

**COMP4801 – Final Year Project**  
**Web based tool for Chinese character evolution**  
**Intermediate Report**

**Siu Lik Hang, Leo (3035125201)**

**Wong Tsz Hing, Chris (3035068908)**

## **Abstract**

Chinese characters has its long history, for a single character or word, it contains multiple meaning and pronunciation. Existing online information of Chinese characters are scattered. When a user wants to search for a specific character or word, they may need to browse several websites to gather the information he is looking for. This project aims to create a website that contains as much as Chinese characters information by opening to users to allow contributing their own knowledge to the website under a flexible database system. This website will contain a character dictionary and a word dictionary. Users can freely input new data to enhance the dictionary. They can also search a character or word by the search function inside this website according to different criteria. This application was developed using html, javascript and php supported by a MySQL database .

Currently, the basic structure of the system has been established. As new ideas were come up, the project scope and schedule are modified. These new functions will be implemented in the second semester.

## **Acknowledgements**

First, I would like to thank my advisor Dr Vincent Lau for guiding and giving insightful feedback to me, which made this project can stick to the right track. I would also like to thank my family for giving me support and advice.

## Table of Contents

<b>ABSTRACT</b>	<b>2</b>
<b>ACKNOWLEDGEMENTS</b>	<b>2</b>
<b>1 INTRODUCTION</b>	<b>5</b>
<b>1.1 Background</b>	<b>5</b>
<b>1.2 Problems</b>	<b>5</b>
<b>1.3 Objectives</b>	<b>6</b>
<b>1.4 Related works</b>	<b>7</b>
<b>2 PROJECT METHODOLOGY</b>	<b>8</b>
<b>2.1 Feature</b>	<b>8</b>
<b>2.1.1 Character directory</b>	<b>8</b>
<b>2.1.2 Word directory</b>	<b>8</b>
<b>2.1.3 Search function</b>	<b>8</b>
<b>2.1.4 Games</b>	<b>9</b>
<b>2.1.5 User contribute function</b>	<b>9</b>
<b>2.1.6 Revision history</b>	<b>9</b>
<b>2.2 System Architecture</b>	<b>10</b>
<b>2.3 Programming Language</b>	<b>10</b>
<b>2.4 Website structure</b>	<b>11</b>
<b>3 CURRENT PROGRESS</b>	<b>11</b>
<b>4 PROBLEM ENCOUNTERED</b>	<b>12</b>
<b>5 FUTURE WORK</b>	<b>12</b>
<b>5.1 Handle multimedia</b>	<b>12</b>
<b>5.2 Audit function</b>	<b>13</b>
<b>5.3 Linking function</b>	<b>13</b>
<b>5.4 Listing view</b>	<b>13</b>
<b>5.5 Smart upload</b>	<b>13</b>
<b>5.6 User management</b>	<b>14</b>
<b>5.7 Security review</b>	<b>14</b>
<b>6 CONCLUSION</b>	<b>14</b>
<b>LIST OF FIGURE</b>	<b>15</b>
<b>REFERENCES</b>	<b>15</b>
<b>APPENDIX</b>	<b>16</b>

<b>Appendix 1: Project Schedule and Milestones</b>	<b>16</b>
<b>Appendix 2: Online resources about Chinese word and characters</b>	<b>17</b>

# **1. INTRODUCTION**

## **1.1 BACKGROUND**

As an era of information, public can receive information they want from internet easily. Even for Chinese dictionary, it has shifted from paper to electronic version. There are websites and online dictionaries that contain information about Chinese characters (Refer to appendix 1). Each website has its own characteristics, for example, some focus on the pronunciation of the character, while some focus on meanings of a single character and some focus on the stroke order of a character.

## **1.2 PROBLEMS**

Firstly, information of Chinese characters are scattered over the internet. If a user want to search for more information of a specific Chinese character, he need to browse through several websites to gather all information he is looking for. Secondly, these online resources may not be updated frequently. As a result, these resources may not accurately reflect the evolution of Chinese characters. Last but not least, new words and phrases increase with times, the existing websites cannot catch up with these changes as they were not designed to change any content inside.

