

COMP4801 – Final Year Project 2017-2018
Department of Computer Science
The University of Hong Kong



WEB BASED TOOL FOR CHINESE CHARACTER EVOLUTION

Group report

Supervisor
Dr. Vincent Lau

Student

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 .

The source code of the project will be available to the public for further improvement. We hoped that others can use our source code as a framework to develop their own web application with the minimal time and effort.

Acknowledgements

First, we would like to thank my advisor Dr .Vincent Lau and second examiner Dr. Beta Yip 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

ABSTRACT	2
ACKNOWLEDGEMENTS	2
1. INTRODUCTION	5
1.1. Backgrounds	5
1.2 Problem	5
1.3 Objectives	5
1.4 Related Works	7
2. PROJECT SCOPE	10
2.1 FEATURES	10
2.1.1 Character Directory	10
2.1.2 Word Directory	10
2.1.3 Search Function	11
2.1.4 List View	11
2.1.5 User Contribution Function	11
2.1.6 Revision History	11
2.1.7 Internal linkage	11
3. BASIC FUNCTION	12
3.1 Main	12
3.2 Log	13
3.3 Create Account	13
3.4 Logged In	14
3.5 Character and Word Dictionary	15
3.6 List View	16
3.7 Edit Data	17

3.8 Batch Upload	18
3.9 Recover deletion of record	19
3.10 Internal linkage	20
4 PROJECT METHODOLOGY	22
4.1 Equipment and technology	22
4.2 Flexible database structure	23
4.2.1 Base tables	23
4.2.2 Attribute tables	24
4.2.3 Static tables	26
4.3 Testing and result	33
5. DIVISION OF LABOUR	34
6. AMENDMENT	35
6.1 Change of batch upload file format	35
6.2 Removal of game feature	35
7. PROBLEMS ENCOUNTERED	36
7.1 Implementation of flexible structure	36
7.2 Limitation of time	36
8. FUTURE IMPROVEMENT AND DISCUSSION	37
8.1 Use of framework in interface design	37
8.2 Support more file format for batch upload	37
8.3 Insert or update data when using batch upload	37
8.4 Serve as a database for generic usage	38
9. CONCLUSION	39
LIST OF FIGURE	40
LIST OF TABLE	41

REFERENCES

42

APPENDIX

43

Appendix 1: Project Schedule and Milestones

43

Appendix 2: Online resources about Chinese word and characters

44

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
As a web based tool, convenient and usability will be the main concerns of this project. Our website provides a simple interface which all the major functions are listed on the first page of the website such that users do not need to spend extra time to understand the structure of the website. Searching or editing data can be finished within few mouse clicks.
2. Providing a platform for users that can contribute on updating the latest Chinese characters information.
For websites that are related to Chinese characters, they were designed not to make any significant changes for the content inside. Therefore, new words and phrases are not expected could be found in these website. Instead of heavily rely on administer solely updating or adding new data to the website, this project provides a platform for users to contribute their knowledge on Chinese character once they have logged into their account in this website.

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. Users are allowed to add a new column 'big5 code' if they wish to store the big5 code in the system. Moreover, multimedia is also supported in this website. Users can upload image, audio or even video materials which are related to Chinese character in order to further enhance this website.
4. Source code is opened to the public to allow improvement on our website and creating other similar database system for public interest.
To make the system more complete in all aspects, source code will be published to GitHub and available for general public. General public can modify the source code to serve their own purpose, for example, create an encyclopedia website. This can be achieved by modifying only part of the backend source code, suit for people do not have much experience on coding.

1.4 RELATED WORKS



WIKIPEDIA
The Free Encyclopedia
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.

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

粵語審音配詞字庫

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.

粵典

words.hk

Words.hk is a website aim to create a Cantonese dictionary by crowd-sourcing, each user can resister as an editor to contribute their knowledge on Cantonese to their database.




Website	 WIKIPEDIA The Free Encyclopedia	 粵語審音配詞字庫		Dynamic Chinese Database (Our project)
Reflect real time update?	✓	✗	✗	✓
Batch upload?	✗	✗	✗	✓
Open source?	✓	✗	✓	✓
Search tool easy to use?	✓	✗	✓	✓

Table 1-Compare of different website

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 SCOPE

In this section, it introduces the features of this project and a flexible database system was used to implement these features.

2.1 Features

This project will act as an open source web based tool for visitors. The major features listed as follow:

2.1.1 Character directory

- Information of a Chinese character, such as pronunciation, radical, component, meaning and stroke, can be checked.

- Different forms of the same character (such as the archaic form and the simplified form).

- New attribute of character can be added or modified by any registered users.

2.1.2 Word directory

- Meaning of a word or a phrase and the usage can be checked.

- 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.

- Same as character directory, registered users are allowed to add or modify any attribute to word directory.

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 and fast retrieval of the searching result.

2.1.4 List view

- List all characters/words according to attribute(s) selected by user.

2.1.5 User contribute function

- Provide a platform for users to update the information on the website. The system can handle input of txt files, videos, images, sounds and hyperlinks.

- Handle batch input with given format.

