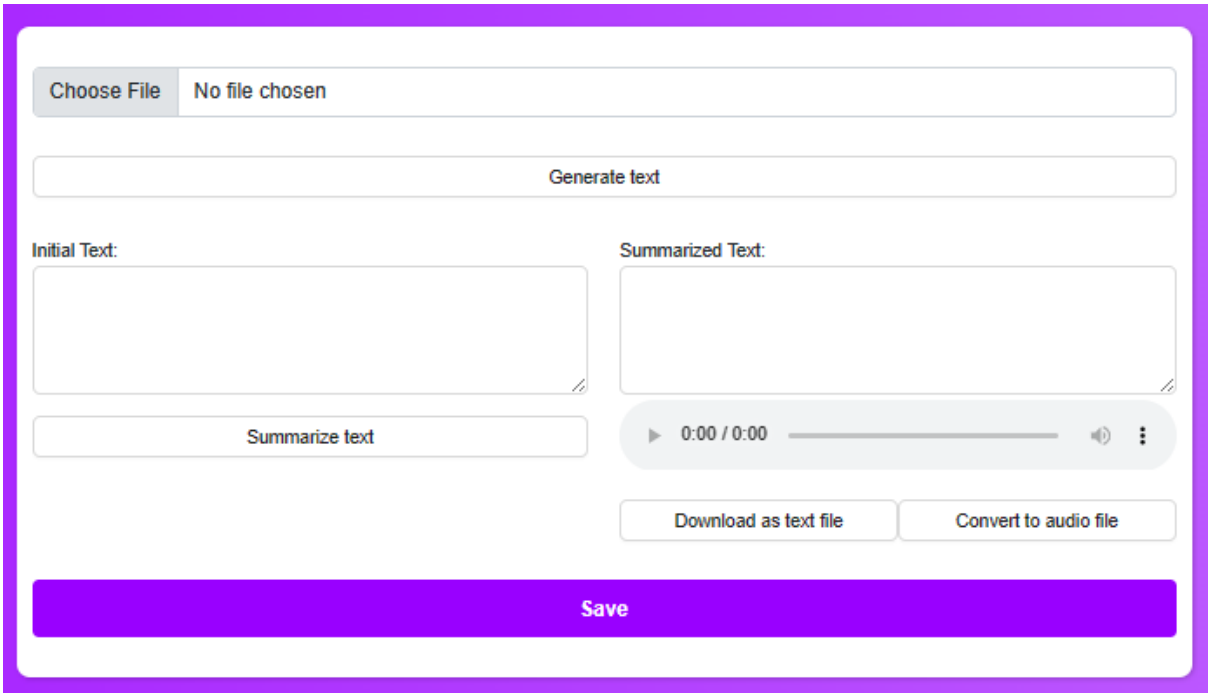
- When data is successfully summarized, we get the following prompt;

- When data has successfully been saved to the database, we get the following prompt;



Recognition rather than recall

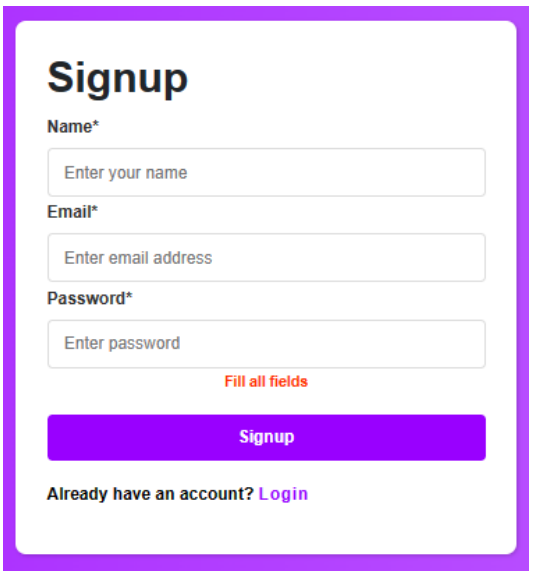
Our goal was to reduce the amount of information consumers need to recall in order to utilise our app. We made sure every page should appear familiar to users so they can navigate it easily to complete tasks. This heuristic was satisfied by properly labelling fields and menu items and putting them in places where they are simple to see and find. We maintained most of our functionality within the following container allowing for a simple and easy experience, with both consistency and familiarity.

