**Online Adblocker**

One Semester Individual Project

Final Report

CM3203

Author: **Oliver James Robert Storey-Young**

Student Number: **C1635943**

Supervisor: **Padraig Corcoran**

Moderator: **Federico Liberatore**

**Cardiff University Computer Science 2020**

# Abstract

When browsing the internet, it is all too common to be bombarded by adverts that will hinder and alter their browsing experience. Over the past few years, the number of adverts and unwanted traffic loaded when you visit a website has increased to a ridiculous amount. Adverts can often be unsafe, collect personal information and have even been used to influence political votes.

In this project I aim to create a reliable and effective adblocker that will be able to detect and circumvent adverts that may be displayed on any given website. Since there are already adblockers available I will spend time researching the techniques and algorithms that have previously been used to help identify adverts on a web page and begin to build my own method of detecting them.

The following report will outline and evaluate thoroughly the process I have gone through to complete this project.

# Acknowledgements

I would like to give a big thanks to all my friends and family who gave continued support and motivation throughout the duration of my project. I would also like to express my gratitude to Dr Padraig Corcoran for his continued guidance and help throughout the project.

Most of this project was completed during the lockdown brought on by the coronavirus outbreak of 2020. I would like to add further thanks to all the NHS staff, key and essential workers who have risked their lives to keep the country running during these times.

Contents

[Abstract 1](#_Toc40106362)

[Acknowledgements 2](#_Toc40106363)

[1 Introduction 5](#_Toc40106364)

[1.1 Project Overview 5](#_Toc40106365)

[1.2 Project Aims 6](#_Toc40106366)

[1.2.1 First Aim 6](#_Toc40106367)

[1.2.2 Second Aim 6](#_Toc40106368)

[1.2.3 Third Aim 6](#_Toc40106369)

[1.2.4 Fourth Aim 7](#_Toc40106370)

[1.3 Project Scope 7](#_Toc40106371)

[2 Background 8](#_Toc40106372)

[2.1 Existing solutions 8](#_Toc40106373)

[2.1.1 AdBlock — best ad blocker “Ex1” 8](#_Toc40106374)

[2.1.2 AdBlock Plus - free ad blocker “Ex2” 8](#_Toc40106375)

[2.1.3 uBlocker - Ad Block Tool for Chrome “Ex3” 9](#_Toc40106376)

[2.2 Evaluation of the current solutions 9](#_Toc40106377)

[2.2.1 The Balance of too many options and not enough options 9](#_Toc40106378)

[2.2.2 Default adblocking effectiveness 10](#_Toc40106379)

[2.3 Constraints and Limitations 10](#_Toc40106380)

[3 Approach 12](#_Toc40106381)

[3.1 Setup 12](#_Toc40106382)

[3.2 Research 12](#_Toc40106383)

[3.2.1 WebRequests 13](#_Toc40106384)

[3.2.2 Storage 13](#_Toc40106385)

[3.2.3 AdBlock Plus 13](#_Toc40106386)

[3.3 Development 13](#_Toc40106387)

[4 Specification and Design 14](#_Toc40106388)

[4.1 Popup 14](#_Toc40106389)

[4.1.1 Users Perspective 14](#_Toc40106390)

[4.1.2 System Behaviour 15](#_Toc40106391)

[4.1.3 Architecture 16](#_Toc40106392)

[4.1.4 Constraints 16](#_Toc40106393)

[4.2 Storing Data 16](#_Toc40106394)

[4.2.1 Users Perspective 16](#_Toc40106395)

[4.2.2 System Behaviour 17](#_Toc40106396)

[4.2.3 Architecture 17](#_Toc40106397)

[4.3 Selecting Elements 18](#_Toc40106398)

[4.3.1 Users Perspective 18](#_Toc40106399)

[4.3.2 System Behaviour 18](#_Toc40106400)

[4.3.3 Architecture 18](#_Toc40106401)

[4.3.4 Constraints 19](#_Toc40106402)

[4.4 Blocking Elements 19](#_Toc40106403)

[4.5 Selecting Domains 19](#_Toc40106404)

[4.6 Loading Domain Filters 19](#_Toc40106405)

[5 Implementation 20](#_Toc40106406)

[6 Results and Evaluation 21](#_Toc40106407)

[7 Future Work 22](#_Toc40106408)

[8 Conclusions 23](#_Toc40106409)

[9 Reflection on Learning 24](#_Toc40106410)

[10 References 25](#_Toc40106411)

[11 Figures 26](#_Toc40106412)

# Introduction

## Project Overview

Websites have become littered with adverts. Adverts can come in many shapes and sizes, but all have one thing in common, to take something from you. As internet traffic has boomed over the past several years so have adverts. Adverts have evolved from simple pictures to embedded frames that can track you, collect your information and can even have malicious intent. Some websites are so heavily plastered with adverts that load times can be reduced and website functionality drastically lowered.

The aim of this project is to allow the user to easily manipulate their own browsing experience by culling adverts they do not want to or should not be seeing. Although ad-blockers already exist online they are often hard to customize for everyday users. I will be creating a chrome extension that can be installed into the chrome browser and then used to block adverts. Users will be able to manually select adverts to be blocked or opt to use predefined ad blocking lists that can be provided or added later to the blocker.

The application will offer the user simple yet powerful tools that allow for the manipulation of the websites they are visiting. Adverts that the user wishes to be blocked will be synced to their google account (this is needed to get the extension) and therefore will persist through any of their signed-in instances of google chrome. The application will offer two main streams of blocking that are element blocking, and domain blocking. Simply put, element blocking will allow for specific items on a web page to be hidden, while domain blocking will allow for adverts to be blocked before they are even loaded. Combining these two approaches will allow for a flexible and effective blocking extension.