- Flexible database structure will be used in constructing the system. Users are allowed to add new attribute to a character. After adding a new attribute, the dictionary layout and the search function will display the new changes correctly.

2.1.6 Revision history

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

2.1.7 Internal linkage

- The pages of words and character can be links together automatically if they are related. The linkage can also be defined manually by user.

3. BASIC FUNCTION

In this section, major functionality of this project will be discussed and supported with the actual picture.

3.1 Main

When a user visits our website, a main page will be shown.

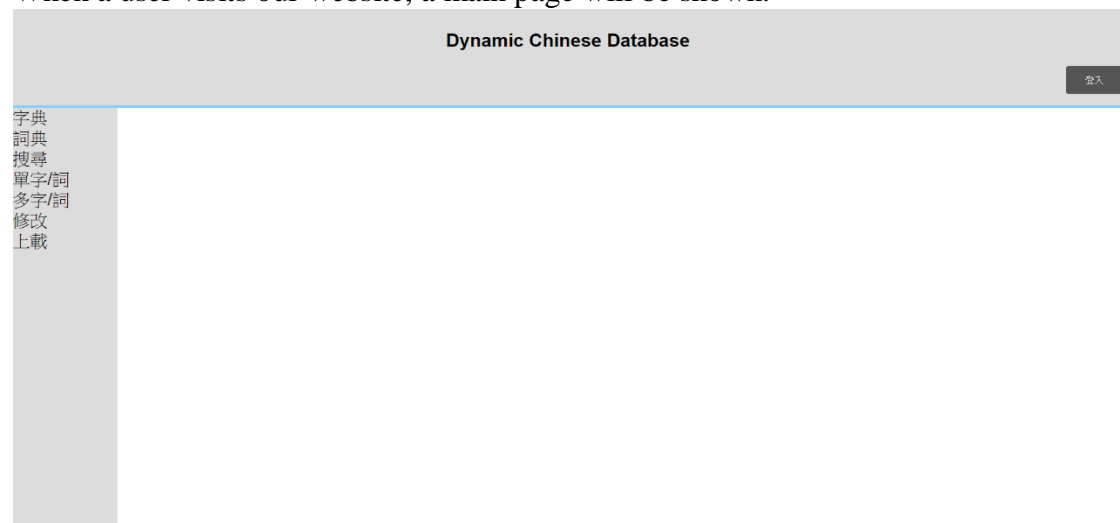


Figure 1-Main page

3.2 Log

All major function, namely “Character dictionary”, “Word dictionary”, “Search character/word”, “List view”, “Edit” and “Upload” are listed on the left-hand side. To use the function of “Edit” and “Upload”, the user must first login the website.

Dynamic Chinese Database Login

用戶名稱:

密碼:

登入

[建立新帳號](#)

Figure 2-Login

3.3 Create account

A user can also create an account by providing their “User name”, “Password” and “Email”

Dynamic Chinese Database Create Account

用戶名稱:

密碼:

重新輸入密碼:

電子郵件:

提交

Figure 3- Create account

3.4 Logged in

Once the user is logged in, “User name” and two buttons “Log out” and “Change password” will be shown on right hand side of the page. For administer one more

button “Message” will be shown, which any user can send message to administer for page issue, account issue or other question.



Figure 4 -After logged in (Normal User)

Figure 5 -After logged in (Administer)

3.5 Character and Word dictionary

If the user clicks “Character dictionary” or “Word dictionary”, 10 random characters or words will be shown.



Figure 6-Character dictionary

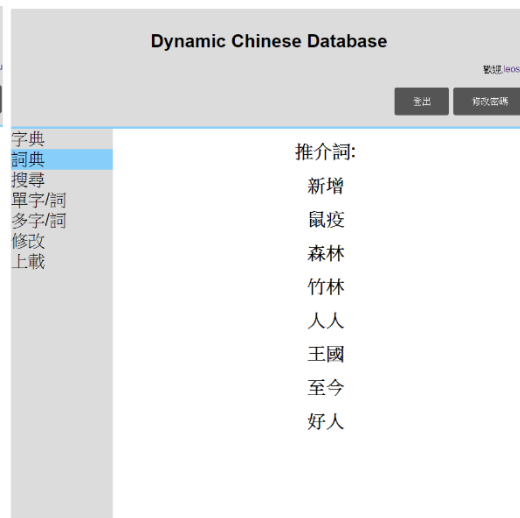


Figure 7-Word dictionary

For the search function, “Search character/word” can search for characters /words which matches with the search criteria.

Figure 8-Search character/word

搜尋結果

倡
唱
娼
田
晶
喝
曝
狷

result

Figure 9-Search result

3.6 List view

The user can also use the “List view” to view multiple attribute of character/word. For attributes that do not contain any content inside, it will be shown in red color of “暫無”.