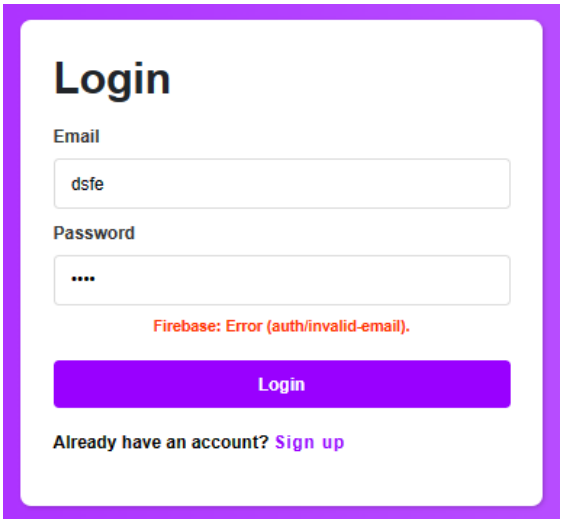


Error Prevention

Error messages were effectively used in our application. We must develop designs that stop user problems in their tracks in order to completely eradicate them. For places where there is a high potential of user error, we added a number of checks to ensure that users supply accurate information. This is done, for instance, on our Login, Register and on our Home pages.

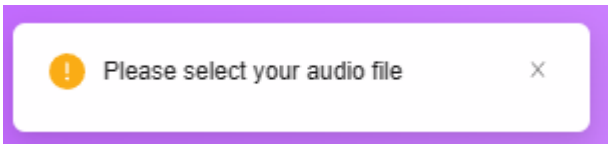
As you can see from the Signup page, if the user fails to supply either the Email or Password they will be presented with a prompt warning them that they must fill all the fields in order to proceed. Likewise, with the Login page if the user enters the incorrect credentials they will be told about invalid authentication email etc.



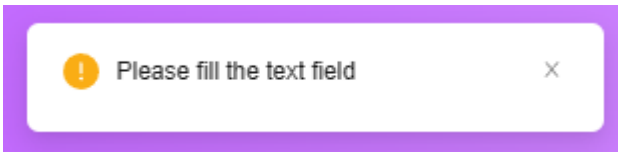


For the home page we have put in place a variety of checks in order to maintain order and provide error prevention. These measures have come in handy as I prevent mistakes from occurring and confusion.

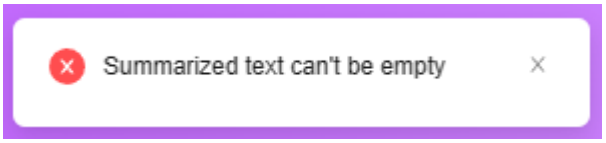
- If the user forgets to choose an audio file before proceeding to generate the text they will be prompted to correct their action.



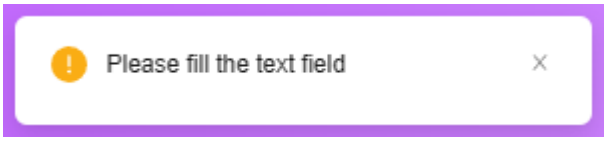
- If the user forgets to generate text before summarising they will be prompted with the following message



- If the user forgets to summarise the text and goes straight to the download options; “Download as text file” or “Convert to audio file”



- If the user decides to Save the empty fields they will be prompted to the following message



User Testing:

Introduction

This report presents the findings of a test that the DCU GDPR committee ethically approved, as well as data gathered after users tested our app. We made every effort to ensure ecological validity because we wanted this test to be as natural as possible due to the nature of our application.

In order to evaluate the usability and efficiency of our web application, user testing is a crucial activity. This user testing session's main objective was to assess our web application SUM-UP that provides the ability to perform extractive summarisation on audio media. To enhance the overall user experience of the program, it was intended to comprehend user viewpoints, detect usability concerns, evaluate the level of summaries, and gather feedback.

The following facets of the web application were to be evaluated during the user testing session:

- User Interface: Assessing the application's interface's style, organisation, and usability.
- Execution of Task: Examining how users interact with the program, add audio files, start the summary process, and evaluate the summaries that are produced.
- Summarising audio Evaluation of the application's audio-to-text conversion and summarising algorithms' accuracy and efficacy.
- Understanding the general happiness, usability, and any difficulties users may have had during the testing process.
- Feedback and ideas are sought from users in order to enhance the functionality and user experience of the program.

Test Participants

For the user testing we gathered a number of DCU students and informed them of the project and its details. We made sure that the participation selection was diverse and representative of the final target audience. We gathered students by sending out a request via academic servers and once we were aware of an indication of interest, we provided further details to the subjects with all the required documentation in order to make sure they were fully aware of what the testing entails. The users had different levels of understanding of the topic of summarization. We informed them of the stages involved within the testing phase and confirmed their participation by having them sign the consent forms.

