Project Brief Form 2022-23

Student Name	**JONAH WASILEWSKI**	*Id Number*	**F122151**
Supervisor	**PETER NORRIS**		
Programme	**Computer Science with Artificial Intelligence Bcs**	*Module code*	**23COC257**

ALL BOXES IN THIS SECTION SHOULD BE TICKED.

☒ The attached project brief is agreed between the supervisor and the student as a description of the proposed project work.

☒ A provisional workplan is attached (on subsequent pages.)

☒ The Ethical Awareness Form (EAF) has been filled and attached (on subsequent pages.)

☒ If any answers on the EAF have been marked as **Yes** (except for the first), then a formal application has to go through LEON (leon.lboro.ac.uk).

Part C students to highlight/mark the column matching their project module code (remove other X's).
Part D (COD290) students to highlight the column matching their part C project module code.

Project Code --> Project Title --> Programme(s) -->	COC251 CS (CS)	COC252 Computing (C+M, ITMB)	COC253 IT (C+M, ITMB)	COC255 CS+Math (CS-Math)	COC257 CS+AI (CS-AI)	COC800 Software (China)
Requires CS content.					X	
Requires maths content.						
Requires AI content.					X	
Requires IT content.						
Requires a real customer.						
Requires Novelty						

Student signature:	J.Wasilewski

Loughborough University

Project Description

I will be researching, designing and implementing a web browser for this final year project. For this to be successful, I will need to either create or use and existing web crawler to create an extensive 'crawl list' containing the URL, title and keywords of websites that have been deemed relevant. Then, when the user enters a search into the engine, an appropriate list of websites are returned to the user by searching through the crawl list to find matches. The easier option is to use an API that can be called whenever needed, but this might be problematic if it doesn't fit into my program correctly and so a newly made web crawler would be a lot more tailored to my needs. This is the fundamental base for the program and is therefore the most important part of implementation. For the AI portion of the project, I will add extra functionality to the crawlers that observe different properties and characteristics of the webpages it finds so that it can then return a more relevant list to the user at the end. This will be done by using machine learning to analyse the user's previous searches and make a user profile. Furthermore, the AI could be adapted to have additionally functionality, for example it could read through the top result and create a small summary for the user.

I will be using python to code this project because it allows me to use some very helpful and insightful artificial intelligence elements as well as being a language that I am very confident in using. Furthermore, although the user interface aspect of the project is not something that I see as incredibly important, I will make use of some library in python that allows a graphical interface to be used that can display the results to the user in a clear and precise manner. Upon early research, a library like Tkinter or PySimpleGui would fit my needs for this. Alternatively, another option could be for me to create the user interface using web based programming languages and display the data to the user that way. Logically this makes sense because all web browsers are typically run on the internet (and hence on a user's web application) and it may make the response time faster because data wont have to be passed from the internet to my python program quite as much.

The early stages of development will involve mostly research for the program so that I can decide the best method in which I can implement the problem. For example, testing out different available APIs or ways in which I can make my own web crawler from scratch. Additionally, a big part of the planning process will be ensuring no ethical issues arise, or at the very least they can be accommodated for. There is an element of data collecting and processing from the user and so I will need to make sure that all this is done responsibly and with the complete knowledge of the user. I will be looking at existing search engines (e.g. google) to draw inspiration and investigate what elements work well and why they are used. This, of course, does not mean copying designs or solutions to problems but it does mean that I will try to incorporate elements that I particularly like or feel that would work very well in my solution to the problem.

A realistic timeline for the project would be:

- Planning and research until the middle/ end of November
- Basic implementation by end of term 1
- Presentable prototype by January submission date
- Continue to add additional functionality (mainly AI elements) and refine search algorithm/ web crawler until the final date

