



# Create a browser extension that provide end-to-end support for online shopping

by Meirong Li, BSc

Submitted to The University of Nottingham

in September 2023

in partial fulfilment of the conditions for the award of the degree of

Master of Science in Computer Science

20341357

School of Computer Science

University of Nottingham

I declare that this dissertation is all my own work, except as indicated in the text

Signature: Meirong Li

Date: 07/09/2023

I hereby declare that I have all necessary rights and consents to publicly distribute this  
dissertation via the University of Nottingham's e-dissertation achieve.

## **Abstract**

In the dynamic landscape of e-commerce, individuals seek tools and solutions to enhance their purchasing encounters. This article examines the notion of creating a browser extension that offers comprehensive assistance for online buying, effectively tackling the varied requirements and difficulties encountered by contemporary e-commerce consumers.

The browser extension that has been developed seeks to offer a comprehensive approach to online shopping by integrating various functionalities, including the ability to compare product features, track prices, and receive personalized purchasing recommendations. By integrating these activities into a unified and intuitive extension, online consumers may streamline their procedures, optimize time management, and enhance their ability to make well-informed buying choices.

Furthermore, this study delves extensively into the technical and security aspects that are necessary for the successful implementation of such an expansion. The study also investigates the potential for compatibility concerns arising from cross-platform needs, with the aim of ensuring the extension's effectiveness and user satisfaction.

In the current epoch characterized by the widespread adoption of online shopping as an essential component of daily existence, the creation of a comprehensive browser extension emerges as a potentially advantageous resolution to furnish customers with the necessary resources for adeptly and effectively engaging in transactions across diverse e-commerce platforms. This study aims to elucidate the potential advantages of implementing this extension, with the objective of enhancing consumers' purchasing experience by offering greater convenience and efficiency.

## Content

<b>ABSTRACT .....</b>	<b>2</b>
<b>CHAPTER 1 .....</b>	<b>5</b>
<b>INTRODUCTION .....</b>	<b>5</b>
1.1 SENSEMAKING.....	5
1.2 AIMS AND OBJECTIVES.....	6
1.3 MOTIVATION .....	6
1.4 DESCRIPTION OF THE WORK .....	6
<b>CHAPTER 2 .....</b>	<b>8</b>
<b>RELATED WORK .....</b>	<b>8</b>
2.1 Data visualization .....	8
2.2 online shopping.....	9
2.3 Literature review .....	9
<b>CHAPTER 3 .....</b>	<b>11</b>
<b>USER DESCRIPTION.....</b>	<b>11</b>
3.1 Target user.....	11
3.2 User requirements.....	11
3.3 Function requirements .....	12
<b>CHAPTER 4 .....</b>	<b>13</b>
<b>METHODOLOGY .....</b>	<b>13</b>
4.1 Overview .....	13
4.2 Data Management.....	14
4.2.1 Data Collection .....	14
4.2.2 Data Storage .....	15
4.2.3 Data Understanding.....	15
4.3 Radar Chart.....	16
<b>CHAPTER 5 .....</b>	<b>17</b>
<b>DESIGN.....</b>	<b>17</b>
5.1 Chrome Browsers Extension.....	17
5.1.1 Early Preset.....	18
5.1.2 Post Implementation.....	20
5.2 Related Processes.....	22
5.2.1 Data Link .....	22
5.2.2 Data presentation .....	22
5.3 Data processing.....	23
5.3.1 Data collection .....	23
5.3.2 Data analysis .....	24

<b>CHAPTER 6 .....</b>	<b>25</b>
<b>IMPLEMENTATION .....</b>	<b>25</b>
6.1 DATA RANGE.....	25
6.2 CONCRETE IMPLEMENTATION.....	25
6.3 DATA STORAGE AND INTERACTION .....	25
6.4 MESSAGE QUEUES.....	26
6.5 CONFIGURATION OF THE USER INTERFACE .....	27
6.6 THE ESTABLISHMENT OF THE BACKEND SYSTEM.....	29
<b>CHAPTER 7 .....</b>	<b>31</b>
<b>EVALUATION.....</b>	<b>31</b>
7.1 USER EXPERIENCE .....	31
<b>CHAPTER 8 .....</b>	<b>32</b>
<b>SUMMARY.....</b>	<b>32</b>
8.1 PROJECT MANAGEMENT .....	32
8.1.1 TIME PLAN.....	32
8.2 CONTRIBUTION AND THINKING .....	33
8.2.1 INNOVATION.....	33
8.2.2 COMPUTER ETHICS AND PROFESSIONAL ISSUES .....	33
8.2.3 REFLECTION AND FUTURE CONSIDERATIONS.....	34
<b>REFERENCES.....</b>	<b>35</b>

# Chapter 1

## Introduction

The realm of e-commerce has had an unparalleled upsurge in prominence, becoming as an essential component of contemporary consumerism. The advent of internet shopping has significantly transformed the retail industry by providing consumers with the ability to browse an extensive range of products and make purchases conveniently from their own residences. Nevertheless, as the realm of online purchasing expands, buyers encounter increasingly intricate hurdles in motives via diverse e-commerce platforms, locating optimal bargains, and guaranteeing a secure and efficient shopping encounter.

In order to tackle these issues, the present effort endeavors to create a browser plugin. This plugin facilitates the provision of decision-making suggestions by utilizing the user's online search data. It offers customized purchasing recommendations encompassing pricing comparison and performance analysis. Consequently, it enhances the user's shopping experience and improves their shopping efficiency. This article aims to explore the process of developing a browser extension that offers comprehensive assistance for online shopping, catering to the many requirements of contemporary consumers.

### 1.1 Sensemaking

In an age of data, people are expected to make decisions or make deep understandings of information from large data sets. Sensemaking may be defined as the cognitive process by which individuals gather, structure, and generate conceptual models of information pertaining to a specific situation they seek to comprehend. During extended durations of research and analysis, individuals often encounter challenges in navigating enormous datasets while attempting to address intricate tasks. Individuals may experience difficulties in recalling their actions, experiencing uncertainty regarding their role within the broader work, and encountering challenges in determining the appropriate course of action to proceed. In this study, we aim to investigate the effects of a specific intervention on the cognitive abilities (Nguyen et al., 2016).

The process of visual analysis in sensory production primarily involves the extraction of pertinent information, the development of an understanding of significant factors and their interconnections, the formulation and testing of hypotheses, the comparison of various alternatives, the decision-making process, and the communication of findings and decision-making procedures to others. According to the second source (Kai, *Visual analytic for Sensemaking* 2023).

## **1.2 Aims and Objectives**

The objective of this project is to :

1. Develop a browser extension that facilitates the extraction of data from internet searches and offers comprehensive assistance for the process of online purchasing.
2. The proposed approach involves conducting an end-to-end analysis, commencing with an online survey aimed at capturing user behavior. This survey will automatically gather the necessary data to establish a comprehensive comprehension of the significant factors and their corresponding relationships. Subsequently, hypotheses will be formulated and tested, and various items, such as different camera models, will be compared. This comparison will encompass evaluating the prices of the various models as well as assessing the advantages offered by each camera.
3. Generate a visual representation for the purpose of comparing several alternatives. This study offers practical suggestions to enhance the user's shopping experience and improve their purchase efficiency.

## **1.3 Motivation**

In the context of the World Wide Web, a proposed solution to address the potential issue of user disorientation caused by vast amounts of data is the implementation of a visually interactive web page extender. The objective of this system is to assist users who are unable to open multiple page tabs in performing various tasks such as selecting a brand of goods, comparing prices, and evaluating product performance. By conducting an analysis, the system generates radar charts to offer users purchase recommendations that align with their expectations. This significantly reduces the time spent on online shopping and enables users to shop more efficiently and sensibly.

## **1.4 Description of the work**

Currently, it is a prevalent occurrence for customers to need the comparison and selection of many things based on various product attributes. For instance, an individual seeking to make a clothing purchase may allocate effort towards evaluating and contrasting various discounts offered on diverse websites, with the objective of attaining optimal cost-effectiveness. Furthermore, a family with offspring who desires to capture images of their recently born infant may invest a substantial amount of time engaging in a comparative analysis of cameras across various online platforms, with the aim of selecting a device that aligns with their specific requirements. In order to address this requirement, the objective of this project is to create a web extender.

The solution developed in this project will mostly serve customers who are seeking various cameras and seeking to get information according to their specific requirements. The program

will offer the capability to aggregate data pertaining to various camera kinds, including pricing, features, and other relevant characteristics, facilitating their comparison.