搜尋結果

字	意思	例句	圖案	超連結
人人	人	暫無	暫無	暫無
好人	Good person/ Good people	暫無	暫無	暫無
好人	Good person/ Good people	他是個好人	暫無	暫無
好人	very good people	www	暫無	http://hk.yahoo
好人	very good people	www	uploads/admin190Sign.jpg	http://hk.yahoo.com
新增	www	qqq	uploads/Lighthouse.jpg	http://www.google.com
新增	ddd	fff	uploads/Koala.jpg	http://ggg
森林	asd	暫無	暫無	暫無
王國	暫無	暫無	暫無	暫無
竹				

Figure 10-Choose attribute that will be listed

Figure 11-List View

3.7 Edit data

To edit data, user can click “Edit” and the current relation of words and characters will be shown. User can perform adding new column, attribute or word/character to the database. For administer, he/she also has the right to rename a column or attribute or even delete column or attribute.

字典
詞典
搜尋
單字/詞
多字/詞
修改
上載

加入新資料

中文字 中文詞語

中文字意思

- 意思 [修改此項目](#)
- 讀音 [修改此項目](#)
- 別字 [修改此項目](#)
- [在此加入新項目](#)

倉頡輸入法

- 倉頡碼 [修改此項目](#)
- [在此加入新項目](#)

筆劃

- 筆劃數 [修改此項目](#)
- [在此加入新項目](#)

中文電碼

- 電碼 [修改此項目](#)
- [在此加入新項目](#)

別字

- [在此加入新項目](#)
- [加入新欄目](#)
- [輸入資料](#)

登出 修改密碼

字典
詞典
搜尋
單字/詞
多字/詞
修改
上載

加入新資料

中文字 中文詞語

中文字意思 [修改此欄目](#)

- 意思 [修改此項目](#)
- 讀音 [修改此項目](#)
- 別字 [修改此項目](#)
- [在此加入新項目](#)

倉頡輸入法 [修改此欄目](#)

- 倉頡碼 [修改此項目](#)
- [在此加入新項目](#)

筆劃 [修改此欄目](#)

- 筆劃數 [修改此項目](#)
- [在此加入新項目](#)

中文電碼 [修改此欄目](#)

- 電碼 [修改此項目](#)
- [在此加入新項目](#)

別字 [修改此欄目](#)

- [在此加入新項目](#)
- [加入新欄目](#)
- [輸入資料](#)

信息 登出 修改密碼

Figure 12-Normal user (logged in)
Figure 13-Administer

修改項目意思

重新命名

重新命名為

移除項目

你確定要作此操作, 移除項目意思?

[返回](#)

Figure 14-Edit column/attribute

3.8 Batch upload

For batch upload function, it supports Unicode file ends with .txt and each line must contain a special character “end”. Then the user has to choose the corresponding meaning of that column, then press upload to upload the data to database.

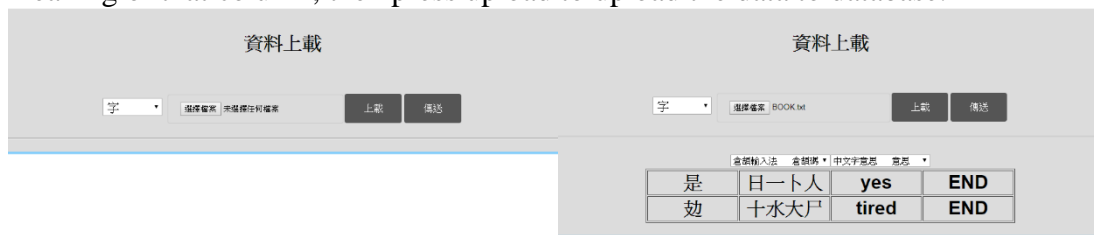


Figure 15-Upload page

Figure 16-File

uploaded to page

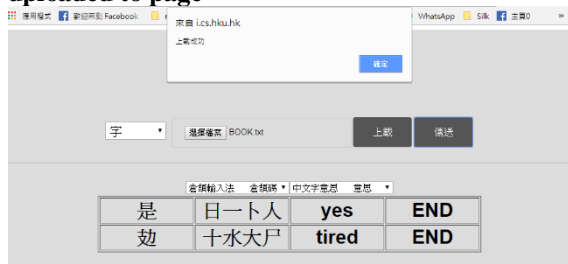


Figure 17-Data uploaded to database

3.9 Recover deletion of record

For all the action that will invoke any changes to database (insert, update or delete) will be recorded and it is recoverable.