The aim of this project is to develop a chrome extension that can be used to block adverts on websites inside of the chrome browser. The finished product will allow the user to manually select elements on the page to be blocked as well as offer a variety of address filtering methods for the user to block advert requests. The extension should be simple to use and easy to understand for the everyday web user. Although similar extensions already exist, the aim is to create a more user friendly and accessible version that will allow users to customise their filtering more simply.

My personal aims while completing this project are to increase my knowledge of JavaScript and to learn how to create and use the chrome extension API. This project will allow me to greatly expand my knowledge in both through use of documentation, trial and error and questioning. By the end I should have a clear understanding of the components that make up a chrome extension and the different functions and API endpoints I can use from chrome.

## Project Aims

The project being undertaken consists of a handful of complex parts that will allow for the overall ad-blocking experience. Therefore, to effectively plan and understand the road ahead I split my project into concise segments. Each segment holds a key role in the overall aim of this project therefore having detailed and well thought out acceptance criteria will help make sure they are met properly. Since I was working with the development process it was easy to make slight changes to my aims as needed whenever I encountered a major problem or discovered a new path somewhere else.

|  |
| --- |
| First Aim |
| The extension allows for the manual selection, storage and blocking of elements from any website the user wishes to block elements. |
| **Acceptance Criteria**   * It is obvious to the user which element they are selecting. * A reference to any item the user selects is stored (across sessions). * When a website loads, stored references are checked and applied to the loaded website to hide elements. * The user can edit the list of blocked elements they have selected. |

|  |
| --- |
| Second Aim |
| Adverts can be blocked via domain filtering selections made by the user. |
| **Acceptance Criteria**   * The user can select domains on an already loaded page to be blocked upon reload. * The user can manually type in domains which they wish to be blocked. * The user can edit the list of blocked elements they have selected. |

|  |
| --- |
| Third Aim |
| Adverts can be blocked via domain filtering lists that the user can load or present to the extension themselves. |
| **Acceptance Criteria**   * The user will be able to specify some sort of path to an already created filter list that they wish to use in their blocker. * The user can choose whether to enable this feature at any time. * The provided list works without conflict towards any other filtering. |

|  |
| --- |
| Fourth Aim |
| The user can easily and without confusion edit the settings of their application. |
| **Acceptance Criteria**   * All settings are loaded into the extension’s popup window, not in additional html files or background pages. * The settings have a good UX and will be understandable to users. * All settings are saved seamlessly upon change |

## Project Scope

The original scope for the project involved creating an extension for chrome, porting it over to other browsers like IE and even looking at blocking domains through the routers firewall. However, as the project took foot, I quickly discovered how hard it was going to be initially with learning another language and API. As the project has drawn to a close the scope has indeed shrunk greatly. The scope of the project ended up just being blocking adverts inside of the chrome browser. Initially I was too ambitious with what I wanted to do and reducing the scope meant; less pressure, better code, more time learning less languages and overall being able to think about how I could improve this smaller focus of mine. Overall, the ideas behind the adblocker have stayed the same, with the ideas for blocking elements being implemented, the domain filtering that I had previously thought about being done through the firewall or router being implemented straight into chrome.

# Background

## Existing solutions

When searching the chrome extension store its easy to see that there is not a shortage of adblockers out there for you to download for free. But what makes them different from each other, why is one better than another one. This short evaluation of existing solutions aims to look at some of the features and ideas behind a few of the adblockers. Gathering and evaluation these will help me make better choices when creating my own adblocker.

I will be looking at some of the options available with each of the blockers as well as carrying out some simple tests to see how well they block adverts across a couple of websites. Most of the adblockers seem to use the same underlying technologies so it will be interesting to see how they differ in action

|  |
| --- |
| AdBlock — best ad blocker “Ex1” <http://getadblock.com/> | 10,000,000+ Users |
| **Features:** Offers a free or premium ad-blocking experience, allow non-intrusive ads, show statistics of adds blocked, acceptable ads, custom filters, anti-circumvention measures, cryptocurrency mining protection. Open source and has a good support suite. + many more options |
| **Pros, cons, and thoughts:** Firstly, there are so many options for this adblocker, this is both good and bad. It boasts great customisation but can be a little overwhelming if you are not totally sure what you are doing. The popup is very simple to use and understand, also quite compact, however does not really have many options on it. The installation process on this one also takes a little longer than the others and does prompt you to donate once it has installed. Following on from the previous point the upgrade button is always displayed on the popup which can be a little annoying. With this adblocker when tested on *reddit.com* some ads were displayed, disappeared, and reappeared.  **Blocked:**  **R= {Claimed: 5, Visibility: Poor} S = {Claimed: 13, Visibility: Poor}** |

|  |
| --- |
| AdBlock Plus - free ad blocker “Ex2” [adblockplus.org](http://adblockplus.org/) | 10,000,000+ Users |
| **Features:** Whitelist website, whitelist page, basic statistics, different languages, tracking blocking, custom filters. Extensive help suite with lots of documentation. |
| **Pros, cons, and thoughts:** This adblocker have a good amount of options, not too many nor to few. The popup is very simple and therefore all the settings are hidden away. The UI for the adblocker is very clean and easy to understand. The adblocker installed fast and is totally free. When loading *reddit.com* no ads we are displayed to me. Overall, this adblocker is very simple to use and has a good amount of options that are easy to understand.  **Blocked:**  **R = {Claimed: 3, Visibility: Poor} S = {Claimed: 12, Visibility: Poor}** |