## Chapter 2

### Related work

#### 2.1 Data visualization

Data visualisation plays a crucial role in the realms of business and data analysis, and it is widely acknowledged that statistical graphics and data visualisation represent relatively recent advancements within the field of statistics. The fundamental stages in the process of data visualisation encompass the formulation of the research enquiry, data collection, data cleansing, chart type and tool selection, data preparation, and ultimately, chart creation. As a result of the expanding volume of data, there is a growing demand for database specialists who can facilitate efficient and precise data processing, given the escalating expectations for speed and effectiveness in data processing. This study investigates several stages of data visualisation, including its types and numerous uses, in response to the increasing demand. Data visualisation plays a crucial role in scientific research by facilitating the identification, creation, and development of insights. It provides substantial assistance in predicting and conceptualizing ideas (Muskan et al., 2022). However, there are scholars who argue that the utilisation of extracts in data modelling and knowledge remains uninteresting, as humans possess the capacity to comprehend a vast array of diverse material. Hence, the utilisation of visualisation tools is crucial in order to discern and elucidate intricate correlations and patterns that may not be readily discernible from the raw data (Mozzafari & Seffah, 2008).

The tool encompasses the radar chart, which is commonly referred to as the spider chart, for the purpose of data visualization. A radar plot is a visual representation technique used to depict multivariate data by plotting three or more quantitative variables on axes originating from a common point. Typically, the relative positions and angles of the axes do not provide significant information. However, one can employ various heuristics, such as algorithms that maximize total area, to categorize the variables (axes) based on their relative positions. This categorization can unveil diverse correlations, trade-offs, and numerous other comparative metrics (Porter & Niksiar, 2018). In fact, radar charts have some defects, including that too many polygons in radar charts will make the readability of the graph poor, and too many colors may lead to the overall graph is not easy to recognize. Second, if the variable is overmuch, the chart is too dense, will the user the readability of the decline. Moreover, the utilization of radargrams with many dimensions or samples may result in visual clutter and heightened interpretational challenges as the quantity of samples escalates.



## **2.2 online shopping**

Due to the ease of Internet accessibility, people can engage in online shopping. Furthermore, the utilization of online shopping platforms has the potential to enhance retail channels and foster heightened consumer engagement in the purchasing process. There are several factors that contribute to the increased convenience and popularity of online buying. The utilization of social media platforms for shopping purposes offers several supplementary benefits to customers. Firstly, it provides them with an expanded array of products, hence increasing their options for purchase. Additionally, it empowers customers to independently seek out the most favorable prices, enabling them to make informed decisions. Moreover, social media shopping facilitates the access to customer reviews, enabling individuals to gauge the quality and reliability of products. Lastly, it expedites the overall purchase process, streamlining and accelerating transactions. As the internet retail sector expanded, business owners gained insights into the factors that drive consumer behaviour in online shopping. The utilisation of social media competitions within online stores allows business proprietors to actively pursue a competitive advantage in the realm of online sales. In order to gain a competitive edge, it is imperative for online business proprietors to possess a comprehensive comprehension of consumer behaviour, particularly as it pertains to their inclination towards engaging in shopping activities on social media platforms (Bong et al., 2023).

Furthermore, the project research has also placed emphasis on identifying the indicators that would impact customer consumption behaviour. The literature demonstrates that within the context of Amazon's review system, several noteworthy aspects can be seen. Firstly, there exists a personal rating, commonly referred to as a star rating, which is facilitated by Amazon within their online marketplace. This feature enables purchasers to articulate their level of contentment with a given product. Another aspect to consider is that customers have the opportunity to provide further feedback on the product through reviews, which can be evaluated as either valuable or not valuable by other individuals. Additional customers have the ability to provide a beneficial rating for these reviews via the assistance rating feature. Utilising the aforementioned data has the ability to enhance market awareness, participation, and the likelihood of success in selecting product design features for the company (Tang & Guo, 2022).

## **2.3 Literature review**

A substantial corpus of scholarly literature exists pertaining to the field of perceptual visualization, which extensively incorporates principles and theories associated with machine learning. Prominent instances encompass the investigation conducted by SensePath, a venture focused on web-based perception in the digital realm. The article posits the proposition of a comprehensive approach to streamline the qualitative analysis procedure, thereby proving advantageous for various qualitative research conferences. These conferences encompass the examination and subsequent analysis of user participation in sensory tasks, with the aim of delving into their sensory processes at a profound level. SensePath is a tool that autonomously

records a user's perceptual activities, encompassing the examination of their sources, and offers assistance for subsequent investigations through a variety of interconnected perspectives. The tool satisfies additional needs that were acquired during the design discussions (Nguyen et al., 2016). The user's text is already academic and does not need to be rewritten. Furthermore, an additional tool called sensemap has been developed to tackle the issue of users being disoriented while utilizing extensive datasets for the purpose of solving intricate tasks during prolonged periods of exploration and analysis inside the realm of browser-based online perceptual behaviours. Individuals may experience memory lapses, lack of awareness regarding their progress in the overall activity, and uncertainty regarding the next steps to be taken. This study develops an enhanced conceptual framework of perception, drawing upon Pirolli and Card's theory, with the aim of providing a more comprehensive understanding of the process of discovery. Specifically, users engage in an iterative process of collecting task-relevant information sources, effectively organising and managing these sources, and ultimately conveying their findings to others. The SenseMap tool is designed to capture the sources of a user's perceptual activities in an automated manner. It accomplishes this by providing a multi-linked view that visually presents the collected information. This tool aims to enhance users' ability to plan and effectively convey their discoveries. The user's text does not contain any information to rewrite (Nguyen et al., 2016).

## Chapter 3

### User description

#### 3.1 Target user

In the preliminary phase, a selection of target groups was established with the aim of examining the distinct requirements of various user segments:

1. **Budget-Conscious Shoppers:** Users who are budget-conscious and actively look for ways to save money while shopping online. They are interested in using price comparison features and finding coupons and promo codes to get the best possible prices.
2. **Frequent Online Shoppers:** Users who prefer the convenience of online shopping over physical stores and make regular purchases through e-commerce platforms.
3. **Impulse Buyers:** Users who enjoy discovering new products and sometimes make impulsive purchases. They might benefit from personalized product recommendations offered by the extension.
4. **Consumers who are very concerned about information security:** Usually, they will give priority to their own information security risks when shopping, and they may hope that the extension program can protect their personal information to a certain extent, respect their personal privacy, and not excessively collect personal data information.

#### 3.2 User requirements

In order to further meet the relevant needs of users, this project identified a part of potential user groups in the early stage of the research, and conducted relevant interviews to obtain the relevant main needs of users, including:

1. The coupons and coupon code search. Users expect the extension to help them find coupons or promo codes related to their current purchase, so as to save the most time and provide better purchase suggestions.
2. Personalized recommendations for different user groups. Users hope that the extension program can recommend relevant products that meet their expectations according to their different needs. So as to give users better purchase suggestions.
3. Project front-end interface is beautiful. Users think that the beautiful front-end interface can better attract them to use the extension program, and make it easier and easier for them to use the related functions of the extension program, thus improving the user experience.
4. Relevant data protection. Users expect that the extension will protect their information security, respect their privacy, and not excessively collect their personal information.

### **3.3 Function requirements**

The tool's interface is composed of three different panels, are respectively introduced panels, information collecting and comparing the panel.

**Introduction Panel:** The main purpose of the introduction panel is to provide an overview of the source, main functionality and intended use of the extension.

**Information collection panel:** The information collection panel displays detailed information about the camera that the user has queried, including price, features, brand, and other attributes related information. The process USES the crawler technology, users don't need to manually enter search information can automatically fill in a form. For instance, when a user is engaged in the process of searching for a telephoto camera manufactured by Canon, the pertinent details such as the brand name, price, and wheelbase of the camera will be displayed inside the designated form. The user has the option to manually include additional qualities if they perceive a lack of information for comparison.

**Comparison Panel:** After the completion of data collection by the information collection panel, the user may go to the comparison screen. The comparison interface utilizes radar charts as a means of delivering feedback on the outcomes of the data analysis. This distinctive characteristic of the application enables users to visually assess and compare the price or performance of the relevant cameras using the radar charts.

