

# UNDERGRADUATE PROJECT PROGRESS REPORT

Project Title:	Design of a web-based price comparison system for an online shopping platform
Surname:	Zhu
First Name:	Xunran
Student Number:	201918010408
Supervisor Name:	Albert Xu
Module Code:	CHC 6096
Module Name:	Project
Date Submitted:	2023/01/10

## Table of Contents

1 Introduction .....	2
1.1 Background .....	2
1.2 Aim .....	2
1.3 Objectives .....	2
1.4 Project Overview .....	3
1.4.1 Scope .....	3
1.4.2 Audience .....	3
2 Background Review .....	3
3 Project Technical Progress .....	5
3.1 Methodology .....	5
3.1.1 Approach .....	5
3.1.2 Technology .....	6
3.2 Testing and Evaluation .....	7
3.3 Design and Implementation .....	8
4 Project Management .....	11
4.1 Activities .....	11
4.2 Schedule .....	13
4.3 Project Version Management .....	13
4.4 Project Data Management .....	13
4.5 Project Deliverables .....	13
5 Professional Issues and Risk: .....	14
5.1 Risk Analysis .....	14
5.2 Professional Issues .....	14
6 References .....	14

## 1 Introduction

### 1.1 Background

In the 21st century, with the gradual rise of Internet technology, online shopping has become more and more popular [1]. In 2020, the sales volume of China's Tmall Shopping Festival reached 498.2 billion yuan [2]. Nowadays, due to different business models, these online shopping platforms often have different prices for the same goods at the same time. For online shoppers, when they make online shopping, they want to buy what they want at the lowest price. This requires online shoppers to compare prices of goods on multiple platforms, so more than one web interface needs to be opened. Obviously, although this method can achieve the goal, the operation is very troublesome and time-consuming. If a commodity price comparison system can automatically collect commodity price information from multiple shopping websites and display the price comparison results, it will provide great convenience for users. The purpose of this paper is to study how to realize this price comparison system.

### 1.2 Aim

In this paper, the data extraction method based on Web and web crawler technology are used to extract information from multiple online shopping platforms through the search function of the search bar of the home page of the system, and then the extracted information is integrated. Finally, the integrated results are displayed on the interface of the system, so that online shoppers can compare prices of similar products on each online shopping platform.

### 1.3 Objectives

(1) Consult and collect related literature of the subject, and conduct background investigation of the subject. On this basis, the project is studied and researched, and the project conceptual model is proposed.

(2) Design price comparison system, and use crawler technology for data extraction and data collection.

(3) Design UI, that is, design the web user interface, so that users can operate the hardware conveniently and efficiently, and realize two-way interaction.

(4) Database design, including requirements analysis, conceptual structure design, logical design (ER modeling), physical design, database successful implementation and database maintenance.

- (5) Design and write back-end code to realize the normal operation and storage of data.
- (6) Test all functions.
- (7) Interpret and deal with product instructions, hazard instructions and possible safety effects.
- (8) Realize the interface showing the price comparison results of commodities on different platforms to users.

## 1.4 Project Overview

### 1.4.1 Scope

The price comparison system of the online shopping platform is based on the web, which aims to directly compare the price of the same commodity on different platforms for users, without the need to open multiple web interfaces, which greatly provides convenience for users. The price comparison system first requires users to register an account. After logging in, a home page will be displayed. The home page can display some different categories of items, such as mobile phones, computers and so on. Not only that, it should also include a search box, which is used for commodity price comparison, when the user enters the search box to want to compare the same commodity, the interface will display the corresponding commodity information and price comparison results. In addition, the results of commodity price comparison can be arranged in ascending or descending order, making it easier for users to filter.

### 1.4.2 Audience

The project's main audience is those who like to shop online and compare prices. With the development of Internet technology, more and more people are more willing to buy the things they love and need through online shopping. In the process of buying, people will compare the prices of the desired items, and this Web-based online shopping platform price comparison system provides great convenience for this kind of people.

