Chinese Language Retention System

By

Joshua McFarland

Submitted to
the Faculty of the Information Technology Program
in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science
in Information Technology

University of Cincinnati
College of Engineering and Applied Science

June 2010
Chinese Language Retention System

by

Joshua McFarland

Submitted to
the Faculty of the Information Technology Program
in Partial Fulfillment of the Requirements
for
the Degree of Bachelor of Science
in Information Technology

© Copyright 2010 Joshua McFarland

The author grants to the Information Technology Program permission
to reproduce and distribute copies of this document in whole or in part.

_________________________    ____________________
Joshua McFarland          Date

_________________________    ____________________
Robert Schlemmer         Date

_________________________    ____________________
Hazem Said, PhD          Date
Abstract

The *Chinese Language Retention System* is a web-based application that allows students of the Chinese language to more effectively manage and study their material. Students are able to input new material into the system and then review them through an interactive interface. The results are stored allowing for dynamic and focused review sessions based on how well known the study material is. By combining aspects of spaced repetition learning and handwriting recognition students are able to achieve a higher degree of long term memory retention compared to that of traditional paper note cards. The result is an easy to use, well orchestrated and effective median for helping to learn the Chinese language.
# Table of Contents

Abstract ........................................................................................................................................... 3  
Table of Figures .............................................................................................................................. 5  
Introduction ..................................................................................................................................... 6  
Product Description ........................................................................................................................ 6  
User Profiles .................................................................................................................................... 7  
  Students ....................................................................................................................................... 7  
  Teachers ...................................................................................................................................... 7  
  Administrators ............................................................................................................................. 8  
Design Protocols ............................................................................................................................. 8  
  Concept Overview ...................................................................................................................... 8  
  Technology ................................................................................................................................. 9  
  Color Scheme ............................................................................................................................ 10  
  Layout ....................................................................................................................................... 10  
  Database Diagram ..................................................................................................................... 11  
Testing Plan .................................................................................................................................. 12  
  Example Test Scenario ............................................................................................................. 12  
  Network Setup (Local Testing) ................................................................................................. 12  
Risk Management ......................................................................................................................... 13  
Proof of Design ............................................................................................................................. 14  
  Login ......................................................................................................................................... 14  
  Dictionary ................................................................................................................................. 14  
  Manage ...................................................................................................................................... 17  
  Study ......................................................................................................................................... 18  
    Writing Practice ..................................................................................................................... 18  
    Pinyin Practice ....................................................................................................................... 24  
Project Planning ............................................................................................................................ 26  
  Budget ....................................................................................................................................... 26  
  Timeline .................................................................................................................................... 26  
Project Deliverables ....................................................................................................................... 26  
Conclusion .................................................................................................................................... 28  
Works Cited .................................................................................................................................. 29
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chinese Language Enrollments</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Conceptual System Overview Diagram</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Color Scheme</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Demonstration Layout</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>Database Diagram</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>Local Testing Network Diagram</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Login Screen</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>New Dictionary Search</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>Dictionary Search Results</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>Management Panel</td>
<td>17</td>
</tr>
<tr>
<td>11</td>
<td>Study Mode Selection Screen</td>
<td>18</td>
</tr>
<tr>
<td>12</td>
<td>New Handwriting Screen</td>
<td>19</td>
</tr>
<tr>
<td>13</td>
<td>Handwriting Recognition 1</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>Handwriting Recognition 2</td>
<td>21</td>
</tr>
<tr>
<td>15</td>
<td>Handwriting Recognition Feedback 1</td>
<td>22</td>
</tr>
<tr>
<td>16</td>
<td>Handwriting Recognition Feedback 2</td>
<td>23</td>
</tr>
<tr>
<td>17</td>
<td>New Pinyin Practice</td>
<td>24</td>
</tr>
<tr>
<td>18</td>
<td>Pinyin Practice</td>
<td>25</td>
</tr>
<tr>
<td>19</td>
<td>Pinyin Practice Feedback</td>
<td>25</td>
</tr>
</tbody>
</table>
Introduction

The enrollment of English speaking students in Chinese language classes has more than doubled since 2002. Figure 1 below shows the significant increase in American college students enrolled in Chinese language courses between 1960 and 2006 (3). This substantial growth has led to an increasing demand for language learning materials and other solutions.