是

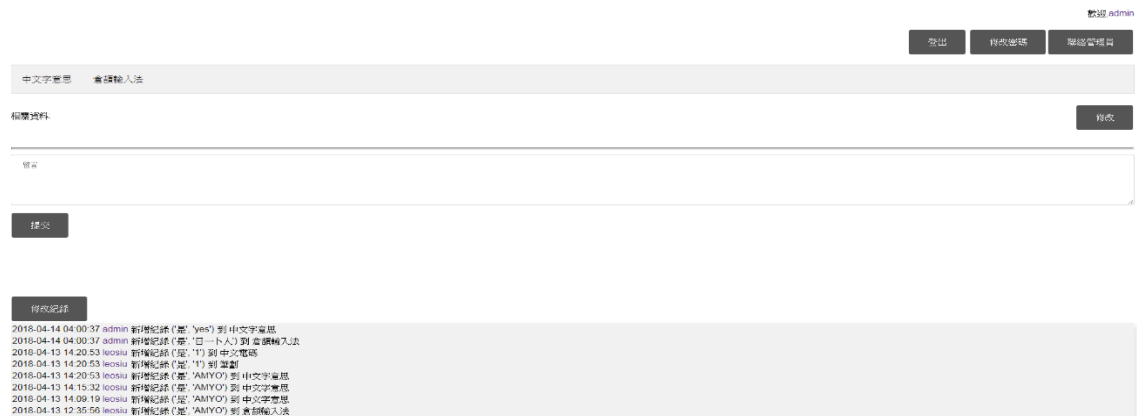


Figure 18-Character “是” with edit record

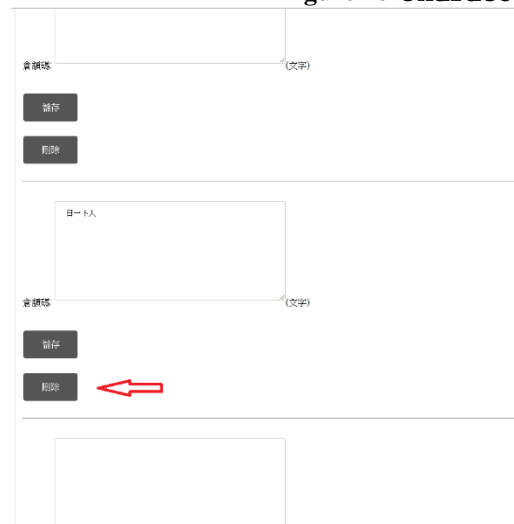


Figure 19-Delete record

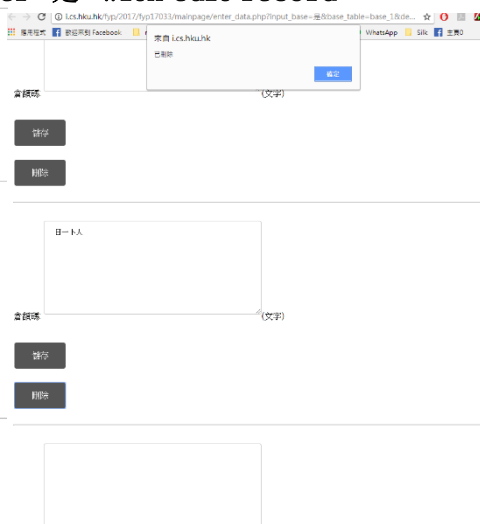


Figure 20-Delete record (continue)



Figure 21-Record deleted and edit record updated

Figure 22-Try to recover record

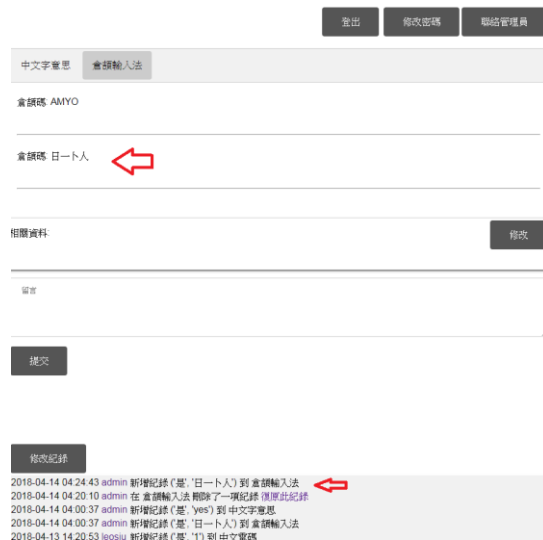


Figure 23-Record successfully recover

3.10 Internal linkage

Internal link refers to the link that point to other pages of this website. With proper internal links, user can easily browse other related page, which creates better user experience.

Auto linkage

In the pages, there are auto linkage to links to other pages.

好

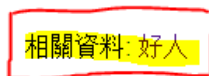
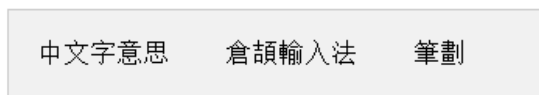


Figure 24: Auto internal linkage

Under the dictionary tabs, the system will search for other words or characters that is related to the current page and provide a hyperlink to access that page. For Chinese words, the system will search for the characters that formed the words. For a Chinese character, system will search for words that is formed by it.

Manual linkage

When user need to create an internal link manually, he need to first type a space, then use 2 @ signs to enclose the character or words that need to be linked.

For example, when we input '123 @abc@', than abc will become a hyperlink to link to page 'abc'. You may find figure 3 to see how this function is used.

他

中文字意思	倉頡輸入法	中文電碼
意思: 讀音: Ta1 例句: 他是個好人		
意思: 讀音: Ta1 例句: 他是個 好人		

Figure 25: Compare normal record and records that contains internal links

User inputs ‘他是個好人’ in the upper record, so the record does not have the links. The lower record contains internal links because user inputs ‘他是個 @好人@’. So the lower record will create a link to the page ‘好人’.