Once all the paperwork was complete we moved onto the actual testing of the application. We booked a room in the library where we could have a private space in order to run the tests and allowed the subjects to tinker with the application and as they did we viewed their responses and behaviours. The application was kept the same on the same device with all users having the option and the same variety of audio files in order to keep everything

consistent and accurate. Once we received confirmation that the participants were satisfied with the amount of time they have spent and were content with usage we concluded the physical tests. The users were then surveyed with the questionnaire that reflects the overall experience and usage of the app along with the comprehensibility, coherence and relevance of the summaries.

Ultimately, we gathered a total of 4 participants for the user testing. The participants' anonymity was maintained with the highest level of confidentiality during the entire process.

Goals & Objectives

We wanted to understand as much as we could about our application. We needed to be aware of the application's advantages as well as, and perhaps more importantly, its flaws. The objective was not to hide issues and faults but to bring them to light.

Survey Questions

1. **Do you believe there is a need for an application such as SUM-UP?**

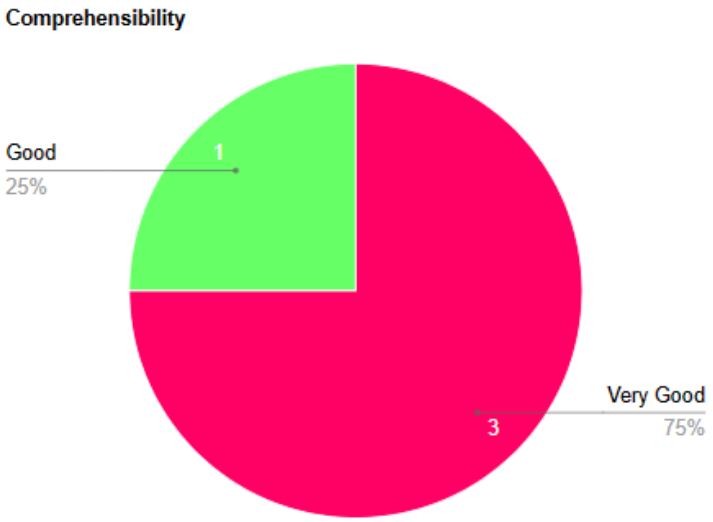
   We were pleased with the survey findings that 100% (4/4) of the respondents thought the application proposal was fantastic and that it was needed in a variety of ways. Some intended to use it for its designed use whereas others had some creative ideas where the application could be of potential use. This demonstrates that users are interested in a program like this, indicating that there could be high market demand for such an application.

2. **What is your first impression of the application SUM-UP?**

   The main comments on first impression of the application were all positive. They focused primarily on the design and layout of the application, particularly mentioning the simplicity in usage.

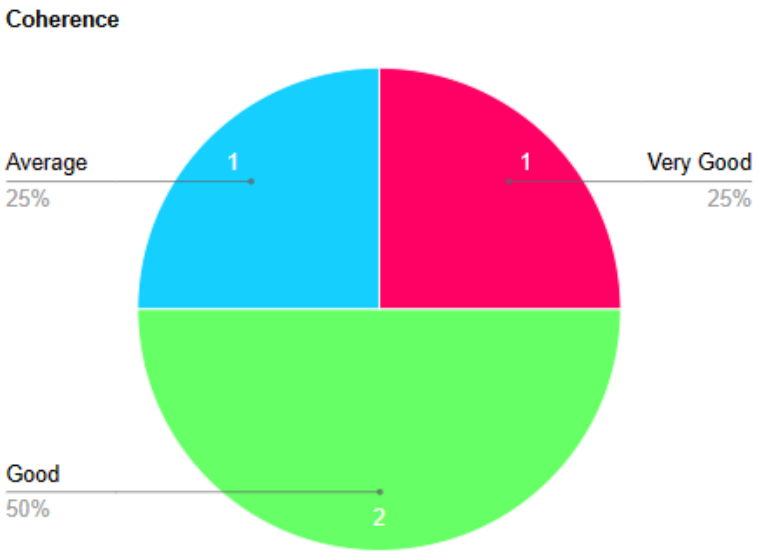
3. **How comprehensible are the summaries?**

   The comprehensibility of the summaries averaged as very good with 75% vote in favour of it. This is important as for a summary to be correctly evaluated it must fall within the criteria of being easily understandable and comprehensible.