# Chapter 4

## Methodology

### 4.1 Overview

This study from the theory of Pirolli and Card found an enhanced consciousness of the concept of relevant framework, aims to provide a more comprehensive understanding of the discovery process. Specifically, users engaged in an iterative process of gathering task-relevant information sources, effectively organizing and managing these sources, and ultimately communicating their findings to others. SenseMap tools designed to capture user perception activities by way of automation. It does this by providing a multi-linked view that visually presents the collected information. The tool is designed to enhance the user planning and the ability to effectively communicate their findings. The user's text does not contain any information that needs to be rewritten. To enhance the understanding of user behavior and preferences, the tool gathers data on the user's page clicks and the accompanying user logs. During the process of data collecting, the data is identified and transmitted to the service, where it undergoes a specific level of processing before being stored in the database.

Furthermore, the establishment of data visualisation is facilitated through the careful selection of suitable data metrics. Based on a thorough examination and analysis of the data presented on the user's browsing page, we have determined that the important indications for comparison will be the camera brand, camera price, camera focal length, camera rating, and other relevant factors. These indications assist users in making informed selections during the process of acquiring things.

The radar graph has been selected as the ultimate visual interface for this project. Based on the data that has been collected and comprehended. Radar charts are used due to their ability to effectively represent several variables and accurately depict the interrelationships among them. Radar charts facilitate the visualisation of distinctions and correspondences among diverse data points, hence expediting the process of comparing products. Simultaneously, radar charts has a visually captivating quality. The distinctive layout of a radar graphic serves to capture the user's focus.

## 4.2 Data Management

### 4.2.1 Data Collection

The data collection for this project involves retrieving the log information of the user. The present assortment exclusively compiles the user's historical log data pertaining to their interactions with the Amazon camera search page, encompassing the URL address and the timestamp of each browsing session.

Upon acquiring the user log data, the team made the decision to employ crawling techniques in order to extract more pertinent information from the webpages accessed by the user. This information encompasses the camera brand, camera price, camera purchase rating, number of camera reviews, and camera pixels.

A web crawler is a software application that autonomously retrieves online pages from the World Wide Web on behalf of a search engine, serving as a crucial element inside the search engine's infrastructure. A conventional web crawler commences by receiving the URL of one or multiple beginning web pages. It proceeds to retrieve the URLs present on the first web page, and while crawling across the web page, it consistently extracts new URLs from the current page and adds them to a queue. This process continues until specific termination criteria of the system are satisfied (*Web Crawler\_Baidu Encyclopedia* 2023). The libraries encompassed inside the system are the request library, parser library, and storage library. This project utilises the request library, which is developed based on urllib, a synchronous HTTP request library that halts programme execution until a response is received from the server before proceeding to the subsequent step. The primary rationale behind selecting the request library for this project lies in its straightforwardness and user-friendliness. It offers an intuitive application programming interface (API) that simplifies the process of making HTTP requests, while also boasting a wider range of features. Furthermore, it is evident that the request library possesses the capability to effectively manage exceptions, hence facilitating the process of obtaining information.

The field commonly used several crawler frames including Scrapy, Crawley, Portia and Newspaper, etc. For this project, the Crawley framework was chosen as the preferred tool for data crawling. Python web crawlers are used to collect, process, and store data. It is often described as programming or scripting initiated from one or more urls (Ma & Zhang, 2021). With web scraping, data can be retrieved in any form depending on the desired context. After retrieving these data, they can be transformed into a more desirable format depending on the motivation of the application. (Abodayeh et al., 2023). The Crawley framework provides some user-friendly features that help get started quickly and speed up development. In addition, it facilitates automatic data crawling by using page structure and predefined rules, which greatly reduces coding requirements. Simultaneously, it provides support for diverse data storage capabilities, enabling the storage of extracted data in various formats such as databases, JSON

files, and others. This feature caters to the requirements of the project while enhancing the adaptability of later data processing tasks.

### 4.2.2 Data Storage

In relation to the storage of data, there exists a plethora of storage methods that encompass database storage, local file storage, cloud storage, distributed storage, cache storage, in-memory storage, API services, distributed storage and processing, as well as data warehouse data marts. The study project has opted for the utilisation of the local file storage approach. The data is stored in the JSON format. JSON is a widely accepted file format and data interchange format that is open in nature. It is used for the purpose of storing and sending data objects, which are composed of attribute-value pairs and arrays (or other values that may be serialized). The data is represented in a manner that is easily understandable by humans. The data format in question is widely utilised for various purposes in electronic data transmission, encompassing web applications that include servers (*JSON Wikipedia* 2023). The selection of JSON as the preferred data format was based on its inherent conciseness in comparison to XML. This attribute facilitates the process of troubleshooting during compilation, hence enhancing the overall ease of debugging. JSON is characterized by its lower magnitude, resulting in enhanced efficiency in terms of writing and parsing. Furthermore, the utilization of compression technology during transmission contributes to additional broadband savings (Chaser's Blog, *A brief summary of the advantages of JSON* 2023).

### 4.2.3 Data Understanding

Gaining a comprehensive understanding of the significance of the data assumes particular significance in order to enhance the efficacy of conveying the outcomes of the studies. The primary objective of this research is around the procurement of a camera, thereby necessitating a comprehensive comprehension of the pertinent camera-related data. To enhance comprehension of the data, a comprehensive investigation was undertaken to ascertain the criteria for evaluating the efficacy of a camera. It was determined that the dimensions of the frame and the level of resolution are pivotal factors in determining the quality of a camera. Simultaneously, empirical research has demonstrated that the conventional approach to facilitate product searches involves the utilisation of ratings. Individuals have the ability to submit ratings that are clearly stated or expressed. The careful balance between the provision of explicit ratings and the level of user effort required to get perceived advantages is crucial in order to obtain accurate findings from explicit ratings on a scale (Ojtáš & Peška, 2014). Hence, the present data metrics incorporate both ratings and the quantity of comments as potential indicators of user interest, serving as a foundation for additional data studies. This study will utilize these metrics as a fundamental foundation for data analysis and afterwards present them in radar analysis charts during a later phase.

## 4.3 Radar Chart

Given the presence of various variables to be compared, radar charts were selected as the preferred visual representation for the final result of this project. The selected criteria for comparison are camera price, camera pixel count, camera frame rate, camera rating, and camera review count, which have been thoroughly investigated and analyzed. The implementation of this module utilizes various libraries, including E-chart. E-Chart is a data visualization chart toolkit that is implemented using JavaScript. It offers a range of visually appealing and interactive charts for representing data in a personalized manner (*Echarts\_Baidu Encyclopedia* 2023). E charts is constructed using HTML5 technology, making it very compatible for the purpose of data visualisation within web applications. HTML5 is a JavaScript framework designed for the creation of dynamic and interactive data visualizations within a web browser environment. The utilization of the Scalable Vector Graphics (SVG), HTML5, and Cascading Style Sheets (CSS) standards is observed. As a widely used data visualisation framework, E-chart facilitates the creation of text and graphical charts, is highly compatible, and offers a wide range of customisation options for chart settings.



## Chapter 5

### Design

The project will be developed as a Chrome browser extension, with an interface that consists of three interfaces: an introduction interface, an information collection interface, and a data visualization interface. This will be based on the knowledge from the literature review and in accordance with user requirements.

#### 5.1 Chrome Browsers Extension

A chrome extension, alternatively referred to as an add-on or plug-in, is a compact software application that enhances the capabilities of a web browser. There are multiple rationales behind the conceptualization of this project as chrome browser extensions. Chrome browser extensions offer convenient accessibility, eliminating the need for users to visit to distinct websites or applications. Improving ease of use has the potential to increase user engagement. In addition, browser extensions improve the user's browsing experience, allowing users to experience the relevant functions of extensions without leaving the browser environment. In terms of development efficiency, browser extensions have the capability to incorporate several functionalities, including but not limited to form filling and data extraction. This proposed approach offers a viable method to address the challenge of enhancing productivity by developing and implementing extensions that automate certain processes, hence optimizing the efficiency of repeated procedures.

The interface of the extension is constructed using the Vue framework, which leverages Vue.js to enable the extension of HTML through the utilization of directives, which are HTML attributes. These directives offer functionality to HTML applications and are available in the form of built-in directives or directives written by the user (Wikipedia, *Vue.js* 2023). The framework is widely recognized in the field of front-end programming using JavaScript, and it has numerous benefits for the construction of online applications. Vue development is characterized by a higher degree of explicit componentization compared to react development. Additionally, it is perceived as being more intuitive and simpler to design and administer. Vue leverages two-way data binding, which offers a more convenient approach to achieving synchronicity between the visual representation and the underlying data. The incorporation of this feature enhances the intuitive nature of the development experience.