## 2 Background Review

Web crawler technology is a program or script that automatically obtains information according to rules. Using this technology can automatically obtain any information within the scope of authority, which lays the foundation for search engines to further process the data, and finally users can get the information they want. Liu [15] mentioned in his

article that web crawler technology, as a program that automatically obtains network information, can download data from the Internet for search engines and transmit a large amount of network data to the local, realizing the function of obtaining data information through data collection, data processing and data storage. During crawling, the crawler program will continue to obtain a new url from the webpage, and then convert the link address to the corresponding IP address of the network server through the DNS resolution of the URL. Then put the URL into the crawl queue and record the corresponding URL until the information retrieved by the crawler meets the stop condition. After that, the captured data is processed, indexed, and stored in the appropriate folder or database. Finally, according to the query requirements, the data is extracted from the database and displayed according to the user's requirements.

Sharma and Gupta[3] report that Web crawler traverses the Web by following the ever-changing, dense and distributed hyperlink structure. In Shi, Shi and Lin's article [4], text-based crawling and web link-based crawling are traditional crawling strategies. Text content-based methods mainly analyze text and URL in web pages, represented by fish search strategy, shark search strategy [10] and optimal priority strategy [11]. Vadivel et al. [5] explained in their article that Web crawler obtains Web pages and other multimedia files from the Web, preprocesses the pages, and extracts information [12].

HTML5 technology is widely used in the development of Web service platform. As mentioned in Zhao's article [6], analyzing interactions is the first step in the design process. By collecting users' ideas, a good foundation is laid for the design of the system. The second step is to identify the view interface. The third step is to list the elements. Four is grouping. The fifth is to create a paper prototype. The sixth is the lo-fi prototype. Seventh, feasibility analysis. In the article Yi,Feng and Li[7], the basic front-end technique is described. These are HTML, CSS, and JavaScript. HTML is its Structured Standard Language. CSS as the standard language. JavaScript is the standard language for behavior.

As can be seen from Wu's article [8], database system is a system composed of database system and its corresponding database management software. DBS is an ideal data processing core developed to meet the needs of data processing. For most users, the only way the user interacts with the database is through the database management system. Attaran and Far[9] pointed out the design methods of different types of databases. Generally, three schema concepts should be followed: conceptual

schema, logically structured data, and physical database design [13]. Edgar. F. Codd, inventor of the relational model, proposed the concept of normalization in the 1970s, including the first, second and third normal forms, namely the currently known 1NF, 2NF and 3NF[14].