|  |
| --- |
| uBlocker - Ad Block Tool for Chrome “Ex3” 400,000+ Users |
| **Features:** Whitelist website, pause adblocker, statistics, allow non-intrusive add items through right click. |
| **Pros, cons, and thoughts:** This adblocker has a similar popup look to *AdBlock — best ad blocker.* It is very clean and shows you exactly what you need to see. Overall, the options available for this adblocker are very poor. However, I believe it provides a simple adblocking tool for the user who does not want any customisation etc. The extension has a much smaller user base. Once again, the options panel is hidden away in another page and is not very eye catching but cuts straight to the point.  **Blocked:**  **R = {Claimed: 4, Visibility: Excellent} S = {Claimed: 8, Visibility: Good}** |

## Evaluation of the current solutions

### The Balance of too many options and not enough options

Options are important when choosing the adblocker for you, the user needs to be in control of the content that is being allowed and blocked from the websites they visit. Finding a balance between needed and extra options is also something to think about as you do not want users to be overwhelmed or confused by the options that are available to them.

Firstly, *Ex1* had a huge range of options available which meant that the user was going to be able to fully customize their browsing and adblocking experience. However, I felt very overwhelmed when looking at the options as I was not sure what a lot of them meant or what they would do for me. Also having the premium and the options available with it constantly pushed in my face was annoying. Overall, I believe this adblocker had to many options available with it.

Secondly, *Ex2* had a good amount of options, it felt like the options that were available had been carefully curated to keep the user informed and not overwhelmed. Just like the previous the options were tucked away in the settings page. Overall, I believe that this adblocker had a brilliant balance between options and simplicity.

Finally, *Ex3* lacked a lot of options but seemed very simple and intuitive to use. This adblocker was missing a lot of features that I felt were needed such as being able to load in a custom filter list. This adblocker could really do with a few more options but if you are looking for something very simple and lightweight this may be a good option.

Overall, I believe that *Ex2* (AdBlock Plus) had the best balance between customization and simplicity with its careful balance of options available. I would have liked to see more options available on the popup though as all three had theirs hidden away in the settings page.

### Default adblocking effectiveness

An important factor to think about with these adblockers is how well they perform once they have just been installed, so no options being changed, the program in its default state.

I carried out a small test across two websites:

*R =* <https://www.reddit.com/>

S = <https://smallbusiness.chron.com/affiliate-marketing-tracking-tools-67773.html>

I recorded how many adverts are supposedly being blocked by the adblocker and then my own view on how well it had done to remove the adverts from the page. To do this I viewed both pages with and without the adblockers to see how well they did.

Firstly, *Ex1* Boasted the highest numbers being blocked but failed to block the ads which we are visible to me on the page. This happened to both websites in which I tested. I assume this means that adverts / tracking data in the background is being blocked but it has not recognised or filtered out the adverts that I am seeing.

Similarly, with the second, *Ex2*, both websites reported having a high number of ads blocked on the page, but all adverts were let though as usual. This is unfortunate for both bigger names adblockers to be letting through so many adverts I rated both websites as poor just like *Ex1*.

Finally, the smaller adblocker *Ex3* pulled through and did a really good job at blocking the adverts. This adblocker reported much lower number of adverts being blocked but then went ahead and blocked all but one of the visible adverts that I could see.

To conclude my adblocking tests I have discovered that the smaller of the blockers with much less options available had come out and blocked many more adds than it is bigger more popular competitors. There are many reasons this could be: Perhaps the other two need some setting up to work more effectively, Block may have a newer filter list.

## Constraints and Limitations

One major constraint I will face when undertaking this task is that I am not well informed with the technologies that I will be using to create my adblocker. To begin with I will be using JavaScript, this is a language I have used previously at university but have a very limited knowledge of. As this language is used often by many people I will preserver with the guidance and documentation that can be found online. Following this previous statement, I will also be learning and dabbling with jQuery, a shorthand for JavaScript that should help make coding simpler. Similarly, I will be working in a new environment to me – Chrome. I have never developed a chrome extension or worked with the API, so I do not know what is unattainable and what is realistic. Again, for all these documentation and further reading will help lessen this constraint.

One constraint that I have encountered that has affected this project is the coronavirus. At the time of writing this report university has been closed for quite some time. Universities have closed, society has shut down. The resources that are available when normally undertaking a university degree are not available. Computer labs, Libraries, and the ability to meet up with supervisors, peers and tutors have all been removed, potentially threatening the outcome of this project. As this project has progressed it has been noticeably difficult, being out of routine and stuck in a weird situation where no one is safe has impacted the ability to work properly will have had slight affects on this and many other projects.

The final constraint to this project is to do with filter lists that are used by adblockers. After doing research it is apparent that adblockers mainly use other people’s lists as filters rather than their own. Some of these lists are public and can be used but are written in schemas and formats that are not fully explained and may therefore be unusable. It is the aim of this project to at least allow for those filter lists to be loaded and used at least, as if they we’re processing the filters correctly.

# Approach

The approach that was taken towards the project was Setup, Research and Develop. This approach allowed for all the core pieces of application to be put into place before cycling back and forth through research and development. As each week came by and work began on the aims of the project, research was made into each individual part and used to develop that piece. This approach worked well as it worked well with the flexibility and granularity of the agile methodology. As the project move forward less and less research was needed, as familiarity grew, which left more time for development.