Within the development interface of this project, three distinct browsing interfaces have been established. These interfaces are categorized as the introduction interface, the information gathering interface, and the data visualization interface. The information collection interface employs a structured format to gather brand information, price information, and function-related details of cameras on the browsing page during the user's purchase process. Additionally,

the interface includes a manual filling function, enabling users to manually input any uncollected data. In addition, the interface also includes an automatic filling function, allowing users to automatically fill in the form without manually entering any unclaimed data. This choice enables users to conveniently gather information and obtain meaningful insights. During this procedure, customers have the ability to obtain pertinent recommendations.

The outlined procedure is as follows:

The user initiates a click action on the webpage in order to engage in browsing and purchasing a camera. This action triggers a web crawler to collect data from a predetermined website. Simultaneously, the interface is activated, allowing the front-end to retrieve the obtained results and automatically populate the form. Once the user has acquired the necessary data, they can proceed to manually transition to the third interface. This interface allows the user to access a radar map that visually represents the information. By examining this chart, the user can make comparisons and derive recommendations based on the data presented.

The second component, referred to as the information collecting portion, serves the purpose of facilitating data analysis by enabling users to automatically extract a substantial volume of data from a website and populate it. The third component of the system is the radar chart for data visualization. Its purpose is to provide users with a concise visual representation of product distinctions, facilitating informed decision-making and enabling the selection of a camera that aligns with their specific requirements. This feature significantly reduces shopping time and enhances the overall shopping experience.

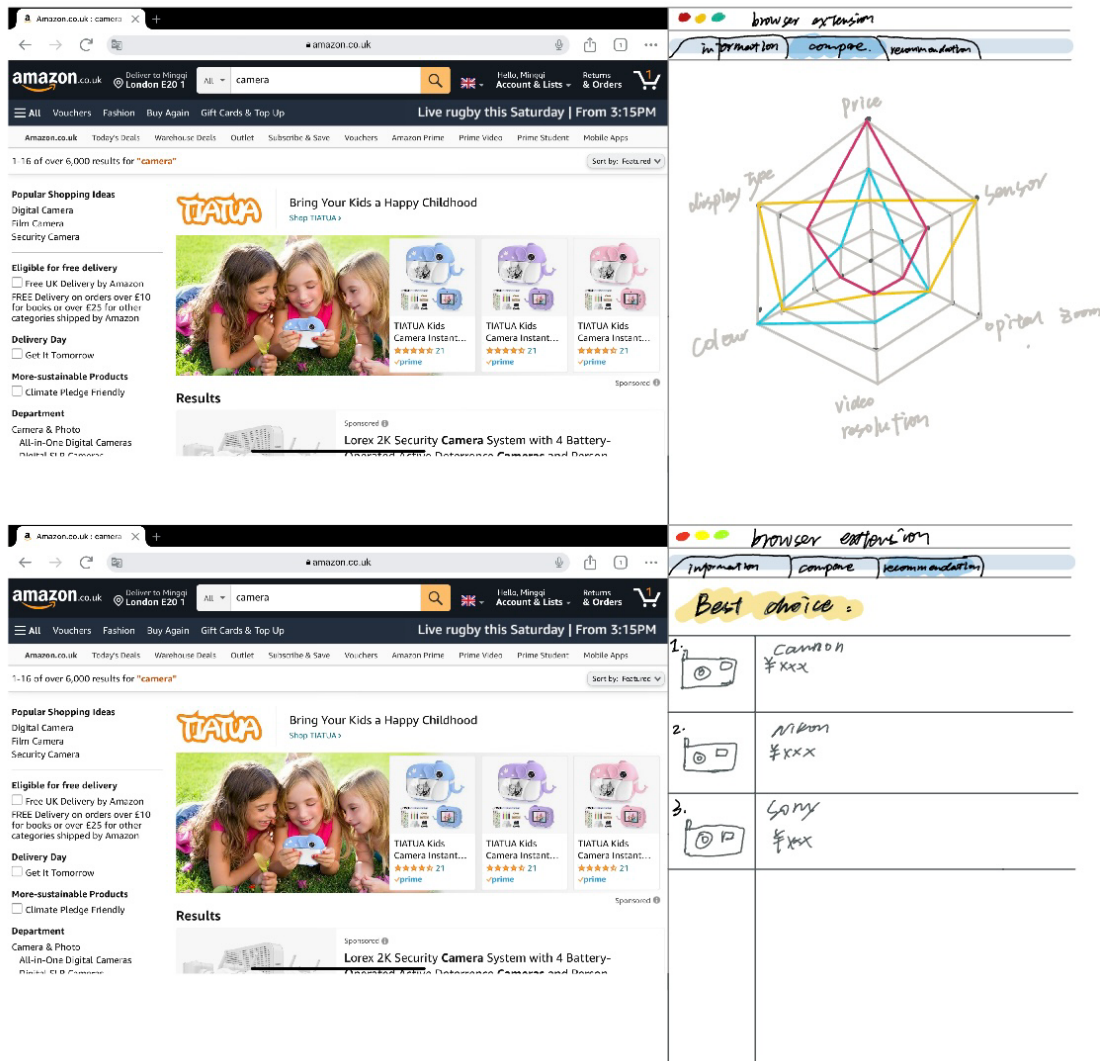
### **5.1.1 Early Preset**

During the first stages of the project, the primary objective for the information collecting page of this extension programme is to incorporate a selection box on the user interface. This selection box will enable users to manually choose the appropriate fields from a drop-down list. Subsequently, users can submit relevant data by clicking the designated "Add" button.

Ultimately, the data gathering process is concluded by presenting the gathered information on the table.

As depicted in the visual representation:





## 5.1.2 Post Implementation

However, throughout the actual implementation process of a project, the collection of user suggestions serves to enhance the project. When designing the information collecting page, it is important to examine the potential impact of allowing users to manually input data using a drop-down menu. This approach may result in an increase in the number of operation steps required by the user, which can lead to a perception of inconvenience and complexity. Furthermore, the available choices within the drop-down menu are restricted, hence exacerbating the constraints on user selection. To address this issue, the project will encounter increased complexity in data gathering due to the utilisation of web crawling technology. This entails the automated retrieval of user browsing data and the subsequent automated population of information collection forms. As demonstrated in the following illustration:

# User History Camera

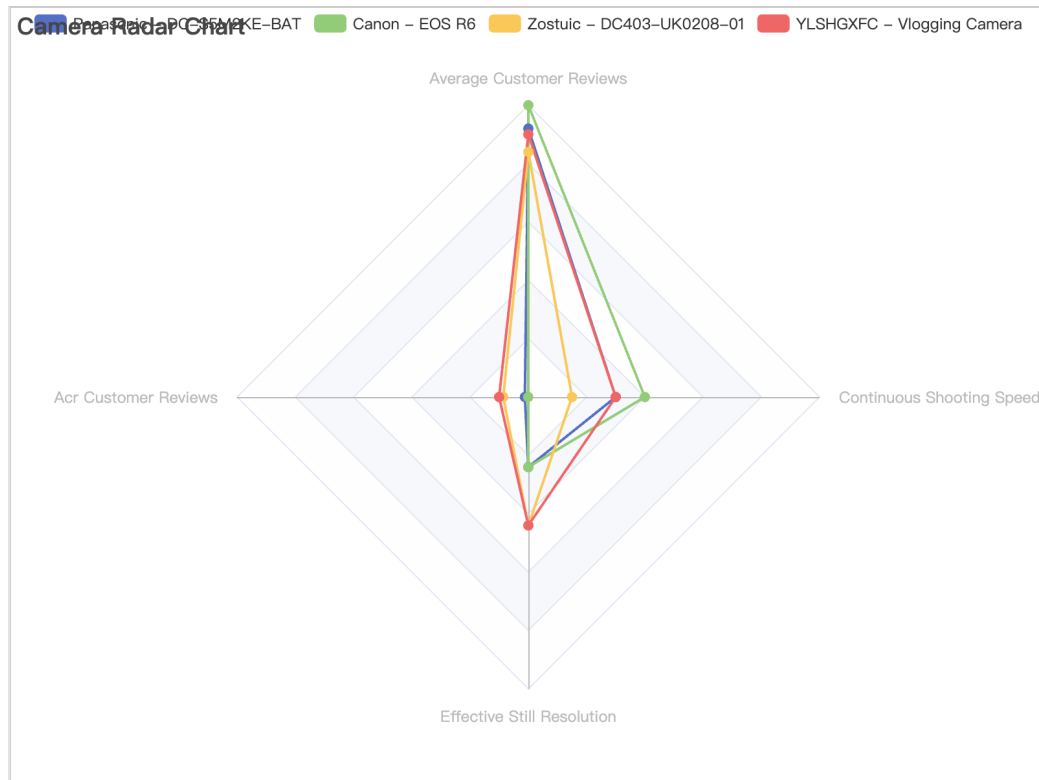
[HOME](#)[COMPARE](#)

