



The University of Hong Kong
Faculty of Engineering
Department of Computer Science
Final Year Project Plan

Web based tool for Chinese character evolution

Supervisor: Dr. Vincent Lau

Name: Matthew Chan Tsz Ho

University Number: 3035269368

Content

1. Background
2. Objective
3. Methodology
4. Related work
5. Scope
6. Feasibility Assessment
7. Deliverables
8. Schedule
9. Risk
10. Reference

1. Background

Chinese character has about 5000 years of history[1]. 5000 years ago, people started using pictures to symbol what they saw. Those pictures are called "oracle bone script" nowadays. Oracle bone script is significant in the study of Chinese history and culture. For example, what a cart looked like in the past can be known by studying the word cart in oracle bone script[2]. Every change in Chinese character may represent a significant change in political power in the history of China. For instance, when the Qin Dynasty united the six kingdoms, the ruler formulated the "small seal script". The evolution of Chinese character made an essential contribution to Chinese history so the data of it should be well preserved. There ought to be a platform storing all the data.

Chinese characters are not merely pictures in the modern era. They have straight rules to be written. For example, the order of strokes of every Chinese character is well defined. On the other hand, the pronunciation of Chinese characters is also special that there may be more than one pronunciation for each word depending on the situation or way to use it. Currently, there are several platforms that can perform the task of online dictionaries or databases. However, those platforms do not have the function of user-contributed content. While there are 4000 Chinese characters and 15000 Chinese words commonly used in Hong Kong[2], updating or adding the data Chinese character by a single group requires much effort. If the idea of user-contributed content is implemented on the database, the information in the database can be more up-to-date and the quantity and coverage of data can also be increased.

2. Objective

This project aims to construct a responsive website about Chinese characters. The basic part of Chinese character, for instance, components, structure, phonetic code, strokes and words will be covered. The data of Chinese Character will be stored in the database and will be displayed to visitors of the website. Visitors can perform several functions including searching the words, exploring related characters and words. Visitors can also register as a member and their preference of using the website can be set. The website also supports multi-media that videos and pictures can be uploaded and displayed on the website. To make the website user-friendly, a responsive layout will be designed that the website has different layouts in desktop and mobile phone in order to make sure the website functions well in different devices.

A suitable database will be designed to suit the use of the website. Not only will the database be able to be supplemented by different users, but it will also be able to expand its fields. To facilitate the efficiency of inputting many Chinese characters, the website allows the reading of text file and the data will be put in the database. The website will construct web pages based on the data in the database. Besides reading the text files, there will also be a page that allows users to contribute their data.

At the latter part of the project after the user-contribute-content function is implemented. A further research on the user's preference on contributing the content will be conducted. How the users contribute data to the website will be examined. With the data collected, the website can be adjusted to be more user-friendly.

3. Methodology

The main programming language required to construct the web tool are Node.js, CSS and JavaScript. There will be a database storing the Chinese characters. The database needs to support dynamic increasing field so NoSQL will be used.

Node.js will generate the required HTML for the basic of the website. CSS will be responsible for making the style of the website and JavaScript will realize the special effects.

Function for users

Chinese Dictionary

Users can treat the website as a Chinese dictionary. Details of Chinese characters, like components, structure, phonetic code, strokes and words will be shown. Pictures and video are also displayed as supplements.

Searching and sorting

Users can search for data they want. They can select the criteria of searching, like by Chinese character or by pinyin. After searching, the result displayed can be further sorted by the number of strokes or by radical, etc..

Preference setting

Users can set their preference of using the website after logging in, including the priority of different fields displayed on the website, searching and sorting criteria, etc..

Related pages

There are hyperlinks inside different pages. The hyperlinks will be linked to other related pages. For example in the page of the Chinese character "女" (women), there will be hyperlink linked to "媽"(mother).

Functions for contributors

Editing and adding page

The editing page allows the user to edit or add the data of Chinese character. Anyone can edit the data of the website. To improve security, the user may be required to register and login before they edit.

Editing History

The history of editing and adding Chinese characters will be stored in the user's database. The user can view his edit history in the profile.

Reading text file

The system allows the reading of text file. After reading the file, the information inside the text file will be transmitted to the database. The text file consists of the information line by line. The format of the information is “field,field,field”, where each field is separated by a delimiter comma. The first line of the text file indicates the name of the different field. The data starts from the second line. The following will be an example:

字,筆劃數,部首

媽,13,女

爸,8,父

UI Design

Homepage

MENU

Name of the website

Introduction of the website

Recommended content

Search result

Menu

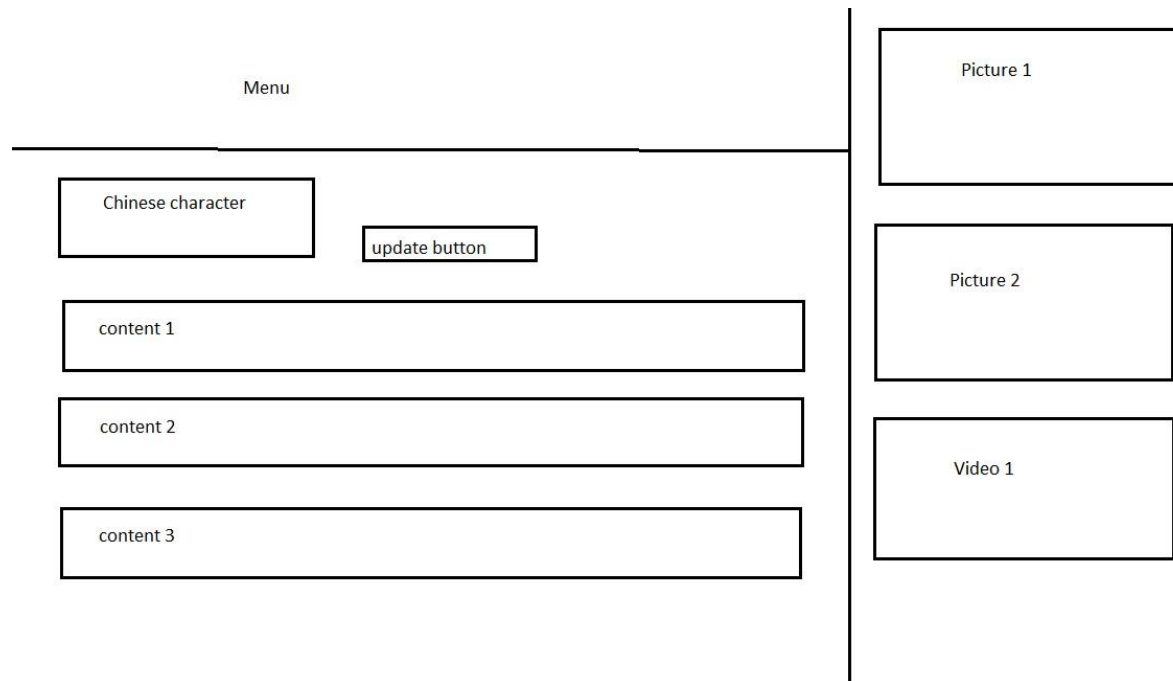
Title

search result 1

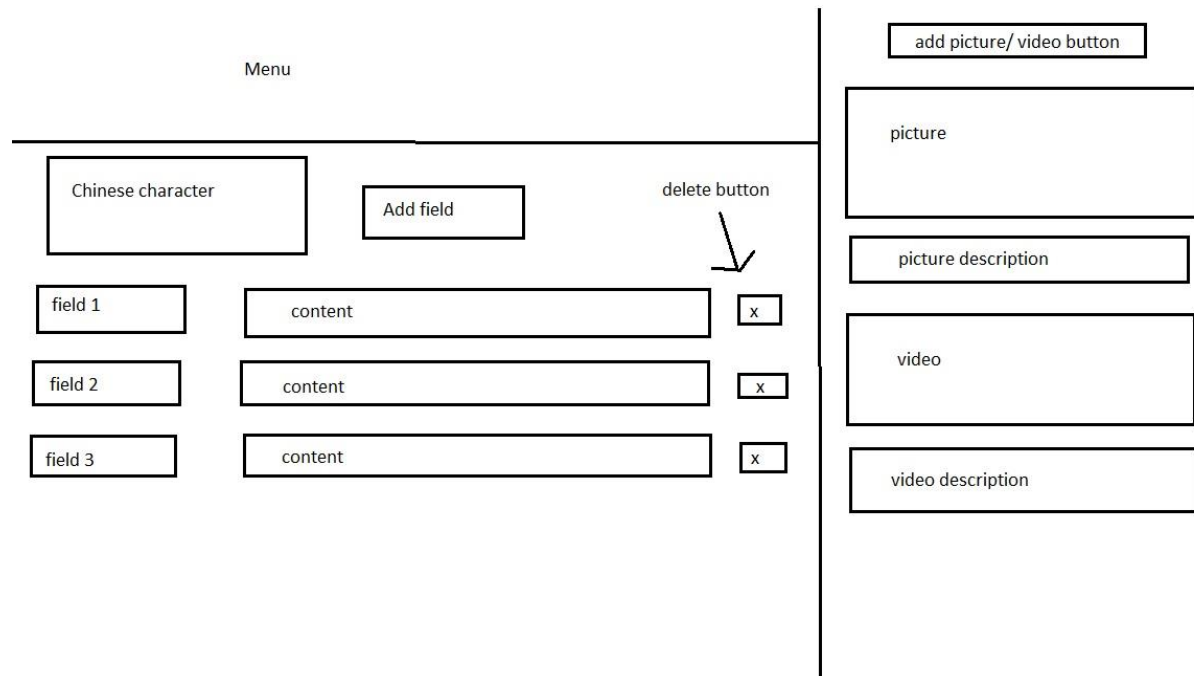
search result 2

search result 3

Main content of each Chinese character



User-input page



4. Related work

Currently, there are a number of platforms that servers the use of a dictionary. Reference to the website can increase the feasibility integrity of the project.