4 PROJECT METHODOLOGY

In this section, system architecture and choice of program language will be included. Then, core of this project, flexible database structure will be introduced and mentioned deeply in this chapter. Finally, testing and result will be discussed and

analyze of the performance of the system.



Figure 26- System flow

The website consists of 3 components – database, server and web interface

4.1 Equipment and technology

HTML, JavaScript, MySQL and PHP were used to build the website in this project.

i. HTML and JavaScript

HTML and JavaScript were used to set up the website of this project. Given that our team has sufficient knowledge on HTML and JavaScript and they are freely to use in nature, HTML and JavaScript were chosen. By combing JavaScript into HTML, a simple and responsive website can be achieved.

ii. MySQL

MySQL is an open source relational database management system that is used for storing data related to Chinese characters or words, or other data related to user management.

Considering that this project will be opened sources to public, others can easy to implement our source code is the main concern of this project .Compared to other open source relational database management system like SQLite and PostgreSQL, MySQL is the most popular one of all the large-scale database servers. It features a rich, open source product that powers a lot of websites and applications online. Also given the fact that getting started with MySQL is relatively easy and developers have access to a massive array of information regarding the database on the internet, so MySQL was used in this project. [3]

iii. PHP

PHP is a widely-used open source general-purpose scripting language that is especially suited for web development and can be embedded into HTML. With the help of PHP, dynamic and interactive application can be achieved and running in satisfactory time without degrading user experience. [4]

iv. Encoding

Since our database need to handle Chinese input, the encoding is important. Otherwise, the input and output will be affected. For both database fields, php sql setting and html character set, we use unicode (utf8) for encoding.

4.2 Flexible database structure

Since we aim to collect different types of data from the public, we need to structure all these data well in order to allow user to search for these data easily in the future.

To achieve this purpose, we adopted a flexible database structure. In the database point of view, users can create new tables and new column for some new types of data instead of only inserting new records.

In our database structure, tables can be classified in 3 types.

4.2.1 Base tables

Base tables are used to define what records are available in the database. If a character or word does not appear in the base tables, the webpage will not create a page for that character or words.

In the system, we now have 2 base tables, called base_1 and base_2. The former is used to handle Chinese character and the latter is for Chinese word.

Table name: base_1

Column name	Data type	Description
Base	varchar(10)	Store the character that have a page in the website.

Table name: base_2

Column name	Data type	Description
Base	varchar(100)	Store the word that have a page in the website.

Table 2-Structure of base table

4.2.2 Attribute tables

Attribute tables are used to store the data related to the bases. They are the part that contribute to the flexibility of the system. When user add new column, create new tables or insert new record, the changes are reflected in the attribute tables.

In fact, for one attribute table, it is represented as a tab in the dictionary pages. Each column in that attribute table is the data type. Therefore, when user adding new tabs or data type, they are creating new attribute tables or adding new columns.

This table show the structure of an attribute table.

Table name: attribute_1

Column name	Data type	Description	
Unique_key	int(11)	The unique identifier of a record. It is the primary key and will auto increase by 1 upon insert.	
Base	vchar(100)	The base that this record belongs to.	
Column_1	vchar(1000)	Store text data for this record.	*
Number_1	double	Store number data for this record.	*
Image_1	vchar(100)	Used for storing multimedia data (such as image, audio, video) for this record. This field records the file name of that data so the webpage can display the data correctly.	*
File_1	vchar(100)	Used for storing file for this record. This field records the file name so the webpage can create a link to the file correctly.	*
Link_1	vchar(100)	Used for storing the hyperlink to external webpage.	*

Table 3-Structure of attribute tables

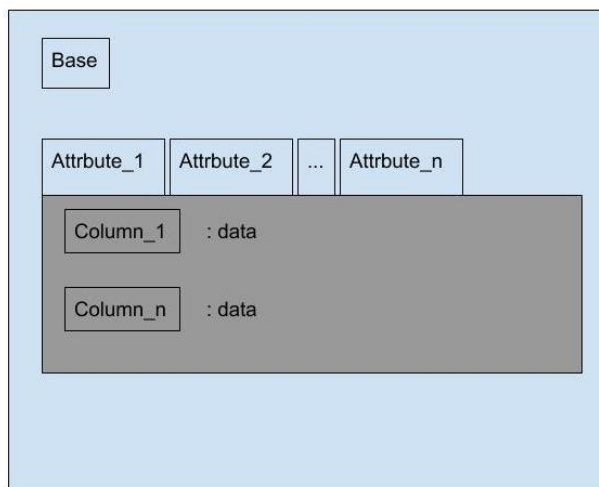
For the starred field, they may or may not be included in that table.

When user request to add a column, he need to choose a format from all the starred column. The naming will be automatically done by the system. For example, the first image column created in the table will be called Image_1, the second will be called Image_2.

The case for table name is similar. The table name is auto generated by the system upon creation. The first attribute table created is called attribute_1 , and the second attribute table is named attribute_2.

Base table and attribute tables can form the pages of this relations.