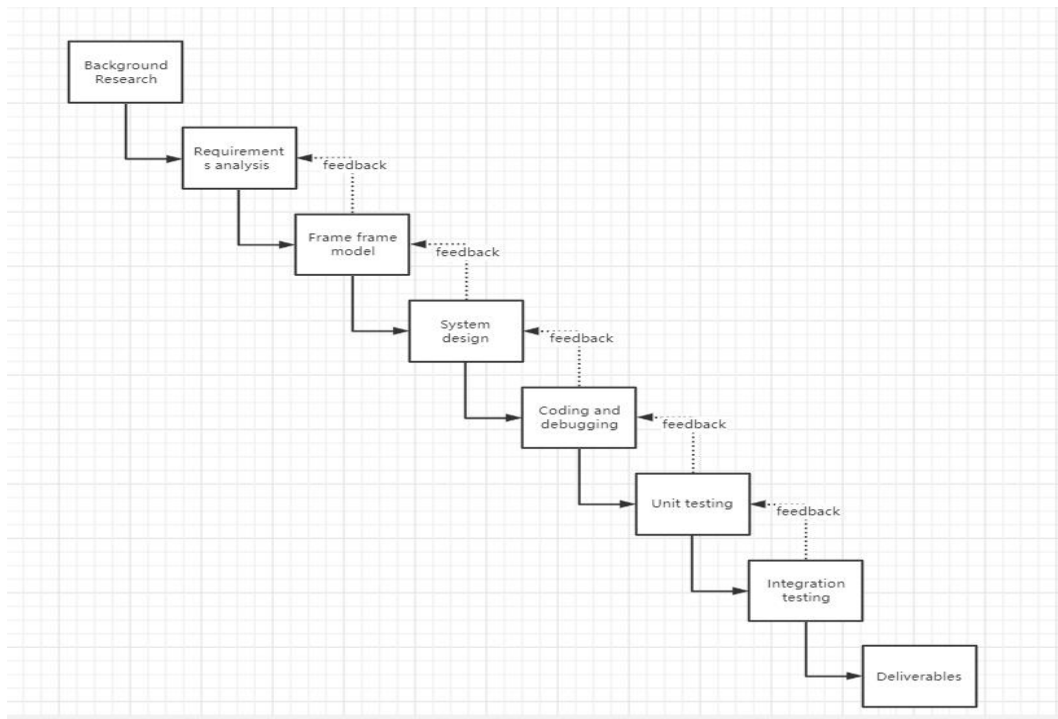
### 3 Project Technical Progress

#### 3.1 Methodology

##### 3.1.1 Approach

This part is mainly divided into three parts, development model diagram, requirements analysis and technical framework.

##### 3.1.1.1 Development model diagram

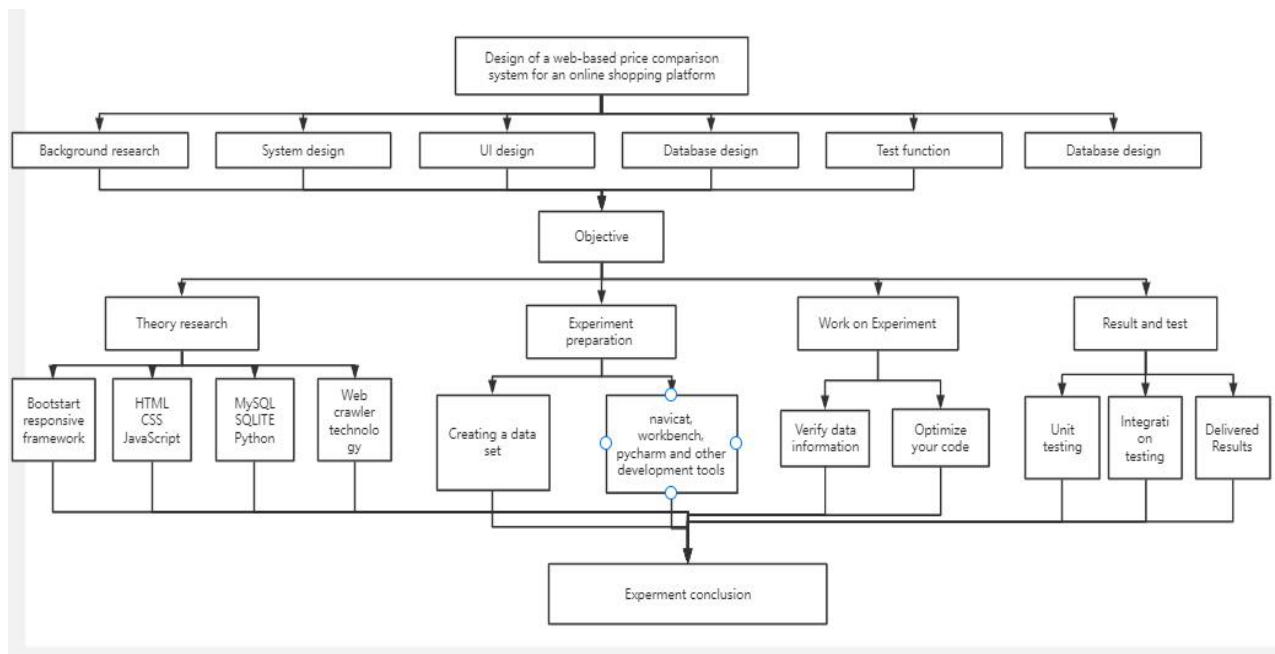


(Figure 1: Development model diagram)

### 3.1.1.2 Requirements analysis

The price comparison system must realize user registration and login, unregistered users should register first, and log in after registration. After login, different categories of commodities will be displayed on the main interface, and there should be a search bar on the main interface, through which data can be obtained and extracted data information can be displayed on the user interface after processing and integration. For example: "commodity price", "commodity name" and so on. To achieve this goal, you need to use crawler technology, build user interfaces and build databases, as well as back-end code implementations. The back-end development software uses pycharm, the framework uses django, and the database uses mysql or sqlite. The user interface can use the bootstrap response framework.