"Lexical Items with English Explanations for Fundamental Chinese Learning in Hong Kong Schools" can search by direct character or character or radical, strokes and pinyin[3]. There are also several ways of sorting the result. Flash is used to show the stroke sequence of the character.

"Chinese Character Database: With Word-formations" is a very detailed online dictionary that includes homophone, pronunciation and word example, etc.[4]. User can search by the exact character or pronunciation. Waveform Audio File is used to play the pronunciation of the character.

Wikipedia is also taken as reference for its user contribute system[5]. There are links in the paragraphs and they will link to related pages. Users can edit the existing pages and can also create new pages. Before a user edits anything, he can log in or his IP address will be shown on the page. The previous contribution of a certain user or IP address can be inquired.

5. Scope

The website will be about Chinese characters. The basic parts of Chinese characters like strokes will be included on the website. However, the field of Chinese characters will not be limited. It can be dynamically increased depending on the input from contributors.

Multi-media format of data is supported in the website. Users can add text as well as upload pictures and videos to the database.

After the user-contributing data system is launched, an analysis of the preference of the user will be conducted. We will investigate how the users contribute the data and the similarity and difference of the data contributed.

6. Feasibility Assessment

This project mainly focusses on web tool development and a central management system. There are several websites that can be taken as reference. Also, the data of Chinese Character is provided so the risk of having insufficient data is low.

Moreover, there is a detailed and reasonable schedule for the project. The workload of building the whole project will not be unbearable if the schedule is strictly followed.

7. Deliverables

Date	Deliverables
30 Sep 2018	Deliverables of Phase 1 (Inception) <ul style="list-style-type: none"> Detailed project plan Project web page
7-11 Jan 2019	First presentation
20 Jan	Deliverables of Phase 2 (Elaboration) <ul style="list-style-type: none"> Preliminary implementation Detailed interim report
14 Apr	Deliverables of Phase 3 (Construction) <ul style="list-style-type: none"> Finalized tested implementation Final report
15-19 Apr	Final presentation
29 Apr	Project exhibition
29 May	Project competition

8. Schedule

Phrase	Plan
September	<ul style="list-style-type: none"> Examine different Chinese characters websites and applications. Those websites' functions and layout will be a reference to this project. Study related programming language. The use of HTML, CSS, Javascript, and NoSQL is needed in this project.
October	<ul style="list-style-type: none"> Design the detailed layout of the website. Decide the color and position of different parts of the website.
November - December	<ul style="list-style-type: none"> Implement the layout of the website. Make the website be able to read text file and generate corresponding HTML.
January	<ul style="list-style-type: none"> Prepare for presentation. Implementation of User-contributed Content. There will be a page where users can contribute data to the website.
February	<ul style="list-style-type: none"> Final Test. Analysis the user-contributed data.

9. Risk

Impact	Probability	Risk	Mitigation
high	high	Tight schedule The time of finishing the project is limited and there are several deadlines for deliverables. There is a risk that the project cannot be finished within the deadline.	A reasonable schedule is planned. With good time management, the risk of finishing the project beyond deadline is reduced.
high	high	Not familiar with the programming language There are several programming languages required to finish the website. Unfamiliar with the programming language may hinder the development of the project	Time will be spent on studying the programming languages. Books and online resources will be referred in case of any questions.
High	Low	Malicious edit of data from users Since the data on the website can be edited by users. Some malicious users may edit false information and may even delete the data.	A registration system will be launched. The trace of the edit will be recorded. If a user's behavior is suspicious, the user will be banned from editing.
middle	low	User not familiar with how to use the website Users may be confused by the functions of the website and do not know how to use the functions.	A guide will be provided to assist users to use the page

10. Reference

[1] 郭小武教授 . (n.d.). 漢字的起源與變遷 .[Online]. Available: https://www.chiculture.net/index.php?file=topic_description&old_id=0601 [Accessed: 2018, Sep 7].

[2] Evolution of Chinese characters. (n.d.). Omniglot.[Online]. Available: <https://www.omniglot.com/chinese/evolution.htm> [Accessed: 2018, Sep 7].

[3] Prof. Tze-wan KWAN (n.d.) 粵語審音配詞字庫 [Online]. Available: <https://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/> [Accessed: 2018, Sep 24]

[4] Hong Kong Education Bureau (n.d.) Lexical Items with English Explanations for Fundamental Chinese Learning in Hong Kong Schools [Online]. Available: https://www.edbchinese.hk/lexlist_en/ [Accessed: 2018, Sep 24]

[5] Wikipedia (n.d.) [Online] Available: <https://www.wikipedia.org/> [Accessed: 2018, Sep 24]