## Setup

The first part to the approach was the setup. During this time, the project had all its core components created and setup which would allow for everything else to flow nicely. To begin with was the task of setting up version management. GIT was chosen as it is easy to use and widely accessible. It was chosen that it would be hosted on a private repository on GitHub. Having the project hosted here meant it could be worked at from any computer on any device and the files would never be lost. The client that was chosen to mediate this was Gi Tortoise as it provides an easy and simple to use interface. GIT played a big role in this project and allowed for rapid development through trial and error.

Next on the setup phase was to create the initial chrome extension. I first had to learn what files would be needed and how they interact with each other. Over the course of a day the barebones application was created and an understanding of what each key individual file did was gained. Over the course of this project the skeleton evolved slightly with new discoveries and ideas coming into play. This played an important part as gaining a better understanding of how each component worked early allowed for progression with more pressing items.

The final Items that was part of the setup was a task board and notes page. Trello was used as an easy way to divide the project into tasks and subtasks that would allow for progress tracking. Trello is an online task board that works well with the agile methodology. Being online, like GIT it is useful because it could be accessed from any device and keep development up to date at any point. To keep notes a work in progress word document was created where anything from failed ideas to thoughts could be placed to help aid the writing of this report.

## Research

Using a language and API that was not so familiar meant a lot of research was needed continually throughout the development. Research was carried out and then put into practice in the development stage before that small cycle started again with the next component. There were three main areas of research that were carried out and are listed below.

### WebRequests

WebRequests is a crucial part of the chrome API that would be needed in the development of the extension. WebRequests allows for user intervention at several locations during the life cycle of an incoming or outgoing request. Using this would mean requests could be filtered and blocked to stop adverts from being loaded onto the webpage from external sources. On the API documentation page there are detailed and descriptive examples of how to use the code in a variety of scenarios. This and the overall chrome API documentation was a big help and made development a lot simpler.

### Storage

The Chrome API has an easy to use storage facility located in Storage.Sync. This storage facility allows for data to be stored inside of the browser, attached to your google account. Using this meant that users settings, choices and block preferences could all be saved easily and loaded to anywhere they log in. The storage API saved lots of time and effort as originally the plan was to save these things to documents. Storage.Sync has two simple functions, Get and Set. Get returns an object and set allows you to set an object. Combined these two powerful snippets helped create basic foundations of how the extension would solely function.

### AdBlock Plus

AdBlock plus is one of the already popular adblockers that is available as a chrome extension. They host lots of material on their website that talks about how their extension works, some of the algorithms and techniques behind the scenes. All their code is open source too. Combining these elements meant there was a good starting point and reference location to help throughout. Although greatly useful some documentation and examples were missing which meant I was not always relying on this and constantly came up with my own solutions.

## Development

The final part of the chosen approach was the actual development. Developing was straight forward and coupled with GIT allowed me to be flexible with my attempts as I had branches, I could always fall back on. Development would differ per task but would normally consist of a front end to back end style approach where possible. Since this extension is focussing on having concise and available options to the user, there was lots of front-end options that needed hooking up. The aim was to create one front end object first, connect it to storage and then hook-up the background logic after. Creating each option like this meant that it would be simple to trouble shoot and each commit was related to just a single item. Finally, as each bigger stage would be completed an overhaul including refactoring, reordering, and minifying could be applied to make the code better. Throughout the process there would be many problems and challenges but thanks to the approach I had selected and tools I had used to help preserve the projects integrity I would be able to work stress free and effectively code the project. Overall, the approach was a good fit for the extension I was creating and worked nicely with the agile framework.

# Specification and Design

The development process of this project is broken down into these key stages

* A popup window that allows the user to interact with the extension.
* A storage system that will allow for user’s settings to be stored and accessed across all sessions.
* A script that will allow the user to interact with their current website to select items that need to be blocked
* A script that will allow previously selected blocked items to be removed from the website
* A script that will allow the user to select domains that need to be blocked
* An importing script that will allow the user to load in a filter that can be used to block domains

## Popup

### Users Perspective

At the core of most chrome extensions is the popup. The popup can be accessed by clicking on the extension’s icon found just after the search bar inside of chrome. The popup is primarily used to allow the user to interact with the back end of the extension. Often you can find settings and statistics inside this panel. Often the popup will link to a background page that allows the user to get further customisation.

With lots of extensions having their settings muddled between both the popup and background page and from my research, it was decided that all the options would be kept inside of the popup to make life simpler. This meant creating and mapping out what would be on this popup was one of the first steps. Once each of the options had been added, each of the scripts and code snippets that would run from and because of those options would be coded.

The popup would be split into three smaller sub sections that would group options and settings together. The groups are split with lines so that the distinction between them can be made. The group of options nearest the top of the popup that the user can see contains the options; *Run on this page* and *Collapse / Hide Items.* The option to run on the page is located right at the top which means upon opening the popup can quickly be used to toggle the adblocker on and off. It uses a checkbox and is therefore easy for the user to understand. The next item is a set of radio buttons that allows the user to select whether blocked items are being collapsed or hidden. This small but concise group allows the user to quickly modify how items are being displayed on the website.