### 3.1.1.3 Technical framework



(Figure 2: Technical circuit diagram)

### 3.1.2 Technology

Software: PyCharm, Navicat

Programming language: Python

Browser: Microsoft Edge

Database: MySQL5.7

Frame: Flask

### 3.2 Testing and Evaluation

According to the design, there are generally two test methods, respectively black box test and white box test. Black box testing is also known as functional testing. This testing method does not consider the logic structure and internal characteristics of the program at all. It only needs to check whether the function of the program conforms to its function description. Black box test is a test carried out on the program interface. Specific black box test case design methods include equivalence class division method, boundary value analysis method, false speculation method, cause-and-effect graph method, decision table driven method, orthogonal experimental design method, etc. Firstly, according to the specification of functional requirements and performance indicators in the user requirements report, the corresponding test requirements report is defined, that is, the highest standard of black box testing is formulated. All the test work in the future will be carried out around the test requirements, and the application that meets the test requirements is qualified, otherwise it is unqualified. At the same time, we should choose the test content properly, arrange the test personnel, test time and test resources reasonably. For the black box test evaluation, the quality, work progress and work efficiency of the application software should be comprehensively evaluated by combining the quantified test coverage area and defect tracking report.

White box testing is also called structural testing. Is a test case design approach that understands what's inside a software system and how it works. The "white box" method has a comprehensive understanding of the internal logical structure of the program and tests all logical paths. By checking the logic structure inside the software, the coverage test of the logical path in the software, and set up checkpoints in different places of the program, check the state of the program, to determine whether the actual running state is consistent with the expected state. The test methods of white box test include code inspection method, static structure analysis method, static quality measurement method, logic coverage method, basic path test method, domain test and so on. The most widely used test method in white box testing is the basic path testing method, which is based on the program control flow diagram, through the analysis of the loop complexity of



control construction, derived the basic executable path set, so as to design the test case method. The general steps are: (1) Draw the control flow diagram, (2) calculate the cyclomatic complexity, and (3) derive the test case.

### 3.3 Design and Implementation

The design idea of the price comparison system is to use Python as its compilation language. The front-end development technology is mainly HTML, CSS and JavaScript framework. JavaScript script language has excellent features such as simplicity, dynamism and cross-platform, so it is more appropriate to choose this framework. The back end uses the Flask framework, which is different from Django. Flask is easy to learn, easy to build prototypes and flexible. As a lightweight framework, the system is more applicable. The database uses MySQL, which has the advantages of realizing data sharing, reducing data redundancy, data independence, data consistency and maintainability.

Through the above design ideas, the system is actually designed. As a commodity price comparison system, first of all, there needs to be a login registration interface. Users without accounts should enter the registration interface in the login interface to register their accounts. After registration, the system will automatically return to the login interface to log in; After logging in, you can return to the main interface. In the main interface, you can see multiple categories of goods and pictures of some goods. Not only that, above the commodity category selection bar, There are two option buttons, "Home" and "Price Comparison". After entering the interface, the system defaults to the home page. If the user wants to compare the price of goods at this time, he should click the "Price Comparison" button to enter the search bar, enter the goods he wants to compare, and click the "Search" button to display the price comparison results on the home page. In addition, in the upper right corner of the entire interface, there is a "logout" button. When the user completes the price comparison, the logout operation is not a good choice.

According to the above design ideas, the preliminary results are as follows:

```

import pymysql
import config
from pymysql.cursors import DictCursor

# 链接数据库
def dbLink():
    db = pymysql.connect(
        host=config.HOST,
        user=config.USERNAME,
        password=config.PASSWORLD,
        database=config.DATABASE,
        port=3306
    )
    cursor = db.cursor(DictCursor)
    return cursor,db

# 直接sql语句
def executionSql(sql,isSubmit=False):
    cursor,db = dbLink()
    if isSubmit:
        cursor.execute(sql)
        db.commit()
    else:
        cursor.execute(sql)
    res = cursor.fetchall()
    cursor.close()
    db.close()
    return res

