Smart Chinese Reader: A Chinese Language Learning Aid with Web Browser

俞毅峰 (Yifeng Yu)
Computer Department, Zhangzhou Institute of Technology, Zhangzhou, Fujian, China

Smart Chinese Reader (SmartCR for short) is a program based on NLP (natural language processing) technology to help you learn Chinese language through deep reading. It provides Chinese word segmentation, Chinese part of speech tagging, Chinese to English translation, example sentence search, and text to speech conversion. It is published by NLPtool.com whose website is http://www.nlptool.com/. You can pick up any Chinese news, stories, or essays interesting to you on internet, and use Smart Chinese Reader to explore them. Compared with dictionary apps, it lets you gain more Chinese language knowledge (meanings and usages of Chinese words, patterns and even rhythms of Chinese sentences) from a text, rather than just to get through the text. It makes your Chinese learning more effective.

Smart Chinese Reader's interface consists of seven tabs: Browser tab, Text tab, Beginner tab, Advanced tab, Examples tab, HSK tab, and Review tab. In addition, there is a popup Chinese dictionary available in most of these views.

1. Browser Tab
You can read any Chinese web page easily in the Browser tab (see Figure 1). Enter its address in the address bar and click on the "Go" icon to go to the page. Then, simply hover your mouse over a word to pull up its definition right on the page. The word's pinyin romanization and HSK (Hanyu Shuiping Kaoshi, Chinese Proficiency Test) level are also provided in a small popup window. To hear the word pronounced, press "Ctrl + z" or click on the "Speaker" icon on the popup dictionary.

Figure 1: Visiting taobao.com with SmartCR’s popup dictionary
When you consider a section in a page worth a better understanding and want to study it in depth, select the section with your mouse, then click on the "Study" button (see Figure 2). This brings up the Beginner and Advanced tabs, where the text is well formatted and easier to read.

![Select a section to study it in depth](image)

**Figure 2: Select a section to study it in depth**

### 2. BEGINNER TAB

You are taken to the Beginner tab (see Figure 3) after clicking the "Study" button in the Browser tab or the Text tab to study a section of text in depth. The Chinese text has been segmented and is rendered at three levels to help you understand and learn it:

- **sentence**: Text is split into sentences in terms of delimiters . , ! : and ;
- **clause**: A sentence is split into clauses in terms of the delimiter ,
- **word**: The lack of word boundaries in orthography is a major obstacle in Chinese reading. SmartCR recognizes words from a clause using a state-of-the-art NLP (Nature Language Processing) technology.
Besides word segmentation, the Beginner tab provides several learning aids for new learners of Chinese to understand Chinese words and expressions:

1) Color part of speech notation. Parts of speech (POS) of Chinese words in a sentence are marked with colors. The syntax structures of Chinese sentences are quite different from those of English sentences. With the color notation, you can familiarize yourself with the Chinese syntax in a short time. The legend of color POS notation is shown in Figure 4.

<table>
<thead>
<tr>
<th>POS</th>
<th>color</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun</td>
<td>bold</td>
</tr>
<tr>
<td>person name</td>
<td>underline</td>
</tr>
<tr>
<td>place name</td>
<td></td>
</tr>
<tr>
<td>verb</td>
<td></td>
</tr>
<tr>
<td>adjective</td>
<td></td>
</tr>
<tr>
<td>adverb</td>
<td></td>
</tr>
<tr>
<td>others</td>
<td></td>
</tr>
</tbody>
</table>

2) Chinese to English Translation: Clicking the "A" icon after a clause, its English translation is provided using the phrase-based statistical machine translation technology (see Figure 5). While moving the mouse over each phrase in a Chinese sentence, the corresponding phrase in the English sentence is highlighted, and vice versa. This allows you to learn translation at the phrase level.

3) Text-to-Speech: SmartCR will read the clause aloud in Chinese if you click the SPEAKER icon after a clause. Smart Chinese Reader uses the state-of-the-art text-to-speech (TTS) technology. Given a sentence, SmartCR can pronounce it in various voices.

3. Advanced Tab

If your Chinese proficiency is at HSK level 4 or higher, you may find the Beginner tab is a little bloated and unnecessary, then switch to the Advanced tab (see Figure 6). The advanced view displays contents in the normal format of Chinese text, i.e., no spaces between words. Users have to segment sentences by themselves. However, if you are uncertain about the segmentation of one part of a sentence, just click on the part. A Chinese dictionary will popup showing the word at the position with its definition.
The learning aids provided by the Advanced tab focus on Chinese text analysis, and improving your Chinese reading skills:

1) **Statistics:** The number of total Chinese words and words at different HSK levels are figured out. This gives you a measure of the difficulty of studying the document.

2) **Color notation of HSK levels:** Furthermore, words at different HSK levels can be marked in different colors.

3) **To-learn and known words highlighting:** To-learn and know words are highlighted in red and blue borders respectively. It shows your mastery of this text visually.

### 4. Examples Tab

The Examples tab can be accessed via double-clicking a Chinese word in the Beginner, Advanced, or HSK tab. This tab provides example sentences containing the current word. Example sentences offer a wealth of information and give you a sense of how the Chinese word is used. Each example sentence is a Chinese-English sentence pair, in which the Chinese sentence is original and the English sentence is its translation by human translator. Because sentences were originally conceived in Chinese, it ensures the authenticity of Chinese word usages.

### 5. HSK Tab

The HSK tab (see Figure 7) provides vocabulary and statistics of your known words for each HSK level (Hanyu Shuiping Kaoshi, Chinese Proficiency Test), including HSK 1, HSK 2, HSK 3, HSK 4, HSK 5, and HSK 6.
To-learn words are highlighted with red border and known words with red border in the vocabulary of each HSK level. It serves as a map clearly showing which HSK words you have been familiar with, which HSK words you haven't, and which HSK words you want to study further. For any HSK word you don't know, you can click on it to bring up the Chinese dictionary to see its definition, and double click to find its example sentences. This let you prepare HSK exam easily.

**ABOUT THE AUTHOR**

Smart Chinese Reader was developed by 俞毅峰 (Yifeng Yu), who is an instructor at Zhangzhou Institute of Technology, China. His study focuses on Natural Language Processing. Based on this technology, in 2014 he developed Smart Chinese Reader to help people read and learn Chinese. The program has been updated through several versions and made great progress since then.

Contact: yifeng_yu@nlptool.com