The following group contains most of the actual blocking interface. These options are primarily made up of buttons and allow the user to store items and domains that they wish to block. Starting at the top of this group is a button that allows the user to start selecting items on the web page to be blocked. When this button is clicked the user will see a visual change on their website that allows them to select items. The next two options both buttons and allow for the user to reset in bulk items that they have previously selected for either the website they are on, or all websites they have data stored for. The next button sits next to a textbox and allows for the user to input a domain that they would like to be blocked. The user must first type in the textbox a domain string, before hitting the button to store it and refresh the page. The input textbox has a placeholder string that denotes the format of the string that needs to be entered. The penultimate option in this section allows for the user to load in an existing filter list via a URL. The option consists of two buttons and a textbox. The textbox can be used to enter the URL, the button on the right labelled Save, stores that URL in the storage and finally the button on the left enables and disabled this option. The save button and input textbox are both toggled on and off via the button to the left that enables and disables this option. The final option in the group is the default filter checkbox. This checkbox can simply be used to enable or disable the default domain filter list that is shipped with the extension.

The final group on the popup contains two dynamically created tables. Both tables look the same and have a header titling each one. The first of the two lists in rows all the elements that have manually been blocked on this page. The user can then click those elements to unblock them and return them to the website. The second table has the same functionality as the first but instead lists all the domains that have manually been blocked. Both tables display a count to the user, so they do not need to manually count how many items are in the lists. Each of the tables has a maximum height too so the tables will be scrolled more so than the popup itself.

Overall, I believe the design is simple and easy to follow for all users. The grouping has allowed for similar options to be put near each other in an obvious way. Coloured and bold text has also been used to emphasis certain parts of the UI and is even used as feedback in some places. All the options fit onto the popup as a scrollbar is used to catch the slight overflow and allow the user to scroll once some items have been added to the tables.

### System Behaviour

The popup works primarily through JavaScript as its back end and uses a mixture of DOM events and the chrome API to act. To begin with the HTML document is loaded with all the options set in their default states. JavaScript DOM events are used to listen for the page being fully loaded and then calls made to the chrome API retrieve all the stored settings options. It is at this point that all the options are populated correctly – meaning the user will see what they have previously saved.

Similarly when the user interacts with the options on the HTML page their actions are listened to by JavaScript, which is monitoring for certain clicks on items, or changed events on checkboxes and will send data to the chrome API to store these changes. When options are changed, the page the user is currently on will be reloaded so that these modifications can take effect. Some options will keep the popup open and therefor will actively modify the contents of the popup when they are changed to ensure the user gets feedback from the interaction. An example of this is when a domain is removed, the popup stays open and the domain will be removed programmatically from the list, without the popup refreshing and repopulating

Another way in which the popup behaves is through messaging. Messages can be sent through chrome to other script. These messages act as a way for the popup.js to interact with the background.js or the JavaScript of the page that is currently loaded. This can be seen through the reload function that will send a message from popup.js to the currently active page and tell it to refresh. Overall, the popup works by getting and setting stored data into chrome storage and notifying other pages when something needs to happen.

### Architecture

The architecture behind the extension is not to complicated and uses the traditional web languages as well as a couple of additional packages. HTML and CSS were both used in combination to create the basic outline for the application. Once this skeleton had been created JavaScript was added in its own file as well as exposing the interface to jQuery to make JavaScript coding more efficient.

JavaScript was heavily used in the popup as a means of loading in data from the storage, which will be talked about later. JavaScript was also used as a method of adding feedback to the user about their actions. An example being when the user enters an incorrect or accepted string into the domain blocking textbox.

Bootstrap was later added to add a cleaner style to the popup. Bootstrap is a premade CSS file that can be added to any project to style elements. The card class from bootstrap was used to house the tables located in the third group of options. Combining bootstrap with our own custom style sheet allowed for quick and simple customisation to the way the popup looked.

### Constraints

As this project has a short deadline and the bulk of the work will be focussing on actual logic and code the UI is not quite as important of a task and therefore will be taking the back foot. This means that the options available may not be as intuitive or easy to understand as I would have liked but will still have the full functionality that I am intending. UI and UX design are primarily driven through user feedback and immense testing and this is not a priority.

## Storing Data

### Users Perspective

Storing data will be straightforward to the user. They will not have to worry about what data is currently stored behind the scenes. This data will be viewed by the user regularly but in its finalised format instead of raw items. This means the user will see settings inside of the settings window, checkboxes ticked etc, rather than interfacing directly with the data itself (seeing a stored true or false etc.) The user may store data only through the provided outlets, again, checkboxes, textboxes, and gestures such as clicks. This means the user is not able to compromise the data by accident or mess with the settings in a way which was not intended. Data about stored domains and stored elements can be seen through the popup. Buttons are also available in the popup to help modify what is stored.

### System Behaviour

The system will work by pushing and pulling data using the chrome storage API. Data in chrome can be stored using Storage.Sync which allows data to be synced to the google account it came from, assuming extension syncing is enabled. The first method is the get method. When using the get method we can provide a string that we would like to look for. This means if the user clicks a button that needs one item, we do not have to get everything and can just retrieve the value they are looking for. All of this can be done using a call back function that will run once the request of data has been retrieved. This process is simple and easy to understand. If retrieving all data from the storage a JSON object will be retrieved, however if it is a specific item it will come in its stored format.

The other way in which the storage API will be used is to store data. Data can be stored by calling the set function and providing it with an object to store. An optional call back can be added to allow extra code execution to be done once the data has been fully set. It is important to note that most of this API is asynchronous and will therefor continue executing other code while it waits for its results. When setting data, we cannot just modify an item that is already stored. To modify an item that we have previously stored we will have to call the get function with a call back. Once this call back happens, we can call the set method and pass in a modified version of the data we received from the get function. This can be very annoying but is not very taxing overall.