![Chinese Language Enrollments](chart)

The primary educational tool used in the classroom environment is still textbooks. In many cases the textbooks in use are rather old and dated. The problem with textbooks is that they are mainly for reference and lack an element of interaction that is important to overall memory retention. By using computer technologies the traditional limitations placed on textbook style learning can be greatly improved upon.

Product Description

The *Chinese Language Retention System* is a web-based tool designed to help manage and increase overall retention in students. It combines interactive note cards with an intelligent management system to continually focus students study time. This means that rather than
continually reviewing materials that the student already knows it focuses attention on the items causing the most difficulty.

The interactive note cards are built taking in several aspects of human computer interaction to speed up the memorization process. By engaging multiple senses in the study process the student becomes comfortable with material faster than traditional methods. Each of the study methods included in this application engages the sense of sight, hearing and touch.

The management system on the backend is built around the methodology of spaced repetition learning. It is constantly evaluating progress and determining the next time items should be reviewed. The overall goal being to review an item just before it will be forgotten and slowly increase the interval of time in between reviews each time. Combining all of these elements into one easy to deploy web-based application allows for a much more efficient studying.

**User Profiles**

**Students**

Students are the most common type of users that interface with the system. They use the system to study new material and review older material. Limited access will be provided to them allowing them to only manage their own vocabulary, search the dictionary and use the built in study applications.

**Teachers**

The teacher role is assigned to allow for a person to manage a group of students. They will have control over the student accounts assigned to them. Through the management interface they can view individual student progress and also add new material to their study list.
Administrators

Administrators can manage general system settings and also create new students and teachers.

Design Protocols

Concept Overview

This project was designed to be very object oriented and can be broken down into several logical pieces. Figure 2, listed below shows the various pieces and how they are linked together. The pieces have been color coded based on the type of object; yellow are interfaces, red are abstracted classes and blue are stored data.
Management Panel: This interface allows users to manage their study data by allowing them to add, delete, edit and sort material.

Handwriting: This interface prompts the user with an English definition and phonetic pronunciation in Chinese. It contains a writing area that allows the user to draw the corresponding character. The handwriting application is also connected to a data file that contains the Chinese character stroke order.

Pinyin: The pinyin interface prompts the user with both the Chinese character and English definition and expects them to enter the corresponding Chinese phonetic pronunciation.

Other: This just signifies that other study interface could be added as objects to interact with the system. The client may request other study applications depending on their particular needs.

Manage: This is a class to manage user data.

Study: This is a class to facilitate studying.

User Data: Stores all of the user’s data which includes vocabulary and the study statistics linked to them.

Dictionary: A comprehensive database containing dictionary entries for over 80,000 Chinese characters and words. It is built by using an open source Chinese dictionary project called CC-CEDICT (2).

Technology

- Apache 2.2.14
- JavaScript/AJAX
- MySQL 5.0.81
- PHP 5.2.11
Color Scheme

The color scheme for the website was chosen to offer minimal distraction while keeping a relaxing and professional look. Refer to Figure 3 shown below and the corresponding numbers for the specific usage of each color. *It is important to note that these are the suggested colors that are used in the application demonstration.*

1. Panels
2. Background
3. Buttons
4. Borders

Layout

The layout in Figure 4 was used to demonstrate the project throughout the senior design sequence. It was built with simplicity in mind and not overwhelming the user with numerous options. The interface has three main sections; navigation, dictionary and content panels. The content panel is where the management and studies features display when selected.
Figure 4: Demonstration Layout

Database Diagram

Figure 5: Database Diagram
Testing Plan

The two major types of testing that were used in the project were iterative and user testing. The iterative testing is a continual process that is done each time the project undergoes a major revision. This type of testing helped debug and mitigate general functionality errors that exist on the backend of the project.