Dictionary pages:



Edit structure page:

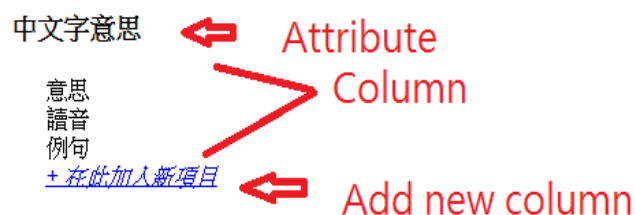


Figure 27: Relationship between a page and the tables

4.2.3 Static tables

These table served the purpose of operating the webpage. The structure of these tables cannot be changed by the user.

Map table

This table are an important table for displaying the Chinese information correctly in the dictionary function. It maps the attribute tables with suitable description. It will translate the table name and column name into user's language. As a result, user can know the actual meanings of each column in the attribute tables instead of meaningless Column_1 and Column_2.

Table name: map

Column name	Data type	Description
Table_name	varchar(50)	The target attribute table this record describe.
Column_name	varchar(50)	The target column this record describ.
Relation	varchar(40)	The base table this record is relate to.

Describ	varchar(40)	The description of that table / column. This description is displayed in the webpage.
Type	varchar(1)	The type of the record. There are 3 possible type. B - This record is used for describing the title of a base table. C - This record is used for describing the title of a column in the attribute table. T - This record is used for describing the title of the attribute table.

Table 4-Structure of map table

Audit log

This table record the changes that user has been made to the system, including modifying attribute tables' structure and editing data. This table store the sql statement that has been run to implement the changes.

This table also contribute to the recover function. The system find the corresponding insert statement for a deleted record to insert back the record in the database.

When necessary, administrator can study the logs for restoring the database. In the most extreme case, this table can serve as a backup of the database. We can build a new database and run all the sql statement in sequence to rebuild the system.

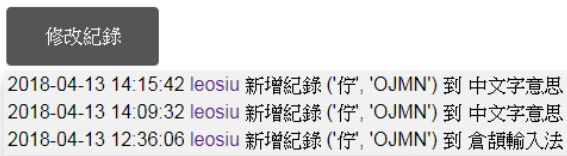


Figure 28: Audit log shown for each

page

In the website, only “Lst up time”, “User id” and “SQL run” will be shown on the corresponding page of that base.

Table name: audit_log

Column name	Data type	Description
Unique_key	int(11)	The unique identifier of a record. It is the primary key and will auto increase by 1 upon insert.
Table_name	varchar(50)	The table that has been modified.
Column_name	varchar(50)	The column that has been modified.
Base	varchar(100)	The base (character / words) this record is relate to.

Action	varchar(10)	The action has been done. There are several type: Add base - Add new record in the base table. Add column - Add new column in the attribute table. Add map - Insert new record in map table. Insert - Insert new record in the attribute table. Delete - Delete record in the attribute table. New table - Create new attribute table. Del column - Delete a column in attribute tables. Del map - Delete a record in map table. Del table - Delete an attribute table. Re column - Rename a column in an attribute table (by updating the corresponding record in the map table). Re table - Rename an attribute table (by updating the corresponding record in the map table).
SQL_run	varchar(200)	The sql statement has been run for this action.
Lst_upd_date	timestamp	The date that this record is created.
IP	varchar(100)	The ip address of the machine that user used to do the changes.
User_id	varchar(50)	The user_id that made the changes.
Remark	varchar(100)	For insert and delete, remark store the record's corresponding Unique_key in that attribute table. It is used for recover the deleted record. For other action, remark store the Chinese description of that action just for convenient.

Table 5-Structure of map table

The base tables, attribute tables, map table and audit log formed our flexible database system to let user manage the data and alter the structure of the attribute tables.

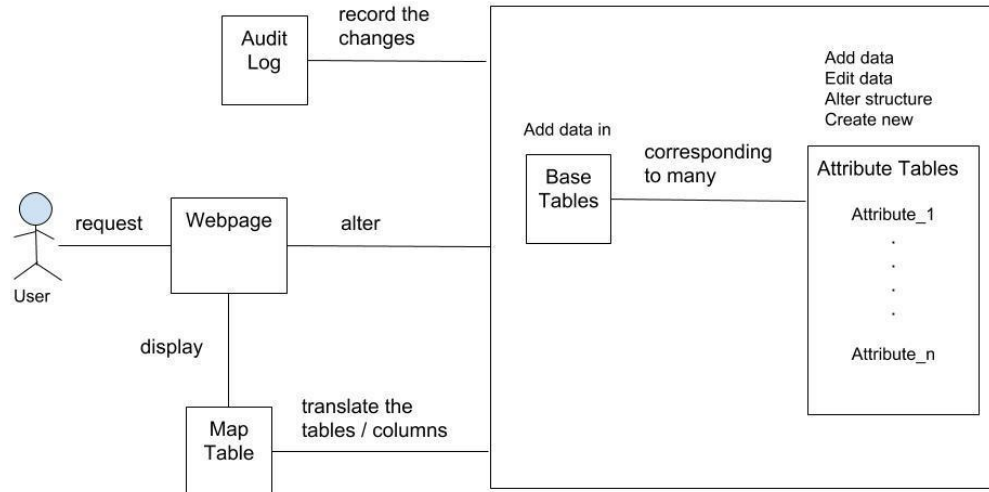


Figure 29 Overview to the whole table structure

Other static tables

There are also other static tables in our system. Although they are not contributing to the flexible structure of the system, they are important for providing operation related functions in our system.