### Architecture

The architecture behind the storage of data is mainly how we format of the data we are storing. Data is stored inside of chrome using JSON objects. JSON (JavaScript Object Notation) works seamlessly with JavaScript and will be easy to index and extract the items I need. The basic Idea is to store three main objects inside of the initial JSON object. These objects will be the blacklist, whitelist, and settings objects. Dividing the initial object up into these three sub objects should allow better data retrieval and a simple to use structure. It is important to note in JSON anything can be stored and nested, so objects in objects is fine.

The blacklist will be split up into two sections. The first section of the blacklist will denote which domains should not be allowed access from the browser and should be an array. The second item will be the elements that need to be blocked on a web page, another array. Storing them like this will mean it is simple to gather all the items which are “blacklisted” or bad and we do not want.

The second part is the whitelist. The whitelist will be used to store which websites and elements should be allowed to run on certain websites. This is important as we need to allow the user to select elements to be blocked but may need to be allowed on another website. It will be connected to the option that asks the user if they want the adblocker to run on this page.

The final part of the storage is the settings object. Quite simply this will be used to store the different settings data. These will primarily be true and false options that allow the user to switch things on and off. Such as using the default filter list or using a web filter list. Other options may use strings for storage or any form they see fit.

### Constraints

There are a few constraints when it comes to data storage. Firstly, is the amount of data I can store. The chrome API boasts a few limitations as to how many items can be stored and how often. These limitations will shape the way in which filters are used by my application and may result in using external API’s as a main form of input rather than saving items directly. Another constraint that arises from what was mentioned in the system behaviour is call backs. Call backs are slightly outdated in JavaScript and when coupled with newer technologies like promises can get complex and confusing. Call back hell exists and means watching how and where call backs are being used will be important to maintain the quality and readability of my code.

## Selecting Elements

### Users Perspective

Selecting elements is a key part of the adblocking experience for the user. Selecting elements allows for the user to manually click an item on the webpage they are viewing to render it blocked. Creating and intuitive and simple interface for this was tricky at first but has turned out well.

For the user to be able to select and element to be blocked they must first open the popup. Once the popup is opened, they must select *Select an element to block* from the second group of options. Once this option has been selected the user will now see their current web page modified as they move their mouse around. As the user moves their cursor over the web page a red dashed border will appear around their currently hovered element. This element selection only allows for items that have a class or ID attached to them. As the user moves their cursor around the border is programmatically removed and applied to whichever element they are not hovering over.

When the user is ready to block an element, they may click to add the classes and, or ids to the list of elements to be blocked on that website. During this time, the user may also want to block an element that is a link or has functionality attached to it. This is not a problem as when the user is in this selection mode links are blocked and therefore will not redirect the user. To further this the cursor will be the normal pointer rather than the hand which indicates a link. Once the selection has been made and the web page will reload, and the selection mode will be taken off revealing the normal website once again. This script is only designed to allow for the selection and storing of elements and blocking will be done by another component.

### System Behaviour

When a web page is loaded the script Select Element is automatically loaded in as stated in the extensions manifest. This script is injected before any of the DOM has begun to load. This script sits idle with a listener waiting for a message telling it to begin. This message is sent only by clicking the *Select an element to block* in the popup window. Once this message has been received the script will be activated.

The script will listen to movements of the user’s mouse and identify which element they are currently hovering over. Whenever the mouse moves this function will be recalled and the domain object recalculated. Every time a new object is found it will be highlighted in an obvious manner for the user so they can see which element the script has calculated they are hovering over. The user is now stuck in this mode until the web page is refreshed which can be achieved by manually refreshing the page or continuing and selecting an element which will automatically trigger the refresh.

### Architecture

The element selection consists of JavaScript and jQuery spread across two files. Once is the *SelectElement.js* file which houses all the code that allows the user to select items. The other file is the *popup.js* file which tells the current page when to go into selection mode. *SelectElement.js* also interacts with the chrome API to store items. CSS is also dynamically injected into the webpages to stop links from being clicked when in selection mode.

### Constraints

One constraint for the current design means that elements without a class or id will not have the ability to be blocked. Currently the plan is to store a reference to the class and or id that the selected element has. This involved going up the DOM until the hovered item has either one. Using this strategy, it means selections bigger than the user’s initial intent may have to be blocked.

## Blocking Elements

### User Perspective

The user perspective for blocking elements should be simple. The process should be done seamlessly upon items they have previously selected. This system should shy away from any extra user interaction and instead this should be handled by the previous, element selection. There should be two related options available to the user. The first is *Run on this page* Which will allow the user to toggle the adblocker on and off for the current web page they are viewing. And the second denotes how elements are hidden on the page. This second option will allow the user to chose whether the space of the adverts is to be left as it was but blank, or for it to be collapsed. These options keep the process simple but offer two powerful choices for the user to make.

### System Behaviour

As the script runs on every page it will always have the most up to date version of the stored elements that need to be blocked. This script will begin to run before the web page has even began loading properly. During this time, it will prepare all the CSS that needs to be injected. This should be done by creating one CSS rule that is applied to all the stored elements. Once This is ready and the page has begun loading it should be added straight to the page allowing it to block all the elements. As this should be done right at the start it should mean the adverts do not get seen at all. Everything to do with this script is automatic and does not need any interaction from the user. A condition will wrap this code firstly telling it whether to run, this will be based of whether the website is on the whitelist or not. Secondly another code snippet will check the settings to either place the visibility or display property on it correctly. An important tag can also be applied add extra presidency onto the generated CSS statement.