```

(Figure 3: Code 1)

```

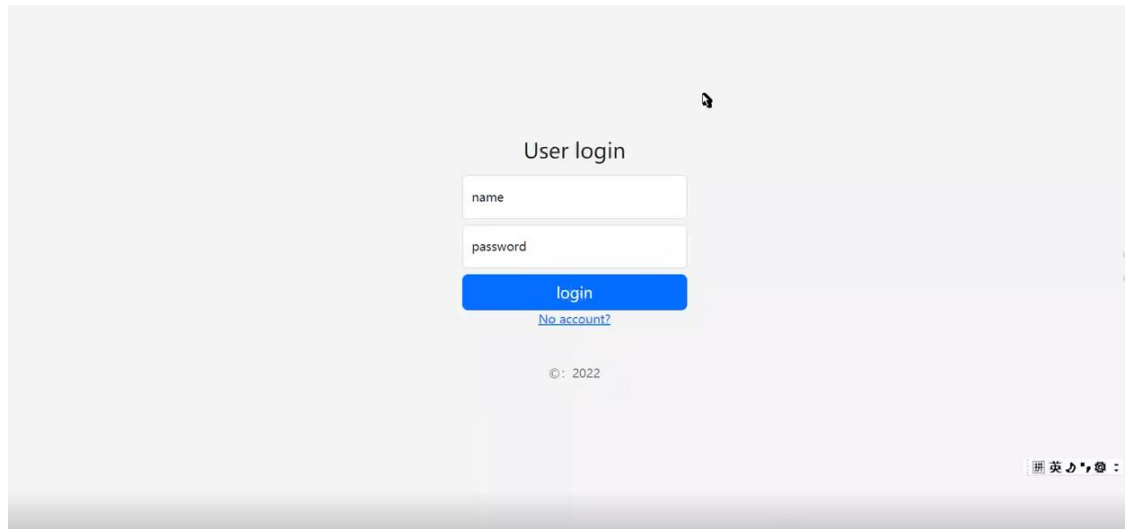
from app import app
from flask import render_template,request,redirect,url_for,session
from database import queryTypeData,queryTypeRandomData,inquire_user,register,queryData
import json

@app.route('/',methods=['GET','POST'])
def index():
    data = queryTypeRandomData()
    if request.method == 'POST':
        types = request.form.getlist('type')
        data = queryTypeData(types)
    return render_template('index.html',data=data)

@app.route('/parity',methods=['GET','POST'])
def parity():
    data = []
    if request.method == 'POST':
        name = request.form.get('name')

```

(Figure 4: Code 2)



A screenshot of a web application's login page. The page has a light gray background. In the center, there is a white box containing the title "User login". Below the title are two input fields: the first is labeled "name" and the second is labeled "password". Below these fields is a blue button with the text "login" in white. Under the button is a blue hyperlink that says "No account?". At the bottom of the white box, there is a small copyright notice "©: 2022". In the bottom right corner of the page, there is a small icon and the text "英办" followed by a colon.

User login

name

password

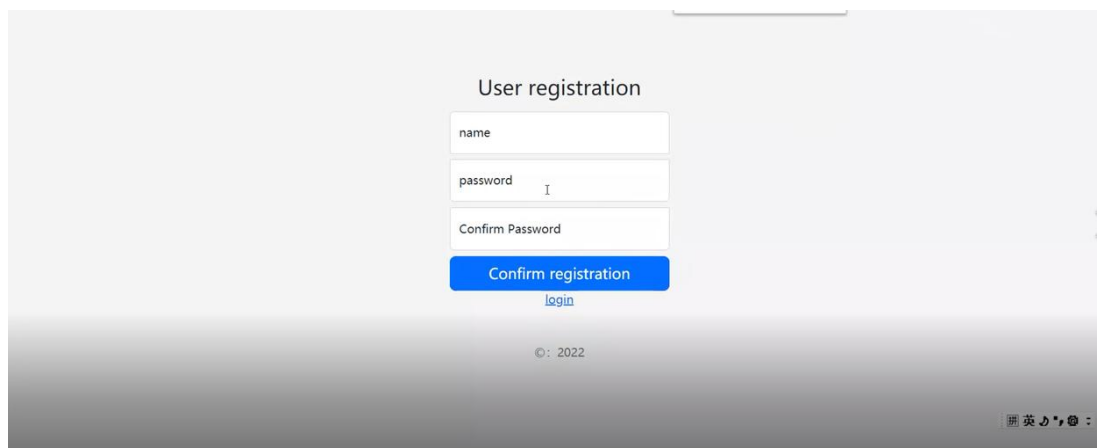
login

[No account?](#)

©: 2022

英办 :

(Figure 5: Login)



A screenshot of a web application's registration page. The page has a light gray background. In the center, there is a white box containing the title "User registration". Below the title are three input fields: the first is labeled "name", the second is labeled "password" and has a cursor, and the third is labeled "Confirm Password". Below these fields is a blue button with the text "Confirm registration" in white. Under the button is a blue hyperlink that says "login". At the bottom of the white box, there is a small copyright notice "©: 2022". In the bottom right corner of the page, there is a small icon and the text "英办" followed by a colon.