### **1.3 OBJECTIVES**

This project aims to create a web tool that contains as much information as possible about Chinese characters which could be updated frequently to show the latest status of the characters, word and phrases. To achieve this, the website will be opened to users to contribute their knowledge to the website.

The goal of this project is to create a website that has the following features:

1. Serving as a convenient tool for searching Chinese characters information
2. Providing a platform for users that can contribute on updating the latest Chinese characters information.
3. Flexible database structure is used to allow enhancement on database. User can input new type of data related to Chinese characters. For example, there are only meanings and pronunciation about Chinese characters at the website originally. User is allowed to add a new column 'big5 code' if they wish to store the big5 code in the system.
4. Source code is opened to the public to allow improvement on our website and creating other similar database system for public interest.

## 1.4 RELATED WORKS

### *Wikipedia*

Wikipedia is a free online encyclopedia with the aim to allow anyone to edit articles [1]. It was launched on 2001 with only an English version originally but now is extended to support more than 299 different languages. With more than 40 million articles and 500 million visitors monthly, exchange of knowledge and culture can be achieved.

However, the open source practice could be possibly a drawback, visitors of Wikipedia cannot identify which part of article is truth, half-truth or even falsehoods.

### *Chinese Character Database: With Word-formations Phonologically Disambiguated According to the Cantonese Dialect*

This database is developed by Chinese University which aims to provide a useful online tool to assist the teaching and learning of the Chinese characters. [2] It includes as much Chinese characters as possible, with their script (how it writes), pronunciations and meaning.

However, this database is not friendly for new learner of Chinese characters as the searching tools require user having adequate knowledge on Chinese characters.

The remainder of this report proceeds as follows. First, the methodology part includes the functionality of the website, the system architecture, choice of program language and the website structure. Next, current progress and problem encountered will be discussed. Finally, future works and a conclusion will be included at the end of this report, summarize the project interim status.

## 2. PROJECT METHODOLOGY

In this section, it introduces the features of this project and a flexible database system will be used to implement these features. Next, system architecture and choice of program language will be mentioned deeply in later part. Finally, the website structure of this project will be included at the end of this chapter.

## **2.1 Features**

This project will act as an open source web-based tool for visitors. The major feature list is as below:

### **2.1.1 Character directory**

Different information of a Chinese character, such as pronunciation, radical, component, meaning and stroke, can be checked. Also, user can check different forms of the same character (such as the archaic form and the simplified form).

### **2.1.2 Word directory**

The meaning of a word or a phrase and the usage can be checked. Also, example sentences are given to demonstrate the usage of the word or the phrase. For some words or phrases, the stories behind are included to explain how they are created.

### **2.1.3 Search function**

Characters or words can be searched by giving different conditions, such as pronunciation, radical, component, meaning, number of stroke and other user input criteria.

### **2.1.4 Games (Low priority)**



Based on the information in the database, games are provided for users to test their understanding on Chinese characters. For example, there will be a game that requires users to answer the meanings of a word. Moreover, game content will be updated automatically when any information at character directory or word directory has been modified.

However, when comparing with other functions, game has a lower importance as it is not the main focus in this project. Therefore, this function was considered as a low priority function. It will be implemented only if all other functions are all finished and having sufficient time.

#### **2.1.5 User contribute function**

This project provides a platform for users to update the information on the website. The system can handle input of csv files, videos, images, sounds and hyperlinks. Moreover, the system can handle batch input.

Flexible database structure will be used in constructing the system. User is allowed to add new criteria of a character. After adding a new criteria, the dictionary layout and the search function will display the new changes correctly.

#### **2.1.6 Revision history**

Since user contribution is allowed, an audit trail will be implemented to check the changes that have been made. Web administrator can rollback the changes if necessary.

## 2.2 System Architecture



Figure 1- System flow

The website consists of 3 components – database, server and web interface (see figure 1).