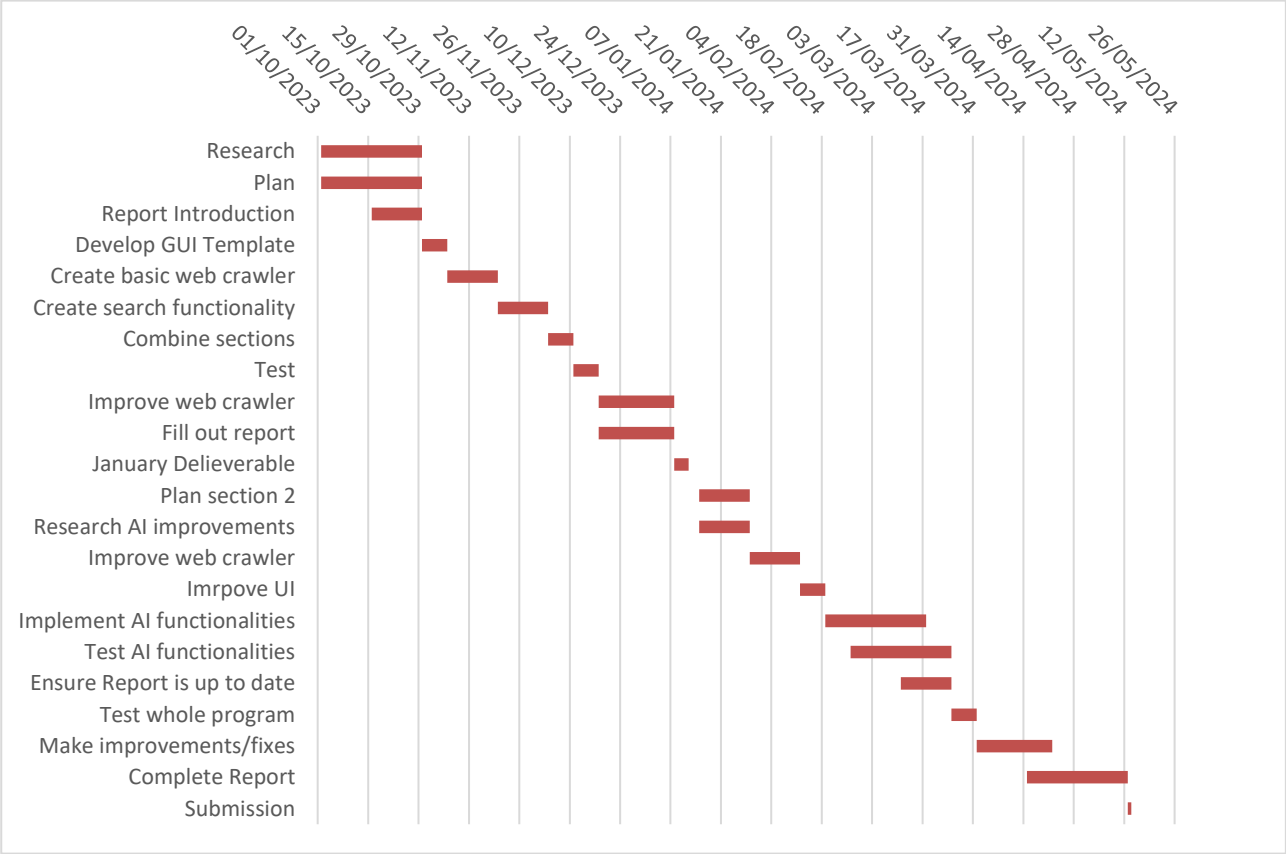
It is important to have a way in which I can measure the success of the project at any stage, especially at the end so that I can reflect on what I achieved. Creating a functional search engine that has at least one AI element to it is the main aim otherwise I won't have fulfilled the proposal. Hence, I will work towards making a search engine that returns relevant and precise information to the user based on their query, and I will do it in such a way that is ethical and returns the results are based on the user profile that is made by the AI. After this, there is lots of extra functionality that I can add to further improve the program. This project obviously has lots of potential but it will be important for me to make sure that I don't get carried away and only make what I have set out

to do otherwise it will be too big to manage and, most likely, I won't be able to finish the project before the given deadline.

The performance of the search engine will be tested by having different users interact with it using different accounts so that the AI can accurately learn what their search habits are and give them a tailored experience. The challenge will be to ensure that a significant number of test users trial the system so that the capabilities of the machine learning algorithms can be fully judged. User feedback after these tests will also be very important so that I can alter the system if any problems are brought to my attention or I can focus on parts that were a success.