Display the item information

brand	model_name	url
Panasonic	DC-S5M2KE-BAT	<a href="https://www.amazon.co.uk/Panasonic-Mirrorless-Unlimited-Recording-DMW-BLK22/dp/B0BSNSQBJ9/ref=sr_1_1_sspa?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-1-spons&amp;ufe=app_do%3Aamzn1.fos.cc223b57-2b86-485c-a85e-6431c1f06c86&amp;sp_csd=d2lkZ2V0TmFtZT1zcF9hdGY&amp;psc=1">https://www.amazon.co.uk/Panasonic-Mirrorless-Unlimited-Recording-DMW-BLK22/dp/B0BSNSQBJ9/ref=sr_1_1_sspa?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-1-spons&amp;ufe=app_do%3Aamzn1.fos.cc223b57-2b86-485c-a85e-6431c1f06c86&amp;sp_csd=d2lkZ2V0TmFtZT1zcF9hdGY&amp;psc=1</a>
Canon	EOS R6	<a href="https://www.amazon.co.uk/Canon-Mark-II-Mirrorless-24-2-megapixels/dp/B0BT4Z94RY/ref=sr_1_4?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-4">https://www.amazon.co.uk/Canon-Mark-II-Mirrorless-24-2-megapixels/dp/B0BT4Z94RY/ref=sr_1_4?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-4</a>
Zostuic	DC403-UK0208-01	<a href="https://www.amazon.co.uk/Vlogging-Rechargeable-Students-Children-Beginners/dp/B0BV6FPNY1/ref=sr_1_3?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-3">https://www.amazon.co.uk/Vlogging-Rechargeable-Students-Children-Beginners/dp/B0BV6FPNY1/ref=sr_1_3?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-3</a>
YLSHGXFC	Vlogging Camera	<a href="https://www.amazon.co.uk/Digital-Rechargeable-Vlogging-Teenagers-Beginners%E2%80%82%E2%80%82Black%E2%80%82/dp/B0BYD8541M/ref=sr_1_2_sspa?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-2-spons&amp;sp_csd=d2lkZ2V0TmFtZT1zcF9hdGY&amp;psc=1">https://www.amazon.co.uk/Digital-Rechargeable-Vlogging-Teenagers-Beginners%E2%80%82%E2%80%82Black%E2%80%82/dp/B0BYD8541M/ref=sr_1_2_sspa?crid=XUB08KS9ZF2E&amp;keywords=camera&amp;qid=1694007624&amp;prefix=camera%2Caps%2C198&amp;sr=8-2-spons&amp;sp_csd=d2lkZ2V0TmFtZT1zcF9hdGY&amp;psc=1</a>

In relation to the configuration of the indicator setting inside the radar chart, notable enhancements have been made in this project. Due to the potential limitation of the radar map's ability to display the index for text, alternative indications such as colour cameras with scheduled late indicators, camera brands, and textual indicators for scoring and commenting on the number of cameras are considered. These metrics have the dual purpose of promoting standardisation in data processing and enhancing the user's comprehension of the radar chart. The specific page is depicted in the accompanying image:

Compare Chart



## **5.2 Related Processes**

### **5.2.1 Data Link**

The purpose of this extension's data link source is to account for the user's generation of visiting websites, clicking, and other associated events. The extension software is utilised to detect the website and transmit the detected time to the service. The service worker assumes a pivotal position in the communication between the server and the browser. Once registered on a website, the service worker gains the ability to intercept all ongoing activities within the website. If a service worker is registered on the website, it has the capability to intercept all ongoing requests made by the website. In order to determine the appropriate course of action, it is necessary to evaluate if a request to the server has to be initiated. If such a request is indeed required, it will be forwarded to the server. Conversely, if the cache can be utilised directly, the response will be retrieved from the cache without any further forwarding to the server. This enhancement significantly enhances the user's browsing experience. Once the request is submitted to the service, the service proceeds to record the relevant events into a queue. Subsequently, the crawling process retrieves and consumes these events from the queue. The Amazon Camera's camera-related product pages are systematically crawled by a web crawler, which retrieves the data, parses the pages, and afterwards stores the information into a database. Two tables are inserted into the database, with one is designated for the purpose of documenting the browsing history of the user, while the second table is intended to store the relevant information pertaining to the products.

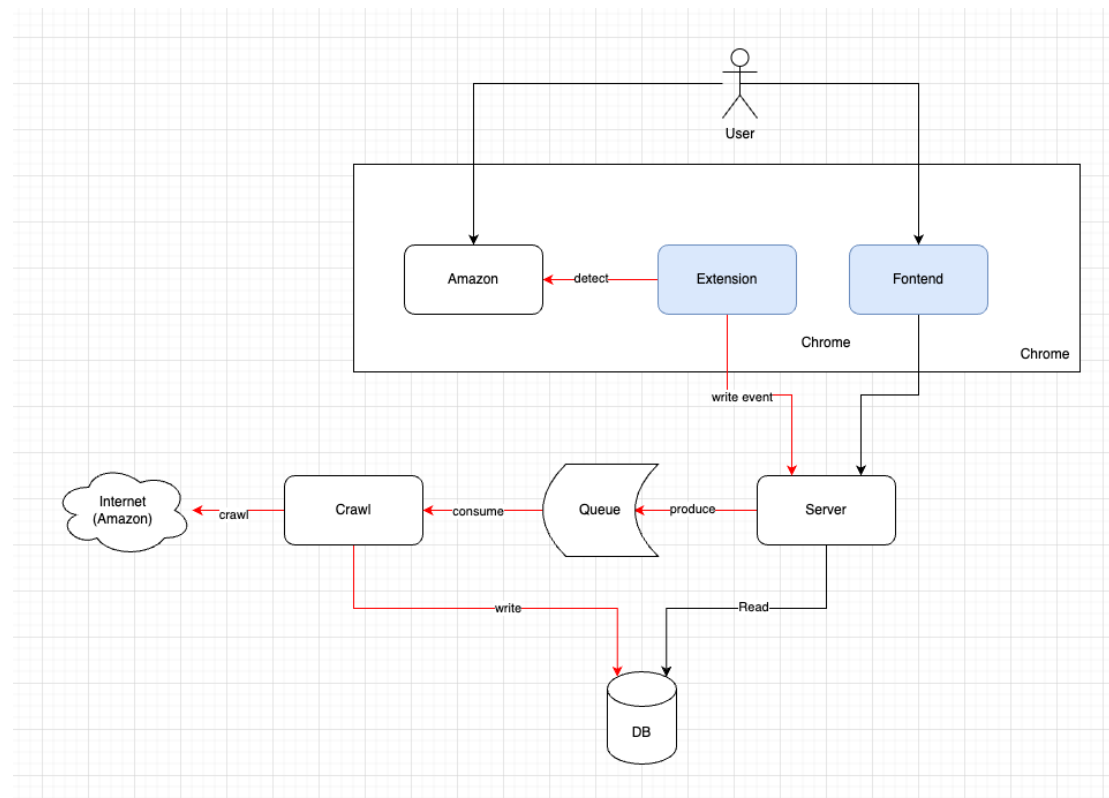
### **5.2.2 Data presentation**

In this instance, the data is temporally regulated by the front-end and is originally configured to initiate a data retrieval process from the service at regular intervals of ten minutes. The service, in turn, provides the user with the most recent historical information for five distinct products that have been accessed.

## 5.3 Data processing

### 5.3.1 Data collection

Prior to data analysis, user log data is preprocessed with the aim of making the data suitable for data analysis. The flow chart of data management is shown in the figure:



The primary data were gathered from the Amazon webpages that pertain to camera-related products. A judgement condition is established throughout the data crawling process to assess the URL's validity, namely by determining whether the URL prefix matches `www.amazon.co.uk`. This is due to the inherent limitations associated with evaluating the URL. Currently, the processing and analysis of data on Amazon is constrained by the limitations of its anti-crawling mechanism, resulting in only a limited quantity of data being accessible for crawling. The primary origin of user logs comprises individuals with the intention to make a purchase of a camera. The collection of camera-related data from the user's history browsing pages is facilitated through the utilization of web crawling technology. This data encompasses several aspects such as the brand of the camera, its price, pertinent specifications, and ratings. Furthermore, with the user's surfing duration, the titles of the visited web pages are also gathered. The designated duration for the execution of a specific task is 10 minutes, during which the user's activity records are kept in a JSON file.