### *Flexible database structure*

Handling database storage is the core of this project. Mysql database is used in this project. Since users are allowed to upload any type of data ranging from image files (.png, .jpg), audio files (.mp3, .wmv), text files(.txt, .docx) or hyperlink to another webpage, to achieve flexibility in storing the data, we provide upload functions that can change the table structure. Differing from ordinary implementation of SQL database, user are free to create new tables and new columns to store data. Hence, the table can be theoretically suitable for storing any types of data and relationship.

Server contains search and fetch of data. At this stage, it will run on hkucs intranet to facilitate searching and fetching of data when it receives request from user.

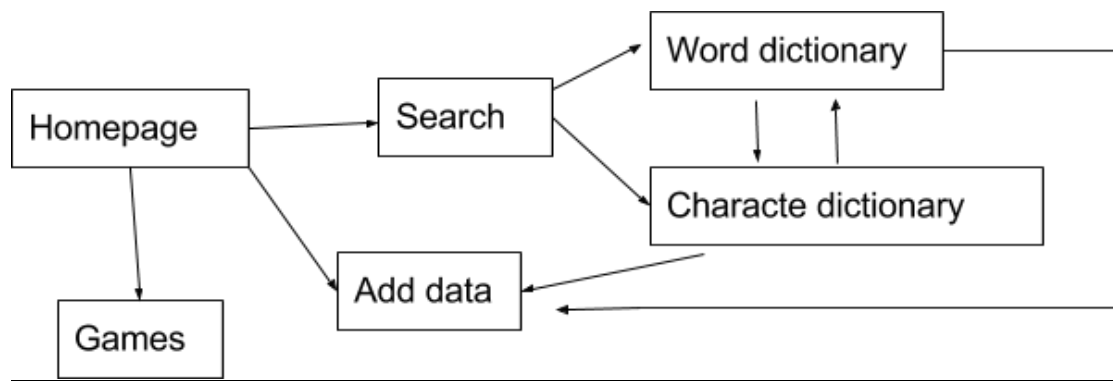
For any devices with internet and browser supported, they can access the web interface with a correct URL.

## 2.3 Programming Language

Html, javascript and php will be used to build the website in this project.

The html and javascript will be used to define the basic layout of a page, serving as a template. When data of a specific character is read, the php script will be called to retrieve data from the database and place the data in the corresponding position.

## 2.4 Website structure



**Figure 2- Basic Flow of the website**

Homepage is accessed when the user firstly visit the website. It contains basic information of the system and links to different functions including game page, the add data page and the search page.

Specific conditions of a character or word can be input on the search page. The system will guide the user to the correct page of the character or word. At the dictionary page, user can access to other dictionary pages that is related to that page.

User can enter the add data page either from homepage or from the dictionary pages. When user accesses add data function from homepage, he can choose to create a new record or edit an existing record. However, when user accesses the add data function from the dictionary pages, only editing existing record is allowed.

### **3. Current Progress**

#### *Design and set up of database*

This is the major part of this project as it heavily depends on the design and setting of database in order to accept any data type uploaded by user. SQL database was chosen (Mysql) as it is supported by most programming language. A database has been set up to support flexible update. Users can create and alter a table and the changes can be shown in the system correctly. To effective manage this structure, more functions and testing will be conducted on next phrase.

#### *Webpage template*

First draft of webpage was created. Showing information, search and input function has been implemented.

#### **4. PROBLEM ENCOUNTERED**

The major challenge of this project is to adopt the database in a flexible structure. A lot of time is spent on designing the table. As the implementation is different from ordinary database structure, some of the traditional management method may not be easy to use. For example, we are still not sure how to do indexing as there may be new tables and columns created. Also, we are not sure how the searching time be lengthen if user keep adding new tables and columns.

More testing will be done to explore the skills we can use to handle this flexible database structure.

#### **5. FUTURE WORK**

Currently, the basic function of this website has been established, like inputting data, searching and listing all information of a character. The coming tasks will be implementing of more advanced function.

As new ideas came up in the first semester, the future plan differs from the original plan in order to fulfill these changes. The updated project schedule is showed at the Appendix. New features are list below:

##### **5.1 Handle multimedia**

The website will be enhanced to handle multimedia data, such as image, audio and video. Users can upload different types of file into the website. Moreover, the system will also behave differently according to the input data type. For example, when the searching field is a number, numerical operator(e.g. greater than) will be provided to the users as a search criteria.