User testing was the more vital piece of testing for this project because it is a project that has a high level of user interaction. The system was tested using actual Chinese language class who were asked to provide feedback based on their experiences. The testing was completed using 5-10 students from the University of Cincinnati’s Chinese program. The main target of this testing was to fine tune default system settings and increase interface usability.

Example Test Scenario

<table>
<thead>
<tr>
<th>Steps</th>
<th>Details</th>
</tr>
</thead>
</table>
| 1. Log into the website using the credentials provided | Username: <your last name>  
Password: <your last name> |
| 2. Go to the manage tab and add (我，你，好，快，了) to the study queue | The words should be listed on the manage page before you continue to the next step. |
| 3. Go to the study tab and begin reviewing the material for 2-3 minutes | Follow the instructions provided by the application to progress through the review process. |
| 4. Log out of the website and complete the feedback questionnaire provided | |

Table 1: Example Test Scenario

Network Setup (Local Testing)

This was network setup for the Tech Expo presentation. It consisted of a Windows XP desktop that has an Apache server installed and configured for both MySQL and PHP. The Apache server was locally hosted. The Windows 7 laptop was connected to the network and used for presentation purposes. The other Windows XP client was a virtual instance located on the server machine for presentation purposes as well.
Risk Management

There are several different risk factors that have taken into consideration for this project. The two most probably to cause real problems are web host failure and/or a computer without the proper updates. Steps have been taken to mitigate these issues at Tech Expo by creating multiple backups of the web server and properly updating prior to the event.

<table>
<thead>
<tr>
<th>Risk</th>
<th>Level</th>
<th>Impact</th>
<th>Probability</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web host failure or outage</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>A local web server will be installed on my laptop and available at all times</td>
</tr>
<tr>
<td>Critical physical hardware failure</td>
<td>Low</td>
<td>High</td>
<td>Very Low</td>
<td>All material will be backed up in a separate locations</td>
</tr>
<tr>
<td>Computer does not have the proper updates</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Install updates or move to another machine for testing</td>
</tr>
<tr>
<td>Exceed the maximum allotted budget</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Select a more affordable hosting package or implement open source</td>
</tr>
</tbody>
</table>

Table 2: Risk Management Plan
Proof of Design

The application proof of design can be broken down into four sections: login, dictionary, manage, and study. Each section will be thoroughly explained and documented using screenshots throughout this section.

Login

One of the key components to this system is being able to uniquely identify users to associate data with. The login screen allows users to sign in using their username/password combination. It encrypts the password and is protected against most SQL injection type of attacks.

![Login Screen](image)

Dictionary

The dictionary search allows the user to search the built-in dictionary from three different criteria as shown in Figure 8. The first is a character search which requires that the user have the ability to type using Chinese characters. The second is the definition search which allows the user to enter and English definition to return results. The third is phonetic Chinese pronunciation or pinyin which uses English characters to search for Chinese sounds.
The dictionary is built using AJAX which means that after the search button is click the page doesn’t need to refresh. This functionality is very useful as it won’t interrupt users other studies if they need to make a quick dictionary search. As shown in Figure 9, there results are loaded below the search box.
The green plus arrow allows the user to quickly add words to their vocabulary from the dictionary. After it is clicked, a message box pops up informing the user that the item has been successfully added. The green arrows won't display next to items that have already been added.
**Manage**

The management screen contains all of the relevant study data organized with the lowest retention percentage first. It allows the user to delete individual items or clear all of the items that they are studying.