### Architecture

The architecture to this script is very simple. As previously mentioned, it loads at the start of the page. It will generate its own custom CSS rule that includes all the items it loaded from the storage. The script will contain a listener that waits for the document to begin loading properly, and a function that will be ran when this happens. This script is the easier part of blocking items and does exactly as you would expect.

### Constraints

A constraint for this system will be to do with how elements were previously stored. As mentioned in the storing of elements, the class and id will be stored of the selected element, or first parent with one of these attributes. This could mean that only a child of the advert is being blocked and that the parent frame will have its own height and width. Although this parent may be blank it will not be collapsible in line with the setting to do so. Another constraint related to other elements with the same name or class. With the current design the blocker will do as it is told and block the ids and classes its been given. However, this does not mean some elements on the page will not have the same class or id and get blocked by accident.

## Selecting Domains

### User Perspective

Selecting domains will allow the user to manually block entire websites they do not like. The user will be able to either type or select domains that they wish to stop traffic from loading from. There will be three ways in which the user is able to do this. Selecting, typing, and loading. Each of these three ways allows for a different interaction between the user and extension but ultimately end with the same goal of blocking domains. Having multiple ways is important as it will allow for a greater convenience top the user.

The first of the three ways that the user will be able to interface with domain selecting will be through a manual selection. The user should be able to identify an advert themselves and add it to the block list much like they can do with elements. With domains it is much easier for the user to see what they are selecting as chrome displays the currently hovered URL against the bottom of the window. The user should then have the option to select the item and mark it as blocked, adding it to the storage.

The second way in which the user will be able to interface with domain selecting will be through typing it out and having it added to storage. Inside of the popup there is a textbox and button combo that allows the user to type in a domain they do not wish to receive traffic from and block it. Domains to be filtered require a specific format with the chrome API and therefor a placeholder and entry feedback system will be in place to help the user type an acceptable string.

The third and final way the user will be able to select domains is through an external source. An option will be available to the user allowing them to save and load in a list of filters from a website. This option again will be displayed to the user inside the popup menu. The user will only need to supply a URL for this to work. The plan is that the user can use lists already created and used by other adblockers. Once the user hits save and ensures the toggle is enabled the loaded domains will be blocked.

### System Behaviour

After any changes are made to the domain’s lists, either than be a new entry is added or loaded the page will refresh so that the domain blockers may take affect on the updated list. Upon the user selecting a domain to be blocked manually the script will automatically convert the URL that was collected into the right format for the chrome API and add it into storage. When the user selects to add their own URL to the block list and types it in the format will be verified and then added once again to the block list. When loading from a web page, the user will first have to enable the option before they enter the URL and hit save. Behind the scenes seamlessly the data will be collected and stored inside of the local session, to be used until the browser is closed. When the browser is reopened the data will be redownloaded from the saved site.

### Architecture

The architecture behind domain selecting is varied per each method previously stated. With the method of manually selecting elements the chrome API can be very handy. Using the chrome API will allow for a custom context menu to be created and used. Using a custom context menu will allow for an extra option to be added to the right click menu. Clicking this item can run code which will allow the domain the click came from to be blocked. This is a simple process and is majorly handled by the chrome API.

When it comes to typing out domains, we are instead restricted by the chrome API. It only accepts domain strings in a specific matching format that it uses. Therefore, the user will have to type it correctly and Regex can be used to ensure this. With little space it is important to make it as obvious as possible what format the domains need to be in for the user.

The final method will use XMLHttpRequest that is shipped with JavaScript. This will allow us send and receive data from other web addressed. The user should enter a URL that loads a list of domains they would like to filter against. XMLHttpRequest can then load the URL and retrieve the contents of the page. Once the data has been retrieved it can be formatted correctly and then applied as a filter to the incoming requests.

### Constraints

A major constraint lies with this part of the project and that is using external filter lists. Many lists of predefined domains already exist but can be hard to implement. To begin with as lists grow the time it takes for them to be filtered and compared will increase making the whole adblocker and browsing experience slower. Another constraint that goes with these lists is implementing them. They all come in their own format with schemes etc and it may not be possible to parse the lists correctly if the documentation of how to do so does not already exist. A final constraint on selection could be making sure the user is selecting the right one.

## Blocking Domains

### User Perspective

The user is not going to see much when it comes to domain blocking, only really that the domains they have selected will not appear. There will be one related option inside of the popup that will allow the user to make changes to blocking. The User will be able to see a list of the domains they have manually chosen to block. This means domains they have typed in or entered through the context menu. The table these are in will allow the user to release domains they no longer wish to be blocked.

### System Behaviour

When chrome loads so does the background script. This is where the domain filter API will be attached to incoming requests. These listeners will take in a list of URLs to check against and block ones that match. These listeners take in a copy of the list so any changes will mean the listeners need to be refreshed. As the user makes changes to the settings, loads more URLS or removes some the background will update its listeners with updated versions of the lists. The blocking can be done in multiple ways. The blocking can either filter out entire domains or filter out domains that match certain character strings etc. This is something I will consider when I begin actually implementing.

### Architecture

The architecture behind this part of the extension is primarily going to be focussed on the Chrome API. The Chrome API has a whole toolkit that allows you to work with incoming and outgoing web requests. I will be using some of these tools to allow me to filter out which domains are being let in. the blocking will be done in the background script which is persistent throughout the duration of your session on chrome.