The speed of the search engine is also very important because I user wants results very quickly and doesn't want to be sat waiting for the results to come through as this is not the case with other engines. A way in which the speed can be improved is to use a txt file to store previously crawled websites so that they don't have to be processed again, rapidly increasing the speed of the browser. Tests for the speed of the search engine are very simple because I can simply record the time that it takes to return results to the user. A comparison can be to the speed of a google search but I do not expect it to be as fast as that so as long as it's relatively close I will consider the test a success.

Workplan with Gantt Chart



Description of Tasks

- Research – Find relevant research articles and real world examples of similar projects to learn the best methods, techniques and approaches. Will help decide how I go about the whole project
- Plan – Use my research to create a structured and detailed plan that I will follow through the development process
- Report Introduction – Create my report template and fill in any sections that I can, mostly areas where I talk about planning and research
- Develop GUI Template – Make a basic UI that I can use as a basis for displaying information to the user, will mostly likely be changed as the project progresses
- Create Basic Web Crawler – Using my research from earlier, I will implement the first version of the web crawler. The main aim will be to get a program that crawls the internet and returns a list of the websites it goes to and find relevant. The selection may not be perfect at this stage
- Create search functionality – Link the UI to a module that will process the user's search query and return a list of websites that are relevant. May use dummy data at this stage
- Combine sections – Ensure that the main components work successfully when linked with one another. Ie the search program can read the list of crawled websites and return a list of websites that relate to the search
- Test – Measure the performance of the search engine against the desired functionality that I want at this point in development
- Improve Web Crawler – Make any improvements needed after the tests

- Fill Out Report – Ensure that the report is up to date and contains all information needed ready for the January deliverable
- Plan Section 2 – Assess my current progress and decide on a plan of action for the rest of the development process
- Research AI Improvements – Look for the best ways in which I can use AI to make my project better and more user friendly. Make a shortlist of techniques that I think will be the best fit for my program.
- Improve Web Crawler – Make a final assessment of the web crawler's success in creating a good crawl list and make any changes/ fixes needed
- Improve UI – Refine the UI to make it look more professional and like a finished product that a consumer would want to use
- Implement AI Functionalities – Incorporate the AI elements into the program
- Test AI Functionalities – Will be done at the same time as implementing the AI elements so that I can see how each of them works straight after adding them in
- Ensure report is up to date – After adding the AI elements to the program, I need to make sure that everything has been documented and well explained
- Test Whole Program – Do a full scale test of the program to look for any errors I may have missed
- Make Improvements/Fixes
- Complete Report

Ethics Awareness Form for Taught Student Projects

Project Title/Topic: Search Engine

All students should discuss with their supervisor whether their project might conflict with the University's ethical principles which can be found in the Ethical Policy Framework.

Students should complete the second column in the table below, discussing with their supervisor as appropriate.

Aspect of project	**Does the project involve this aspect?** **(Yes / No)**	**If Yes, follow the process(es) below**
Analysis of secondary or pre-existing human data which **does not** require ethical review based on the Guidance Note: Studies Using Secondary or Pre-Existing Data	**Yes**	No further ethical review required.
Analysis of secondary or pre-existing human data which **does** require ethical review based on the Guidance Note: Studies Using Secondary or Pre-Existing Data	**Yes**	Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations involving human participants	**Yes**	Follow the Code of Practice on Investigations involving Human Participants. Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations involving activity falling under the Human Tissues Act	**No**	Follow guidance from the Human Tissue Act Licence Sub-Committee

		If necessary, complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations with military applications or using dual use technologies	**No**	Complete the Review Process for Projects Involving Research with Military Applications or Dual Use Technologies.
Investigations involving animals or animal cells/tissues	**No**	Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations involving accessing security sensitive material (e,g, online terrorist content or material).	**No**	Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Possible conflict with ethical principles partially or wholly outside the above.	**No**	Forward a study description to the Dean of School or the Responsible Person (see Ethical Policy Framework for list)

Student Declaration

I confirm that I have discussed the ethics awareness form with my supervisor and, if appropriate, followed the relevant guidance / made the relevant application.

Student name: Jonah Wasilewski .....

Student ID number: F122151.....

Signature: J.Wasilewski .....

Date: 16/10/2023.....

Supervisor Declaration

I confirm that I have discussed the ethics awareness form with my supervisee and, if appropriate, requested that they follow the relevant guidance / make the relevant application.

Supervisor name: Peter Norris.....

Signature: PN.....

Date: 26/10/2023.....