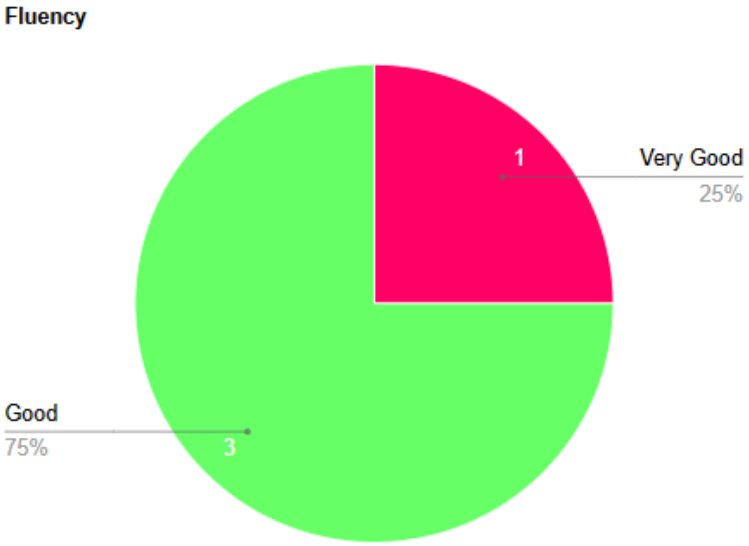
4. **How coherent are the summaries in reflection to their source?**

   This question provided a good indication of how well our summariser performed as this is also a key evaluation metric. The results showed that the majority of the participants believed that the coherence was above average.



5. **How fluent are the summaries in terms of grammatical correctness?**

   Correct fluency is vital as without it, it can reduce the credibility of the overall summary. The results showed that the majority believed that fluency was good, with one vote for very good. No participants rated the fluency as average, slightly poor or poor.



6. **Do you believe you would benefit from an application like SUM-UP? If so please provide reasoning.**

   From our user testing we have found that the majority of the people believe they would benefit from our web application SUM-UP. Users have indicated that the application would help them ***"with note taking and summarising of videos and audios."*** Others have mentioned that they can now ***"enjoy more media content that I***

normally would have passed up on due to length or time". One of the users mentioned that it would help them in terms of educational purposes, in that being able to summarise recorded lecture audios for a quicker/efficient understanding. One student mentioned that they have a correspondent who has learning disabilities and uses a microphone to record audio of lectures for later study use. For these types of situations SUM-UP becomes a fundamental tool for your educational arsenal.

What is important to make note of is that not all of the users who we surveyed found our application beneficial. One mentioned that due to the lack of support for a variety of audio/video files, they do not find the application useful. As the application only accepts one format, it would limit its uses.

7. **Do you feel there are any negatives of using an application such as SUM-UP?**

   Few of the users mentioned the significant amount of time needed for the application to process the audio to text conversion. They mentioned that for bigger files the application tends to stall a while before producing a result. They mentioned that in today's technological world where users have access to instantaneous quick media, in the form of videos, music etc, peoples attention span is significantly reduced and therefore they become less interested in waiting.

   One user mentioned that the application could use more features that can further develop the application and incorporate more creativity.

8. **What was your experience whilst using the application in terms of usability?**

   Majority of the users found the application easy to use & navigate, reliable, clear in its instructions and without any complex or confusing options or subsections. Mentioning that "Everything made sense". They mentioned that the application provided clear labelling and it made user testing very do-able.

9. **Would you use or recommend using SUM-UP to others?**

   All but one of the users said that they would recommend the web application.

10. **What changes would you implement into the application? What would you like to see in future developments?**

    Many of the users mentioned the same changes, which included a quicker initial translation of text, they mentioned that for longer audio files it would take a significant amount of time to convert it. Some further mentioned providing more features to do with customisation or even video summarisation.

Operation Time Testing:

After receiving the results from the User Testing, we noticed that the main flaw that most of the participants mentioned was the Operation time for generating the text from the audio file. In order to further understand the reasoning behind it we decided to run some tests and time the audios based on the lengths.

Audio Length (min)	Operation Time (min:sec)
1	00:20
2	00:40
3	00:50
4	01:10
5	01:20
7	02:00
9	02:35
16	03:50

The results showed that there was a correlation between the length of the summaries and the duration of the operation time. However, we did notice from the results that the quality of audio also played a significant part in the operation time. The clearer the audio quality the quicker the summary would occur and it would exceed the expected times. A potential solution to this issue would be to introduce segmentation.

Conclusion:

The testing went well in general. We obtained a wide range of findings. These consequences were not totally positive, and we learnt a lot from the criticism. What it accomplished was determined via testing. That was done to bring to light issues that we would not have noticed otherwise. Doing the Heuristic testing first was a good decision as it provided great results from the comments received from the User testing. The Technical Specification contains further information about future intentions to make these enhancements.