### Constraints

Blocking domains can come with a few constraints. To begin with there is an issue where the popup itself could become filtered, therefore rendering the whole extension useless until re-installed. As blocking domains can use pattern matching small parts of the URL could match and therefore make it filtered. Another constraint is just the overall war between adblockers and advert makers. Blocking domains on a basic level is not so tough but advert makers are making it harder to do so. There are many ways in which adverts are bypassing domain blocks and it will be hard to work against and combat them all. There is a lot of unknown and only working on this project will help reveal the effectiveness.

# Implementation

# Results and Evaluation

# Future Work

As the project has progressed code has become messy, comments lost, and bits of code duplicated. Some features did not make it in because of time restrictions and difficulty to implement them so here I will look at some of the things that were missed out on and would benefit being added. The program currently works well and does not have any really pressing issues. As l have been learning JavaScript and the Chrome API as the project has progressed earlier bits of code still lie around and could easily be improved.

## Full validation

Validation is something that was only touched on slightly during the project. Currently there is only validation when a user enters a domain string to be added to the domain block list. This validation works well and gives decent feedback to the user in the form of a green or red outline to explain the result of their entry. There is also some validation behind the scenes to stop duplicate items being added to the blocklists but not under all circumstances. Validation like this would be useful in all scenarios and help the user understand how their actions are impacting the extension. Luckily, this application does not have to many points of entry for the user so there are not tonnes of things that need validating.

### Domain Loading

The option to load domains straight from a URL is very important and useful in the extension. However, this feature can be quite dangerous and easily break the application if not used correctly. It would be a good start and important point to add perhaps a limit to which sites the URLS can be from. Perhaps limit it to only pastebin files that are raw (so it is just plain text) or as some other adblockers do, provide a drop down of premade lists. If the option to allow any sort of URL was kept then ensuring that all the contents of the file are acceptable would be a good form of validation. This would stop errors happening during parsing.

### Manual Domain Typing

Currently the user must type in a domain following the schema that is displayed in the placeholder text. This is then validated once entered. It would be better if instead the user could enter any domain and that be validated and then converted into the correct format. Although the format is obvious as it is shown, it would be easier and be less error prone if the user just needed to type a normal domain. The current validation could also be updated to reflect the acceptability of the string on each keypress.

## Error handling

Currently the extension does not have any error handling in it. I have not come into any ground-breaking or application stopping bugs yet. As we are primarily storing items and retrieving them any data that breaks the application while running could probably be solved by a browser restart.

### Error logging

Error logging could add to the application to allow the user to see what errors they are having. These errors could be displayed in a custom debugger window and will allow for advanced users to fix problems themselves. Error logging could also be used to send important information about problems straight to the developer if the user has opted in for their data to be shared. Having error logs can benefit everyone. Currently the application does not record or alert the user of any problems and this can be confusing and could be frustrating if it breaks entirely.

### Try Catch

Currently most of the code runs as normal in the flow it has been created for. It is assumed that the data going in will be right but not every entry point or condition has been through out. Try catch allows for code to be executed without breaking a program. The try statement allows for code to be ran and if nothing is wrong it continues as normal. The catch can be used to alert an error or fix an issue if the try has failed. JavaScript does not rely on user typed data types and therefore works well to cast and convert datatypes where it can but can sometimes fail, this is a scenario where try and catch is useful. Another place could be when connecting to a supplied URL if it is inaccessible, to try again or alert the user.

## Code optimisations

As this project progressed old code and new code have collided and been made to work. Existing structures that were put into place have remained but could be done better. Some components could do with a complete rewrite and may be better of with one. Code however has been refactored where possible and therefore is efficient still. It is believed that further optimisation would help increase the effectiveness of this adblocker. As many parts of this are dealing with requests and run before pages are loaded every millisecond of improvement will help make the extension better.

### Core components

As stressed previously this project has been a giant learning curve and many things have been built upon the initial code. Currently many parts of the code are being duplicated and are jumbled up. Much like an object-oriented language I believe this project could benefit from having its core and reusable components stripped out and added to their own JavaScript files. JavaScript is very robust and allows for code to be split up very easily. Doing this would help make code more readable and the overall structure of files in the extension more concise. Having less duplicated code would also mean a smaller file size of the overall extension.

### Improved data fetching

Currently in the project data is being get and set all over the place. Most methods pull data whenever they are called. While this works well it means that code is being duplicated constantly and fetching data that is likely already being stored in the extension’s current memory. It would be much more efficient to pull the data in once and pass it to every script that needs to use it. Then updating it at any location and setting the original getter to update its values. Finding a way to implement data fetching with promises would allow for a better code structure too. Currently there have been race conditions and problems with call backs that have made other parts of the code structure less clear. Finding a better solution to these race conditions and call back problems would help the tidiness and efficiency of the code.

## Integrating with EasyList

EasyList is one of many free to use lists that contains bad domains and elements for websites. Currently the list is around 2MB and contains hundreds of thousands of entries. During my project I tried to implement this into the program but was not able to after having troubles understanding the full schema it uses. EasyList has its own format for URLs to be written in which allows it to filter domains precisely, with specific attributes, letter patterns, extensions etc. This list would allow domain blocking that is much more on par with the current solutions that implement this. Implementing with EasyList would however mean adding major optimisation onto how each filter is checked. With tens of thousands of potential matches, and this happening before every request goes out it would need to have no major impact to the performance of the web page.

# Conclusions

# Reflection on Learning

# References

# Figures