User registration

name

password

Confirm Password

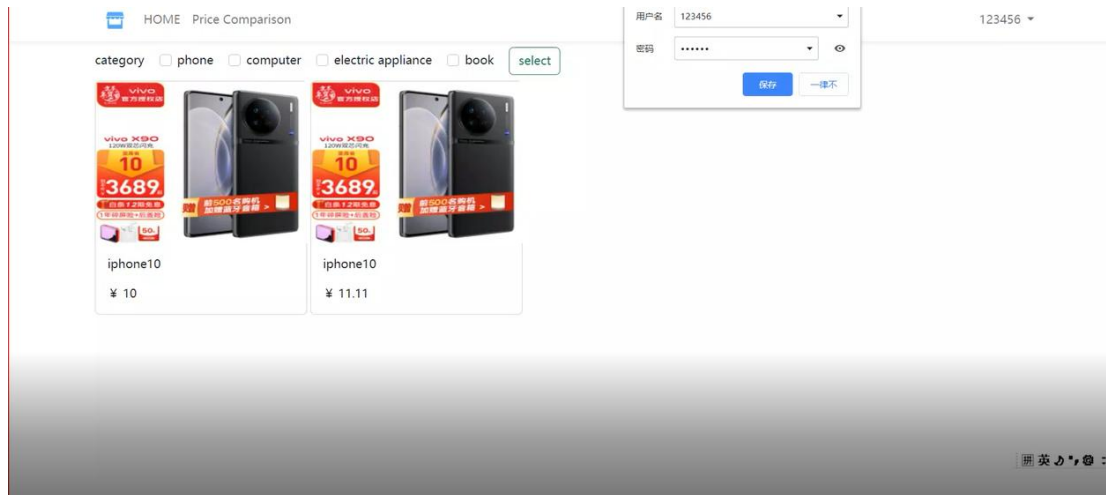
Confirm registration

[login](#)

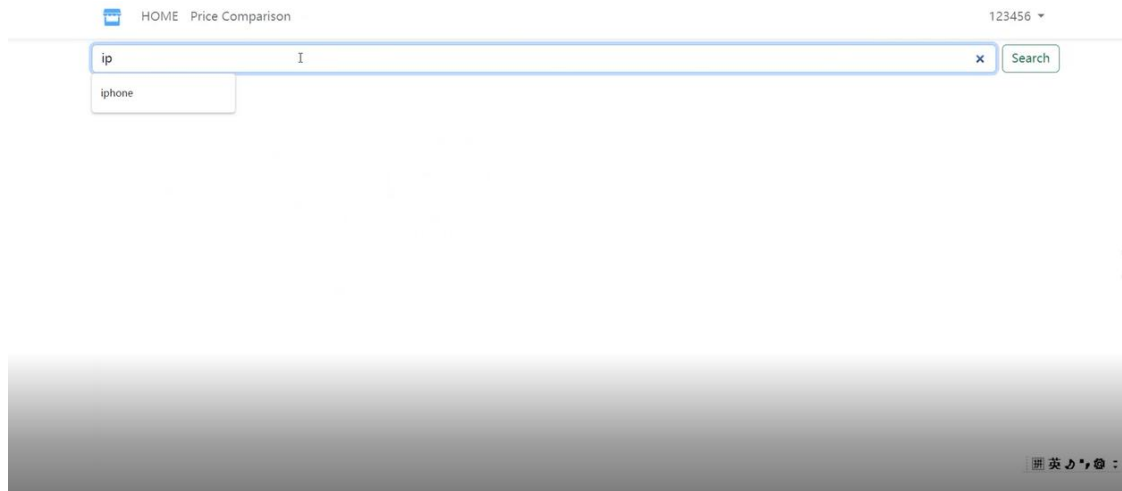
©: 2022

英办 :