![Management Panel](image.png)

**Figure 10: Management Panel**

<table>
<thead>
<tr>
<th>Character</th>
<th>Pinyin</th>
<th>Definition</th>
<th>Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>猫</td>
<td>mao1</td>
<td>CL寳 zhì, cat</td>
<td>NEW</td>
</tr>
<tr>
<td>高兴</td>
<td>gao1 xing4</td>
<td>willing (to do sth), in a cheerful mood, happy, glad</td>
<td>45.5%</td>
</tr>
<tr>
<td>你们</td>
<td>ni3 men5</td>
<td>you (plural)</td>
<td>71.4%</td>
</tr>
<tr>
<td>你</td>
<td>ni3</td>
<td>you (informal, as opposed to polite 尊 [zūn])</td>
<td>72.7%</td>
</tr>
<tr>
<td>我们</td>
<td>wo3 men5</td>
<td>we, us, ourselves, our</td>
<td>75.0%</td>
</tr>
<tr>
<td>太好了</td>
<td>tai4 hao3 le5</td>
<td>very good</td>
<td>75.0%</td>
</tr>
<tr>
<td>查查</td>
<td>kan4 kan5</td>
<td>to take a look at, to survey, to examine, (coll.) pretty soon</td>
<td>81.8%</td>
</tr>
<tr>
<td>中国银行</td>
<td>Zhong1 guo2 Ym2 hang2</td>
<td>Bank of China (BoC)</td>
<td>83.3%</td>
</tr>
<tr>
<td>五</td>
<td>wu3</td>
<td>five, 5</td>
<td>85.7%</td>
</tr>
<tr>
<td>九</td>
<td>jiu3</td>
<td>nine, 9</td>
<td>85.7%</td>
</tr>
<tr>
<td>八</td>
<td>ba1</td>
<td>eight, 8</td>
<td>100.0%</td>
</tr>
<tr>
<td>四</td>
<td>si4</td>
<td>four, 4</td>
<td>100.0%</td>
</tr>
<tr>
<td>妈</td>
<td>ma1</td>
<td>mother, mom, ma</td>
<td>100.0%</td>
</tr>
<tr>
<td>二</td>
<td>er4</td>
<td>two, stupid (Beijing dialect), 2</td>
<td>100.0%</td>
</tr>
<tr>
<td>中文</td>
<td>Zhong1 wen2</td>
<td>Chinese written language, Chinese writing, Chinese</td>
<td>100.0%</td>
</tr>
<tr>
<td>六</td>
<td>liu4</td>
<td>six, 6</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

You have a total of 16 items and your overall retention rate is 85.1%
The power in the management feature is being able to view the retention percentage. The retention percentage is derived directly from the results of studying. It is calculated by factoring in the percentage correct and is also modified by the number correct in row or streak. The idea is that the system will more frequently select the lowest five retention percentages to study until they increase. At the bottom of the Figure 10 it shows overall review statistics. If the overall retention rate falls below 60% it is a good indicator to stop adding new material and focus on increasing the percentages of the current material.

**Study**

The study area allows for the user to choose which mode of study they wish to engage in. The available study modes could vary depending on the particular implementation of the project and need of the client. For the purposes of this implementation a study mode of writing and pinyin practice have been created.

**Select Mode:**

![Writing Pinyin](image)

*Figure 11: Study Mode Selection Screen*

**Writing Practice**

Writing practice allows the user to practice drawing and recognizing written Chinese characters. At first they are prompted with the phonetic Chinese pronunciation, English definition and the audio pronunciation. This is enough information for the user to be able to write the Chinese character if they know it.
As the user begins to draw the character a list of possible answers is dynamically being generated, as seen in the lower right corner of Figure 13. This possible answer list contains suggestions based on the current entered strokes and also throws in a few unrelated characters.
When the user finally finishes writing the character and sees it in the suggestions list they can then select it. It is moved from the suggestions box to just below the question prompt. The user can then continue writing characters until they have built the word. They can also clear the canvas and also clear the current word if a mistake has been made.
Figure 14: Handwriting Recognition 2

After the user has finished building the work and next is selected the answer is checked.