After the original data was collected, it was chosen to be stored in the form of file storage. In the database, two forms were set up, one for the user's browsing history and the other for information about the product. Regarding the execution of the web crawling process, this

project uses BeautifulSoup, a Python package for parsing HTML and XML documents (including tags with formatting errors, i.e., non-closed tags, hence the name tag soup). It creates a parse tree for parsing pages that can be used to extract data from HTML(Wikipedia, *Beautiful Soup (HTML parser)* 2023). It also automatically converts input documents to Unicode encoding and output documents to utf-8 encoding.

Beautifulsoup has many features and was chosen because it provides an intuitive API that makes parsing web pages very easy. BeautifulSoup fixes tag nesting errors and handles broken or under-specified html and xml. it supports multiple parsers and has a high degree of flexibility, i.e., it allows data selection and extraction to be done according to tags or attributes or by textual content. content for data selection and extraction. In addition, BeautifulSoup can convert the parsed data into various data structures for further analysis and processing.

After efficiently collecting the required data, the collected data is rigorously analyzed.

### **5.3.2 Data analysis**

This section mostly discusses the process of selecting indicators for data analysis. Initially, the primary indicators were selected based on considerations of the prevailing buying patterns and psychology of customers, as well as the brand, price, associated links, and focal length of the camera. After careful deliberation over the use of multi-dimensional data, the decision was made to employ radar charts as a means of presenting the outcomes of data analysis. Simultaneously, the limitations of radar charts, including the inconvenience of displaying analytical findings with textual data, were also considered. Ultimately, the decision was made to employ camera pricing, camera pixel count, camera frame rate, camera rating, and number of reviews as the primary variables for the ultimate data visualisation.

The project utilises the E charts icon library for the implementation of the radar chart presentation. The E charts library offers numerous advantages as a widely used data visualisation framework. The icon library provides support for a diverse range of icon kinds. In addition to the radar charts utilised in this particular project, it encompasses various other chart types such as line charts, bar charts, pie charts, scatter charts, and so forth. Echarts offers extensive customization capabilities and a wide range of setup choices, facilitating the customization of chart aesthetics and functionality. Furthermore, it possesses interactive features and exhibits cross-platform compatibility. E charts is constructed using HTML5 technology, making it very compatible for the purpose of data visualisation within web applications.

In summary, the utilisation of radar chart presentations serves as a crucial factor in facilitating users' ability to assess and evaluate the performance of different categories, hence enhancing shopping efficiency.



## Chapter 6

### Implementation

#### 6.1 Data Range

The primary data source utilised in this study is derived from the page information of Amazon consumers engaged in camera searches. To streamline the process of data collection, the initial design of the buttons pertaining to data input by users was substituted with a functionality that automatically retrieves the information from the active browsing page. The implementation of this approach guarantees the currency of the data, as well as the user-friendly nature of the system.

#### 6.2 Concrete Implementation

This section is dedicated to elucidating the particularities of the methodologies employed in the present study.

In this instance, the primary tool employed was the `urllib` module from the Python library. The Python function `urllib.request` is utilised for retrieving URLs. The library offers a straightforward user interface through the `urlopen` function, enabling the retrieval of URLs utilising various protocols. Additionally, it provides a more intricate interface to handle typical scenarios such as basic authentication, cookies, proxies, and other related functionalities. The module offers a straightforward interface for fundamental URL modification, significantly enhancing the ease of managing URLs. Simultaneously, the software exhibits a high degree of extensibility and may be integrated with additional libraries to accomplish more intricate objectives. One such approach is to utilise the `'urllib'` library in order to transmit an HTTP request. Subsequently, the `'json'` library can be employed to analyse and react to the data contained inside said request. The `urllib` module possesses numerous features that make it a compelling candidate for implementation in this project. However, it is not without its limitations. For instance, it lacks support for complex functionalities. However, for the purpose of this project, only a few fundamental URL manipulation tasks are required, making `urllib` the most advantageous choice.

#### 6.3 Data storage and Interaction

In order to do data analysis, it is important to obtain the pertinent datasets containing user log information. Due to this rationale, the project established a similar form for the purpose of data collecting. To optimize user experience and increase user efficiency, the form has been

designed to automatically populate the appropriate user browsing dataset, eliminating the need for manual input of such information by the user. The form contains data pertaining to the brand of the camera, the specific model type of the camera, as well as the URL information of the user's browsing website. Upon retrieval of the dataset request, the associated data is promptly updated.

The database chosen for this project is based on the utilization of `sqlite3`. SQLite3 is a database management system that is designed to be embedded within other applications. SQLite3 is considered a lightweight database management system due to its minimalistic design, which eliminates the need for external database servers or configuration files. Consequently, SQLite3 consumes little system resources. Furthermore, the utilization of the database does not necessitate intricate preparation. Simply incorporating the library files into your application enables its functionality. Furthermore, it should be noted that `sqlite3` operates as a single-user database, wherein each database connection functions independently from one another. This characteristic serves to mitigate the challenges associated with multi-user concurrency. SQLite3 offers transaction functionality that ensures the consistency and integrity of the data. Simultaneously, `sqlite3` exhibits cross-platform compatibility and adheres to the ACID (Atomicity, Consistency, Isolation, Durability) principles, hence ensuring data consistency.

It is important to acknowledge that while `sqlite3` generally fulfills the requirements of this project, its ability to effectively handle substantial data volumes or highly concurrent applications may be limited, hence hindering its optimal performance.

In terms of front-end interaction, the team made the decision to utilize the bower package management. The package manager in question has resemblance to the node and npm systems. The outcome is contingent upon the use of node and npm. One of the primary factors influencing the selection of bower is its user-friendly nature. Bower exhibits a high degree of cross-platform compatibility, enabling its execution on several operating platforms. Moreover, it employs a flat dependency structure, thereby mitigating conflicts and iterations that may arise across dependencies.

Nevertheless, it is imperative to acknowledge that Bower has ceased to get maintenance and is no longer the preferred tool for contemporary front-end development practices. The utilization of npm has superseded the use of Bower as a software development tool.

## 6.4 Message Queues

The concept of message queuing was introduced in view of the possible complexity of having to execute many processes sequentially when dealing with interfaces. Message queuing is a mechanism used for inter-process communication or communication between multiple threads within a single process. It involves the sequential processing of a series of inputs, usually from the user, using stored queues in the software. Message queuing provides an asynchronous communication protocol where each entry in the queue contains a comprehensive description

of the information such as timestamps, input device type and specific input parameters. Therefore, the sender and receiver of a message do not need to interact with the message queue at the same time. Communications are stored in the queue until accessed by the receiver (Wikipedia, *message queue* 2022). Message queues have numerous advantages and are an effective tool for solving this problem. Message queues facilitate decoupling of applications and support inter-application communication. The sender need to send a message to the queue, without having to know the receiver, thus minimizing the interdependence between applications. Message queue can in many workload distribution between the client, to ensure that different service or work nodes in a fair way of management work. This feature enhances the scalability of the system. In addition, the message queue to realize asynchronous communication, the sender does not need to wait for a response from the receiver. Instead, messages are sent quickly to the queue, allowing execution without interruption. This feature greatly improves the response ability of the system. Message queues are highly fault-tolerant, facilitate data transfer across different networks and protocols, and can handle delayed jobs or time-sensitive operations. In addition, message queues can provide temporary storage when the data generation rate of producers exceeds that of consumers, as a means of buffering data and preventing data loss.