(Figure 6: Registration)



(Figure 7: Home Page)



(Figure 8: Price Comparison)

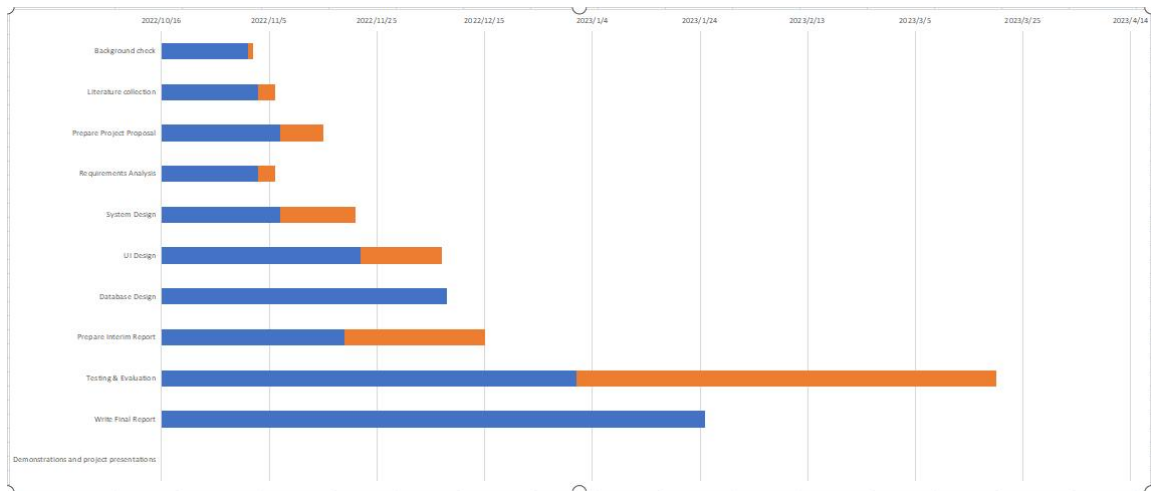
## 4 Project Management

### 4.1 Activities

Objective 1: Study existing system models	Ac1: Background check
	Ac2: Collection of literature
Objective 2: Design of price comparison system	Ac1: Understand web crawler technology
	Ac2: Using web crawler technology to develop the system

Objective 3: User interface design	Ac1: Design the user interface, mainly including login and registration homepage, as well as each interface contains pictures and interactive buttons
	Ac2: Code
Objective 4: Design of database	Ac1: Create a data set according to the model, collect and store the data
	Ac2: Simple processing and analysis of data
	Ac3: Will be presented to the user through the user interface
Objective 5: Test all function	Ac1: Unit test each part of the system randomly for several times and observe the results
	Ac2: Test the whole system randomly for several times and observe the results
Objective 6: Show the comparison results	Ac: Successfully display the price comparison results of the same commodity on different platforms

## 4.2 Schedule



(Figure 9: Gantt Chart)

## 4.3 Project Version Management

The project will use GitHub as a version management tool to create a separate folder called Project for adding and storing code and files, and categorizing each part of it.

## 4.4 Project Data Management

Data related to the project will be stored on GitHub in a folder called Project and updated as the developer progresses until the project is introduced.

## 4.5 Project Deliverables

### 4.5.1 Project Proposal

### 4.5.2 Project Weekly Report

### 4.5.3 Project Progress report

### 4.5.4 Project Final Report

### 4.5.5 Project Code

### 4.5.6 Project Presentation

### 4.5.7 Price comparison system of online shopping based on web

## 5 Professional Issues and Risk:

### 5.1 Risk Analysis

The risk of the project lies in the protection of user data and access to information. The system uses the web crawler technology to crawl to each shopping website for information acquisition. It will obtain the price of the goods in the website, store information, location, etc. If it has not negotiated with the shopping website, it will obtain the encrypted or inaccessible information of the website, which will touch the legal issues, which should be avoided or effectively solved. In addition, for users, frequent search for products will leave data. If user data is leaked, some "people with intentions" will steal more information according to the user's preferences, which is also a problem to be solved.