Figure 15, shown below is the feedback returned to the user if the answer is correct.
Figure 15: Handwriting Recognition Feedback 1

Figure 16, shows the feedback provided if the entered answer was incorrect. It will give the user the correct answer and then move on to the next item.
In summary the writing practice mode is extremely powerful due to the number of reinforcing senses used in the process. It uses three of the five main senses; touch, sight and hearing.
Pinyin Practice

Pinyin practice prompts the user with the Chinese character and English definition which is enough information to then enter the phonetic Chinese pronunciation or pinyin.

The user is given a text field which allows them to manually type the pinyin. It is important to note that the number at the end of each group of letters is a tone mark. The Chinese language has five different tones which signify different vocal inflections that can alter the meaning.
Similar to writing, pinyin practice also provides feedback to whether or not the answer entered was correct. Figure 19 shows the feedback provided when the incorrect answer is entered.
After the user has answered the question the word is read aloud to help reinforce the pronunciation of the term. The study mode engages two of the five senses; sight and hearing.

**Project Planning**

**Budget**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Cost Incurred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Hosting</td>
<td>Hosting package including: Apache, MySQL and PHP</td>
<td>$79.99</td>
</tr>
<tr>
<td>Dell Inspiron Laptop</td>
<td>Personal laptop used to complete developing</td>
<td>$0.00</td>
</tr>
<tr>
<td>HP Pavilion Desktop</td>
<td>Used for local hosting of the project at Tech Expo</td>
<td>$399.99</td>
</tr>
<tr>
<td>Adobe Dreamweaver CS4</td>
<td>Professional web editor used in the development process</td>
<td>$0.00</td>
</tr>
</tbody>
</table>

**Total:** $479.98

**Table 3: Project Budget**

**Timeline**

<table>
<thead>
<tr>
<th>ID</th>
<th>Task Name</th>
<th>Start</th>
<th>Finish</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Research Topic</td>
<td>9/23/2009</td>
<td>12/14/2009</td>
<td>11.8w</td>
</tr>
<tr>
<td>5</td>
<td>Audio Prompts</td>
<td>1/4/2010</td>
<td>3/12/2010</td>
<td>10w</td>
</tr>
<tr>
<td>7</td>
<td>Debugging/Testing</td>
<td>2/22/2010</td>
<td>5/21/2010</td>
<td>13w</td>
</tr>
</tbody>
</table>

**Table 4: Project Timeline**

**Project Deliverables**

The following is list of the original project deliverables set in the design freeze. Below each of them is a brief explanation of how they have been met.

- Interactive JavaScript-based Chinese study application
  - Ability to play audio with proper Chinese pronunciation
Handwriting recognition
Intelligently prompt the user with material to review

The study application recognizes user handwriting based on the Chinese stroke order and will generate a list of potential answers. Audio files have been added to the project allowing the application to dynamically read all dictionary items in Chinese. The backend management system uses the gathered item study statistics to select the items that need to be reviewed the most.

- Fully searchable web-based Chinese dictionary
  - Ability to play audio with proper Chinese pronunciation
  - Add new material to the study list
  - Searchable by definition, pinyin and Chinese characters

By importing the CC-CEDICT into a MySQL database users can search over 90,000 dictionary entries using multiple search methods (2). Each of the entries is linked to the proper audio files and can be added directly to a user’s personal vocabulary study list.

- Easily manage study materials

The management panel allows the user to view all added items sorted by retention percentage. It is built into an easy to read table allowing for manual deletion of unwanted items.

- Administration module allowing for easy configuration of general application settings

The administration panel allows for configuration of some of the system’s basic settings. A user with administrator privileges can manage users and also modify dictionary entries. The formula used to calculate retention percent and item degradation can also be modified.
Conclusion

The *Chinese Language Retention System* is an effective method for learning and retaining Chinese study material. By effectively using human computer interaction and spaced repetition learning methods it creates a powerful environment for maximizing the learning experience. The application itself is very adaptable and could be easily implemented into pre-existing websites hosted by educational institutes.
Works Cited