Log in related tables

To enable user management functions, we need to have tables related to user account. User_info table stores all information about a user.

Table name: user_info

Column name	Data type	Description
User_id	varchar(50)	The unique identifier of a user. User can login using this username.
Email	varchar(100)	The email address of the user.
Status	varchar(1)	The status of the user. A - This account is active. L - This account is locked.
Date_create	timestamp	The date of creating this user account
Passwd	varchar(100)	The password of the user. When user login, system will compare the input password with this field to

		check if the password is corrected. For security purpose, the value of this field is hashed. No one can get the actual password of the user using this field.
Role	varchar(10)	The role of the user. Currently, there are 2 roles: User - normal user Admin - Administrator According to user's role, the system has different behaviour. For example, some function may only available to administrator.
Lst_upd_date	timestamp	The last update date of this record. When user change his/her password, this field will be updated.

access_Log table record the date and time a user try to login the system.

Table name: access_Log

Column name	Data type	Description
User_id	varchar(50)	The unique identifier of a user.
Login_result	varchar(20)	The result of login. Success - The user login successfully. Fail - The user failed to login the system.(Typing wrong username or password).
Access_date	timestamp	The date of this login attempt.
Access_ip	varchar(100)	The ip address of this user who try to login.

Table 6-Structure of login related table

Communication related tables

Both these tables , admin_log and comment stored the message for users. The former saves all the message sent to administrator, and the latter saves the message in the discussion area under the dictionary page.

Upon loading the dictionary page, the system will retrieve the message related to that page.

Table name: admin_log

Column name	Data type	Description
id	int(11)	The unique identifier of the message.

Issue	varchar(100)	The type of the message.
Content	varchar(1000)	The message user has input.
Poster	varchar(100)	The user who post this message.
Time	varchar(100)	The time when the user send the message.

Table name: comment

Column name	Data type	Description
id	int(11)	The unique identifier of the message.
page	varchar(100)	The page that this message is related to
post_name	varchar(100)	The user who post this message.
time	date	The time when the user send the message.
content	varchar(1000)	The message user has input.

Table 7-Structure of communication related table

4.3 Testing and result

Performance

To become a useful web based tool, performance would be the main concern. Testing was based on time needed to load a page when fetching large amount of data or searching through all data from database.

	Test 1	Test 2
Data amount	5283 entries	5283 entries
Upload time	~20 seconds	~20 seconds
List all entries with list view (5475 entries)	~16 seconds	~4 seconds
Indexing the base field?	NO	YES
Search for words/characters	~20 seconds	~2 seconds
Load one dictionary page	~10 seconds	<1 seconds

Table 8-Testing result after using batch upload

With the use of indexing the base field, search time and time needed to load the whole page can be reduce to a satisfactory level which does not degrading much on user experience. However, uploading time still cannot be shorten as insertion of all data cannot be skipped as we cannot determine which data entries is useful or not, so insertion of all data would be better than skip part of them.

5. DIVISION OF LABOUR

To develop this system, Chris and I each responsible for part of the functions.

Tasks	Wong Tsz Hing, Chris	Siu Lik Hang, Leo
Database design	.	.
System design	.	.
Interface design	.	.
Web interface		.
Dictionary function	.	.
Internal linkage	.	
Search function	.	
Listview function		.
Data management function	.	.
Data upload function		.
Structure management function	.	
User management related function	.	
Communication related function		.

Table 9-Division of work

6. AMENDMENT

6.1 Change of batch upload file format

For the initial scope, .CSV file would be the default file to be used in batch file. However, after several testing, we find that saving .CSV file in utf-8 is not support among all version of EXCEL, which will cause the file content corrupt and hand to retread back the original data. After discuss to my teammate Chris, we decide to change the default file format to .TXT file in Unicode with every row ends with an indicator "END".

6.2 Removal of Game feature

As we arrange game feature as low propriety feature in interim report and we barely implement all functionality within the schedule, so game feature was removed.

7. PROBLEMS ENCOUNTERED

7.1 Implementation of flexible database structure

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, instead of using id as the unique key, base was used as the unique key. Also, searching time will be lengthen when large amount of data is added to database, it was solved by indexing the base field in “Attribute table”.

7.2 Limitation of time

Apart from the final year project, my teammate Chris and I still have to deal with other coursework. These result in a tight schedule together with a broad project scope. Learning from this experience, if similar project will be carried in future, meetings should be held regularly to report the development progress and adjust the schedule according to the progress. Also, we should consult more frequently to our supervisors for seeking advice and improving the functionality of the project.

8. FUTURE IMPROVEMENT AND DISCUSSION

8.1 Use of framework in interface design

At the beginning of this project, we did not consider to use any existing framework in interface design and we were advised to use framework in interface design at the middle of the project, but it is hard to integrate the framework into our project as most of the coding had to be reviewed and did the coding again. As a result, our website interface does not look attractive enough. For the next project, a framework should be chosen at the beginning of the project.