### 5.2 Professional Issues

The social problem of this project is mainly the protection and processing of user data. The leaked data may be used to make "interested people" obtain more information. In terms of ethics, the project refers to the professional code of conduct, which meets the requirements and will not cause any legal, social, ethical and environmental problems.

## 6 References

- [1] Han'guk Chōnja T'ongsin Yoñ'guwoñ, Han'guk Chōngbo Sahoe Chinhūngwoñ, Global IT Research Institute (Korea), & IEEE Communications Society. (n.d.). *The 11th International Conference on Advanced Communication Technology : ubiquitous ICT convergence makes life better! : ICACT 2009 : Phoenix Park, Korea, Feb. 15-18, 2009 : proceedings.*
  
- [2] Guo, Q., Yu, W., Ma, Z., & Kong, J. (2021). User Behavior Monitoring Mechanism of Online Shopping Processes based on Hierarchical Colored Petri Net. *Proceeding - 2021 China Automation Congress, CAC 2021*, 6118–6122.<https://doi.org/10.1109/CAC53003.2021.9728033>
  
- [3] Sharma, S., & Gupta, P. (2015). The anatomy of web crawlers. *International Conference on Computing, Communication and Automation, ICCCA 2015*, 849–853. <https://doi.org/10.1109/CCAA.2015.7148493>

[4] Shi, Z., Shi, M., & Lin, W. (2016). *The Implementation of Crawling News Page Based on Incremental Web Crawler; The Implementation of Crawling News Page Based on Incremental Web Crawler*. <https://doi.org/10.1109/ACIT-CSII-BCD.2016.74>

[5] Ieee. (2012). *2012 International Conference on Advances in Engineering, Science and Management*. IEEE.

[6] Haitao, Z. (2021). Research on Mobile Web Front-end Design Based on HTML5 Technology. *Proceedings - 2021 6th International Symposium on Computer and Information Processing Technology, ISCIPT 2021*, 290–293. <https://doi.org/10.1109/ISCIPT53667.2021.00065>

[7] Yi, N., Feng, X., & Li, C. (n.d.). *2017 7 th International Conference on Communication Systems and Network Technologies The Design and Implementation of the Front End of the Art Play Library System*. <https://doi.org/10.1109/CSNT.2017.54>

[8] Wu, Y. (2017). Design of user database resource management system based on web. *Proceedings - 2017 International Conference on Computer Technology, Electronics and Communication, ICCTEC 2017*, 364–367. <https://doi.org/10.1109/ICCTEC.2017.00084>

[9] Attaran, H., & Far, A. H. (2011). A novel technique for object oriented relational database design. *Proceedings of 2011, 10th IEEE International Conference on Cybernetic Intelligent Systems, CIS 2011*, 128–132. <https://doi.org/10.1109/CIS.2011.6169147>

[10] M. Hersovici, M. Jacovi, Y. S. Maarek, et al, “The shark-search algorithm. An application: tailored Web site mapping,” *Computer Networks and ISDN Systems*, 1998, pp. 317-326.

[11] F. Menczer, G. Pant, P. Srinivasan, “Topical web crawlers: Evaluating adaptive algorithms,” *ACM Transactions on Internet Technology (TOTIT)*, 2004, pp. 378-419.



- [12] V. Shkapenyuk , T. Suel, "Design And implementation of a High Performance Distributed WebCrawler", fn fCD£. 2002.
- [13] R. Kaula, and J. Chin "A database approach towards flexible manufacturing: A conceptual framework, " in Computers and Industrial Engineering, Vol 24, No 2, pp. 131-141, Apr 1993
- [14] P. K. Janert, "Practical Database Design, Part 2; Normalization, History, and Event Logging, " From <http://www.devx.com/ibm/Article/20859/4047>, Accessed 2011
- [15]Liu, R. (2021). Network Crawler Technology Based on Python Under Information. In: Chang, JW., Yen, N., Hung, J.C. (eds) Frontier Computing. FC 2020. Lecture Notes in Electrical Engineering, vol 747. Springer, Singapore. [https://doi.org/10.1007/978-981-16-0115-6\\_227](https://doi.org/10.1007/978-981-16-0115-6_227)