##### **5.2 Audit function**

As a database application, this system will log all the changes the user has made. An audit function will be implemented to record all the sql queries run in the system and allow users to roll back changes when necessary.

### **5.3 Linking function**

The system will be able to create linkage between related characters and words. For example, users can access to another page via a hyperlink to view another related page like synonym of that character or word. This function can create better user experience and reflect the relationship between data in the database.

### **5.4 Listing view**

For easy viewing the data, we will provide another method for displaying the data. Currently, the data are shown only when users click on a character. After implementing this new function, users can click on an attribute and the system will create a list of all characters and the corresponding attribute. It helps user to do auditing and comparison between massive amount of characters in a single page.

### **5.5 Smart upload**

A batch upload function will be created for users to input a large amount of data at once. Users can input those data by uploading a CSV file. The system will identify whether it should create new records or update existing records.

### **5.6 User management**

Users are not allow to edit the data unless they have login the system. Therefore, a user authentication function will be created to serve for user to login. Moreover, user

statistic will be recorded, such as the pages that has modified by users. The system will also ban a user if he or she is reported destructive repeatedly. Finally, a channel will be created and provided for user to communicate with administrators or one another.

### **5.7 Security review**

In the testing part, besides functional testing, we will also perform some security testing to eliminate security risk in our system. One of the criteria we will like to test is whether an attacker can perform SQL injection to attack the system.

## **6. CONCLUSION**

The basic part of this project such as set up of flexible database and build a website template has been achieved. During the project, as new ideas came up to support the flexible data structure, the project schedule has been modified.

This project is a new attempt to change the current practice on inserting data into database with a given schema. If this project can achieve it successfully, the final objective: source code is opened to the public to allow improvement on our website and creating other similar database system for public interest, can be achieved. Public can use our project as a reference when they try to build an expandable database.

## LIST OF FIGURE

Figure 1- System flow

Figure 2- Basic Flow of the website

## REFERENCES

[1] Retrieved October 23, 2017, from <https://en.wikipedia.org/wiki/Wikipedia>

[2] 使用凡例. Retrieved October 23, 2017, from <http://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/guide.php>

## APPENDIX

### Appendix 1: Project Schedule and Milestones

Timeline	Task
September 2017	<ol style="list-style-type: none"><li>1. Detailed project plan</li><li>2. Project webpage</li></ol>
October to December 2017	<ol style="list-style-type: none"><li>1. Design and set up database</li><li>2. Website template</li></ol>
December to January 2018	<ol style="list-style-type: none"><li>1. Search function</li><li>2. Interim report</li><li>3. User management</li></ol>
February	<ol style="list-style-type: none"><li>1. Handle multimedia</li><li>2. Smart upload</li></ol>
March 2018	<ol style="list-style-type: none"><li>1. Audit function</li><li>2. Linking function</li><li>3. Listing view</li><li>4. Testing</li><li>5. Security review</li></ol>
April 2018	<ol style="list-style-type: none"><li>1. Final report</li><li>2. Final presentation</li></ol>
May 2018	<ol style="list-style-type: none"><li>1. Project exhibition</li></ol>

### Appendix 2: Online resources about Chinese word and characters



Provide meanings of words and pronunciation in mandarin :

<http://www.zhongwen.com/>

Provide games for user to learn the special features of Chinese character:

<http://www.dragonwise.hku.hk/dragon2/>

Provide articles to introduce special features of Chinese character:

<http://chineseffect.com/lessons/characters/>

Provide information of Chinese character which primary students have to learn:

[http://www.edbchinese.hk/lexlist\\_ch/](http://www.edbchinese.hk/lexlist_ch/)

Provide pronunciation of Chinese Characters:

<http://humanum.arts.cuhk.edu.hk/Lexis/Canton/>

Provide information about stroke-order:

[http://stroke-order.learningweb.moe.edu.tw/mobiles/word\\_query\\_form.tiles](http://stroke-order.learningweb.moe.edu.tw/mobiles/word_query_form.tiles)