8.2 Support more file format batch upload

Currently our system only supports .TXT in Unicode, if a user having a file in other format, a lot of work may need to be done to change the to the default format. To support a real “smart upload”, the system should be able to support as many files format as possible, ie. .CSV, .DOCX or even upload multiple image or audio within only clicking once.

8.3 Insert or update data when using batch upload

For the implementation of the batch upload function, all the entries will be treated as insertion to the database. It will be more efficient for the early stage when the system is in use as the database does not contain many data inside, it is not likely to have duplicate data in the database. By when the amount of data grows to a considerable amount, ie 5000 entries, insertion data via batch upload will have a higher chance of generating entire duplicate data or partial duplicate data, which does not contribute much to the dataset but only occupying the space of database, and it will be time consuming to delete the duplicate data or combine with the existing data manually.

We had also consider that update of data should be allow when using batch upload, however, the issue would be when to update the data. Updating the existing data can be relied on several criteria:

1. Last update time of the target base

Data may need to update when the last update time is long (ie 1 year), but the reason behind this may be it has been validated and no long need to update.

2. Value of the data (upload by batch upload or edit via user input)

For updating data inside a base, we can consider that data input via user input (edit the data directly) is more valuable than batch upload. If we allow batch upload can directly update all the data, these valuable data may be removed.

With these consideration, as a decent conclusion has not drawn, the system only supports insertion of data to database given the fact that not all the data uploaded via batch upload contribute value to dataset.

8.4 Serve as a database for generic usage

In this project, our database only serves to store the relationship between Chinese characters or words, with Base_1 as characters and Base_2 as words, but we discovered that it can also serve for other purpose by simply changing the value of base_1 and base_2 or even adding more “Base table”.

Example like set up a customer management system for internal use.

Base_1 now change to “Normal customer” and Base_2 to “VIP”, the system administer only need to set up table in MySQL server and change of the coding in the PHP part, once all of these were set up, he/she can upload the customer data via batch upload and display them by the source code provided by us. This can help to reduce the time and effort needed to put in setting up a system.

With our source code as the basic framework, we believe that it can serve as a database for generic usage.

9. CONCLUSION

With the cooperation and support by teammates, we developed a flexible database system with a corresponding web application. This application can be used to provide a platform for public to contribute their knowledge in Chinses characters and words.

The source code of this project will be opened to the public. We hoped that others can use our source code as a framework to develop their own web application with the minimal time and effort.

LIST OF FIGURE

Figure 1-Main page	12
Figure 2-Login	13
Figure 3- Create account	13
Figure 4 -After logged in (Normal User)	14
Figure 5 -After logged in (Administer)	14
Figure 6-Character dictionary	15
Figure 7-Word dictionary	15
Figure 8-Search character/word	15
Figure 9-Search result	15
Figure 10-Choose attribute that will be listed	16
Figure 11-List View	16
Figure 12-Normal user (logged in)	17
Figure 13-Administer	17
Figure 14-Edit column/attribute	17
Figure 15-Upload page	18
Figure 16-File uploaded to page	18
Figure 17-Data uploaded to database	18
Figure 18-Character "是" with edit record	19
Figure 19-Delete record	19
Figure 20-Delete record (continue)	19
Figure 21-Record deleted and edit record updated	19
Figure 22-Try to recover record	19
Figure 23-Record successfully recover	20
Figure 24- Auto internal linkage	20
Figure 25- Compare normal record and records that contains internal links	21
Figure 26- System flow	22
Figure 27- Relationship between a page and the tables	26
Figure 28- Audit log shown for each page	27

Figure 29-Overview to the whole table structure
29

LIST OF TABLE

Table 1-Compare of different website	8
Table 2-Structure of base table	24
Table 3-Structure of attribute table	23
Table 4-Structure of map table	24
Table 5-Structure of audit log table	25
Table 6-Structure of login related table	30
Table 7-Structure of communication related table	31
Table 8-Testing result after using batch upload	33
Table 9-Division of work	34

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>
- [3] Retrieved April 14, 2018, from <https://www.digitalocean.com/community/tutorials/sqlite-vs-mysql-vs-postgresql-a-comparison-of-relational-database-management-systems>
- [4] Retrieved April 14, 2018, from <http://php.net/manual/en/intro-what-is.php>

APPENDIX

Appendix 1: Project Schedule and Milestones

Timeline	Task
September 2017	<ol style="list-style-type: none">1. Detailed project plan2. Project webpage
October to December 2017	<ol style="list-style-type: none">1. Design and set up database2. Website template
December to January 2018	<ol style="list-style-type: none">1. Search function2. Interim report3. User management
February	<ol style="list-style-type: none">1. Handle multimedia2. Smart upload
March 2018	<ol style="list-style-type: none">1. Audit function2. Linking function3. Listing view4. Testing5. Security review
April 2018	<ol style="list-style-type: none">1. Final report2. Final presentation
May 2018	<ol style="list-style-type: none">1. Project exhibition

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://chineseeffect.com/lessons/characters/>

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

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