Realize that technology is very important every coin has two sides, the message queue is a typical example. While this approach has a lot of advantages, but it also has some shortcomings. Using the message queue will add to the complexity of the system. To ensure the reliability of message transmission, the queue server must be effectively managed and maintained. The operation and maintenance of a message queue system requires the allocation of additional resources and costs, including hardware, software, and monitoring tools. In addition, the task of ensuring data consistency within message queues is a major obstacle. Especially when messages are sent to different users, it becomes a more prominent problem to maintain data consistency.

```
{
  "url": url, // detected url
  "user_brower_time": user_brower_time, // user brower time
  "n_try": 0 // add one if failed to proce
}
```

## 6.5 Configuration of the User Interface

In relation to the front-end design of this project, the selected programming languages include HTML, CSS, and JavaScript. The Hypertext Markup Language (HTML) is a widely accepted standard markup language utilized for the creation of web pages. HTML is a fundamental technology that is frequently employed, in conjunction with CSS and JavaScript, by several websites for the purpose of crafting user interfaces for web pages, online apps, and mobile applications (Wikipedia, *HTML* 2023). The CSS language is a computer language that is utilized for the purpose of incorporating styles, such as fonts, spacing, colors, etc., into structured documents, specifically HTML documents or XML applications. This language is defined and upheld by the World Wide Web Consortium (W3C) (Wikipedia, *CSS* 2023). The

JavaScript language is a programming language that encompasses multiple paradigms, including object-oriented programming, imperative programming, and functional programming. It is a high-level interpreted language that relies on functions such as prototypes and headers. The software offers a range of techniques for manipulating text, arrays, dates, and regular expressions (Wikipedia, *JavaScript* 2023). The three aforementioned languages are universally recognized and supported by prominent web browsers as standard languages for the purpose of web development. This ensures that the extension functions effectively across a diverse array of browsers and devices. Simultaneously, a substantial and engaged developer community exists to support these programming languages, offering an extensive array of documentation, tutorials, and open-source tools for constructing online apps that can function across several platforms. Consequently, this support network facilitates efficient development by reducing the expenditure of time and resources. Simultaneously, these languages provide a significant degree of adaptability, enabling the tailoring of software to suit the requirements of a particular project, as well as the creation of unique user interfaces and interactions. Furthermore, there is a diverse ecosystem comprising frameworks like as React, Vue.js, and others, which greatly expedites the progress of this undertaking. HTML, CSS, and JavaScript possess numerous advantages in the realm of front-end development; nevertheless, they are not without their own set of obstacles and restrictions. When confronted with intricate front-end application challenges, one may encounter issues such as prolonged loading durations and sluggish page rendering. Simultaneously, JavaScript operates on the client side, rendering it susceptible to potential security vulnerabilities and similar risks. In the process of development, it is imperative to carefully select appropriate tools and frameworks that effectively address the specific challenges in accordance with the project's needs.

The initial phase of this project is the selection of the dependent package for the jQuery ECharts library. The thorough description of the exact reasons for selecting E charts can be found in section 5.2.2. The framework version being utilized is Chrome Extension Manifest 3, which represents a significant milestone in the development of our extension platform vision. The latest versions have demonstrated enhancements in performance, while the introduction of Manifest V3 has brought forth a multi-threaded architecture enabling extensions to execute parallel processing in the background. Consequently, this advancement has resulted in improved performance and increased responsiveness of extensions. This feature enhances the user-friendliness of extensions without causing any noticeable decrease in browser performance. Manifest V3 offers significant security upgrades aimed at mitigating the potential threats posed by rogue extensions and enhancing user security. Simultaneously, Manifest V3 has undertaken a rewrite of the extension's application programming interface (API) with the aim of enhancing clarity and ensuring greater consistency. It is important to acknowledge that Manifest V3 also brings about certain modifications, like the elimination of backend pages and certain current APIs. However, it might be argued that Manifest V3 is the most suitable option for this project.

## 6.6 The establishment of the backend system

The backend of this extension application is implemented using the Python programming language, with the Flask framework being selected as the service framework. Flask is a lightweight web development framework written in Python and is built upon the Werkzeug toolkit. The Flask framework can be likened to a kernel, as it serves as the core foundation. However, the majority of additional functionalities necessitate the utilization of third-party extensions, which must be developed. The Flask framework does not provide built-in support for databases or form validation tools by default. In addition to Flask and Django, there exists a Python web framework that can be considered lightweight when compared to Django. Django and Flask present different characteristics. Django is famous for its comprehensive properties and the Flask is famous for its lightweight and simple design. Flask implements some of its functionality through customizable third-party extensions. Flask's extension system supports a wide variety of functions, which makes it highly customizable. This property is the most important feature of Flask. Flask framework provides developers with considerable flexibility, they can be adjusted according to the specific needs of the project application of the structure and components. The Flask is very suitable for simultaneous development service and API. It is able to effectively deal with HTTP requests, and support the implementation of a RESTful architecture. Flask is well suited for the development of microservices and APIs, handling HTTP requests and communication efficiently while providing RESTful architectural style support. In addition, the Flask can be easily installed in many web server, so it can be used for deployment of cloud hosting. The Flask framework benefits from an active development community that provides extensive documentation, tutorials, and community tools. As a result, developers can easily get support and knowledge to help them use Flask. In general, Flask is a robust and user-friendly web services framework suitable for projects of different sizes.

Despite being a widely utilized framework for web services in Python, Flask exhibits certain restrictions. Although Flask is known for its simplicity, it may be essential to incorporate additional extensions and components in order to accommodate large and intricate projects. However, this might result in the development of complex and challenging-to-maintain projects. The Flask framework lacks an inherent object-relational mapping (ORM) system. Consequently, in order to connect with databases, it is necessary to configure a separate ORM library. The basic feature set of Flask is somewhat restricted, necessitating its augmentation with additional functionalities. The basic feature set of Flask is quite limited, requiring the usage of extensions to incorporate typical functionalities such as user login, authorisation, and form processing, among others. Simultaneously, inside the flask ecosystem, there exists a multitude of extensions and plug-ins. However, it is important to note that these extensions and plug-ins exhibit variations in terms of their quality and maintenance. Consequently, it is imperative to exercise caution while selecting the appropriate plug-in for your project. The utilization of technology encompasses two distinct facets, necessitating the selection of an acceptable framework structure tailored to the specific requirements of diverse projects. Flask is deemed adequate for fulfilling the requirements of providing assistance for this project.

A comprehensive introduction of the project's crawler utilizing the request and beautifulsoup4 libraries can be found in chapter 5.2.1, which covers relevant content. The database utilized in this context is SQLite3, which is extensively elucidated in chapter 6.3.

## **Chapter 7**

### **Evaluation**

#### **7.1 User experience**

For this study, a sample of five individuals who volunteered were chosen to participate in the extension programme. The purpose of their involvement was to gather data and gain a comprehensive understanding of their user experience through conducting interviews.

The evaluation findings from users indicate a favourable perception of the system, with a majority of participants expressing that it is user-friendly and facilitates their understanding and operation. Additionally, users find that the system offers assistance in making informed purchase selections. However, the page design exhibits a certain level of simplicity, lacking in aesthetic appeal. The findings of this study suggest that the radar map interface is user-friendly and has the potential to enhance the shopping experience. Furthermore, it is crucial in boosting shopping efficiency.

Furthermore, the participants expressed that the utilisation of real-time data in development programmes is an advanced technological approach that significantly diminishes the time required for manual data collection, hence facilitating convenient data collection. Furthermore, the chosen metrics for comparing cameras offer significant assistance to consumers in their decision-making process. Many users perceive that camera ratings and the quantity of reviews are particularly helpful in guiding their purchasing choices, even if they possess limited expertise regarding camera performance.

However, users also propose the idea of a project aimed at identifying areas for improvement. For instance, users may encounter challenges while accessing numerous sources of commodities information, as these sources may have certain constraints. Furthermore, the only available visual representation for user reference is the radar chart, which can be considered very repetitive. Users express a desire for an increased availability of data analysis graphs to serve as points of reference.

The research outcomes are limited due to the minimal quantity of data collected in this project. The extended programme did not assess the efficacy of the function in attaining the intended outcomes.

## **Chapter 8**

### **Summary**

#### **8.1 Project management**

At the onset of the project, it is crucial to take into account the user's requirements pertaining to shopping. Factors such as the price of the commodity, the relevant features of the goods, suggestions and comments regarding the goods, and other influential factors will ultimately impact the user's purchasing decision. Given the specified needs, it would be prudent to contemplate the development of this extension. Several challenges were also experienced throughout the duration of the project. In the initial stages, the integration of real-time data retrieval was not executed. However, a supplementary feature was introduced in the form of a button within the pick box, enabling users to manually input data. Subsequently, in light of enhancing user convenience, the decision was made to accomplish this level of performance through the implementation of live data crawling, which would automate the population of the data collection form. Consequently, there is a significant enhancement in the efficiency of user use. Furthermore, it was intended to incorporate parallel coordinate visualisation in the data display page, alongside the existing radar chart. However, due to time constraints during the project's development, this particular requirement could not be implemented. In subsequent research endeavours, it is imperative to progressively enhance the associated prerequisites in order to facilitate additional advancement.

##### **8.1.1 Time Plan**

In order to complete the project within the scheduled time, good time management is the key. For better arrangement of time, at the beginning of the project made a time schedule. According to the schedule, and under the guidance of my tutor Xu Kai, I successfully completed the project.

In order to ensure the smooth progress of the project, the tutor Xu Kai held a group meeting once a week, and the meeting minutes were saved in github. In general, the project went according to plan, but the time limit was limited, and some functional requirements were not implemented in time.



## **8.2 Contribution and thinking**

### **8.2.1 Innovation**

The project's originality is centred on the implementation of automated user data collecting and the utilisation of radar charts. The automated gathering of user data significantly decreases the duration of data collection for users to compare categories, hence enhancing the user's utilisation and purchasing experience to a certain degree. The utilisation of radar charts facilitates the integration of significant data pertaining to various categories. These charts can be assessed against specific criteria, enhancing the value of the analysis outcomes and enabling them to serve as valuable references for consumers. Consequently, they can provide shoppers with guidance in their purchasing decisions, including category recommendations, ultimately assisting users in making more informed choices. This visualisation method partially addresses the challenging issue of analysing multi-dimensional data analogies, and serves as an effective means of presenting data evaluations. Its applicability extends to various domains like business, scientific research, and information technology.

### **8.2.2 Computer ethics and professional issues**

The objective of this study is to enhance consumers' sensory decision-making abilities when evaluating and contrasting several types of products. To a certain degree, this system offers pertinent assistance and recommendations to users in their decision-making processes, potentially influencing consumers' approach to everyday shopping choices. During the study phase of this project, careful consideration was given to not just the technological challenges but also the potential legal and ethical issues that may develop as a result of implementing this extended approach. The project's development adheres to the principles of data protection, non-discrimination, and customer privacy.

From a legal standpoint, this solution aims to guarantee compliance with the data protection rules and regulations of the United Kingdom, specifically in relation to the collection and processing of data. Furthermore, with regards to social influence, this project takes into account the variations among user groups and incorporates interview assessments of diverse user groups to enhance the programme. This includes exploring the possible impact on disadvantaged and marginalised groups. Furthermore, from an ethical standpoint, the study took measures to ensure that users were provided with comprehensive information regarding the practises of data collecting and processing throughout the data collection phase. This approach aimed to safeguard the users' entitlement to be informed about the behaviour pertaining to data collection.

This programme serves as an expansion of an existing initiative that assists users in data analysis and offers pertinent decision-making support. Its target audience encompasses individuals with a broad interest in the subject matter. However, due to the high readability of

the radar chart, even individuals lacking experience in data analysis can derive benefits from the decision-making assistance provided.

### **8.2.3 Reflection and future considerations**

The findings of this study demonstrate that while there exists a current capability to autonomously retrieve user browsing data from the Amazon Camera webpage, there is considerable potential for enhancement.

In subsequent iterations, there is potential for enhancing the project. An illustration of this concept involves expanding the data collecting capabilities of various websites beyond the confines of solely extracting information from a single website through automated crawling processes. In this manner, to a certain degree, consumers are able to compare a greater number of categories and offer decision-making recommendations that align more closely with their expectations.

Additionally, the comparison page can incorporate several data analysis visualisations in addition to radar charts to cater to the diverse comparison requirements of customers. This would enable consumers to autonomously select their preferred data analysis charts for viewing. In this manner, it is more advantageous to allocate distinct consumption decisions to individual consumers.

Overall, although the project successfully fulfilled its intended purpose in terms of enhancing user experience, further research and improvement are still necessary. To effectively mitigate any constraints in this study. Naturally, it is imperative to select a methodology that facilitates the handling of extensive datasets and ensures enhanced security for storage purposes.

## References

- Nguyen, P.H. *et al.* (2016) ‘SenseMap: Supporting browser-based online sensemaking through analytic provenance’, *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)* [Preprint]. doi:10.1109/vast.2016.7883515.
- Kai, P. author By (2023) *Visual analytic for Sensemaking, Human-AI Teaming*. Available at: <https://kaixu.me/2021/01/31/analytic-provenance-for-sensemaking/> (Accessed: 08 September 2023).
- Muskan *et al.* (2022) ‘Data Visualization and its Key Fundamentals: A Comprehensive Survey’, *2022 7th International Conference on Communication and Electronics Systems (ICCES)* [Preprint]. doi:10.1109/icces54183.2022.9835803.
- Mozzafari, E. and Seffah, A. (2008) ‘From visualization to visual mining: Application to Environmental Data’, *First International Conference on Advances in Computer-Human Interaction* [Preprint]. doi:10.1109/achi.2008.29.
- Porter, M.M. and Niksiar, P. (2018) ‘Multidimensional Mechanics: Performance Mapping of Natural Biological Systems using permuted radar charts’, *PLOS ONE*, 13(9). doi:10.1371/journal.pone.0204309.
- Bong, S.L. *et al.* (2023) ‘Analysis of consumer behavior in online shopping on social media’, *2023 International Conference On Cyber Management And Engineering (CyMaEn)* [Preprint]. doi:10.1109/cymaen57228.2023.10051027.
- Tang, J. and Guo, J. (2022) ‘An analysis method for online shopping platform comments based on NLP-AHP: Taking Amazon as an example’, *2022 3rd International Conference on Education, Knowledge and Information Management (ICEKIM)* [Preprint]. doi:10.1109/icekim55072.2022.00234.
- Nguyen, P.H., Xu, K., Wheat, A., et al. (2016) ‘SensePath: Understanding the sensemaking process through analytic provenance’, *IEEE Transactions on Visualization and Computer Graphics*, 22(1), pp. 41–50. doi:10.1109/tvcg.2015.2467611.
- Web Crawler\_Baidu Encyclopedia (no date) Baidu Encyclopedia. Available at: <https://baike.baidu.com/item/%E7%BD%91%E7%BB%9C%E7%88%AC%E8%99%AB/5162711> (Accessed: 08 September 2023).
- Ma, T. and Zhang, Z. (2021) ‘Medical Consultation System based on Python Web Crawler’, *2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT)* [Preprint]. doi:10.1109/cecit53797.2021.00134.

Abodayeh, A. et al. (2023) 'Web scraping for data analytics: A beautifulsoup implementation', 2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU) [Preprint]. doi:10.1109/wids-psu57071.2023.00025.

JSON (2023) Wikipedia. Available at: <https://en.wikipedia.org/wiki/JSON> (Accessed: 08 September 2023).

A brief summary of the advantages of JSON (no date) A brief summary of the advantages of JSON\_Benefits of json\_Dream Chaser's Blog-CSDN Blog. Available at: [https://blog.csdn.net/IT\\_DREAM\\_ER/article/details/60576049](https://blog.csdn.net/IT_DREAM_ER/article/details/60576049) (Accessed: 08 September 2023).

Ojtáš, P. and Peška, L. (2014) 'E-shop user preferences via User Behavior', *Proceedings of the 11th International Conference on e-Business* [Preprint]. doi:10.5220/0005102300680075.

*Echarts\_Baidu Encyclopedia* (no date) *Baidu Encyclopedia*. Available at: <https://baike.baidu.com/item/ECharts/55931618> (Accessed: 08 September 2023).

Vue.js (2023) Wikipedia. Available at: <https://en.wikipedia.org/wiki/Vue.js> (Accessed: 08 September 2023).

Beautiful Soup (HTML parser) (2023) Wikipedia. Available at: [https://en.wikipedia.org/wiki/Beautiful\\_Soup\\_\(HTML\\_parser\)](https://en.wikipedia.org/wiki/Beautiful_Soup_(HTML_parser)) (Accessed: 08 September 2023).

*message queue* (2022) *Wikipedia*. Available at: <https://zh.wikipedia.org/zh-hans/%E6%B6%88%E6%81%AF%E9%98%9F%E5%88%97> (Accessed: 08 September 2023).

*HTML* (2023) *Wikipedia*. Available at: <https://zh.wikipedia.org/wiki/HTML> (Accessed: 08 September 2023).

*CSS* (2023) *Wikipedia*. Available at: <https://zh.wikipedia.org/zh-hans/CSS> (Accessed: 08 September 2023).

*JavaScript* (2023) *Wikipedia*. Available at: <https://zh.wikipedia.org/zh-hans/JavaScript> (Accessed: 08 September 2023).

