3
Keynote Speech:

"Digital Tools for Chinese Language Learning and Teaching:
CKC Code and its Online Dictionary"
By Dr. Esther S. C. Chan & Dr. K. H. Tse
The Hong Kong Institute of Education

When we study, it is inevitable to look up the dictionary. The traditional searching methods of the Chinese dictionaries are mainly based on the radicals, stroke sequence, or spelling. For those who have used dictionaries would have encountered the frustration of not knwoing how to find specific characters. For example when you are using a dictionary with the radical searching, what are the radicals of "凹, 五, 七, 九"? For dictionaries with the stroke searching method, what are the stroke sequence or the total strokes of "凸, 龖, 龜, 龘"? For dictionaries with the spelling searching method, what are the spelling of "蘒, 夼, 卣"?

With the development of information technology, it is much easier to look up the specific character from the on-line dictionary or CD-ROM. But what about when we do not know how to input this character? It seems we have to learn at least one Chinese input method so as to further our study. Among all the different input methods, which one is the easiest to learn?

Chinese input methods can be categorized into 3 modes: keyboard input, handwriting recognition and speech recognition. To people who are acquainted with Chinese handwriting, handwriting recognition input is the easiest way to handle. The disadvantage of it is the low input speed. Besides not all computers are equipped with   writing pads. Speech input is also easy to use but most of the program can only recognize one user's voice. That means you can use your own computer only. And if you do not know the pronunciation of the character, it is difficult to use. That is why keyboard input is the most popular method.

Chinese keyboard input methods also can be divided into 3 categories: by encoding, by pronunciation and by character structure. The encoding method (Chinese Commercial Code/Chinese Telegraph code) was used in the old days for transmitting Chinese text over telegraph. Since the the coding is based on the words' radical and total strokes, it is extremely difficult to remember. The pronunciation methods, such as Pinyin, Bopomofo, Cantonese and Hokkien, is based on the rules of phonetic spelling. If users have already mastered the rules, it is easy to use. However, if users do not know how to pronounce the character, e.g. 汞, it is impossible to input. The character structure methods such as Wubi[1], Cangjie[2] and Dayi[3], can solve the problem of pronunciation. However, the encoding rules and all the character components laid out on a standard QWERTY keyboard are difficult to learn.

---

[1] http://en.wikipedia.org/wiki/Wubi_method
[2] http://en.wikipedia.org/wiki/Cangjie_method
[3] http://en.wikipedia.org/wiki/Dayi_method

The CKC Chinese input method is based on the character structure as well but it only uses the 10 digits for encoding. It makes use of the characteristic of the square shape of the Chinese characters and partitions the characters into 4 parts: upper left corner, upper right corner, lower left corner and lower right corner. Each part is represented by a digit that means CKC uses up to 4 digits to represent a character. Users do not need to know how to pronounce it or even how to write it correctly because each character is only perceived as a picture that is divided into 4 parts.

CKC Character Encoding Rules

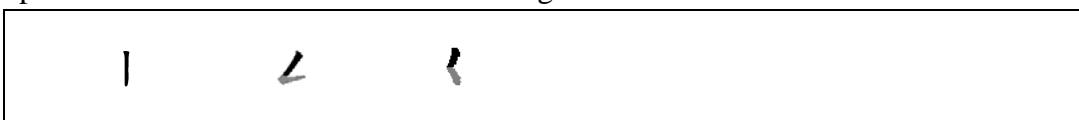The strokes or the shapes of the Chinese characters components are classified into 10 groups.

"0" represents a left-falling stroke or a stroke with left-turning.

"1" represents a horizontal stroke or a horizontal rising stroke written from left to right.

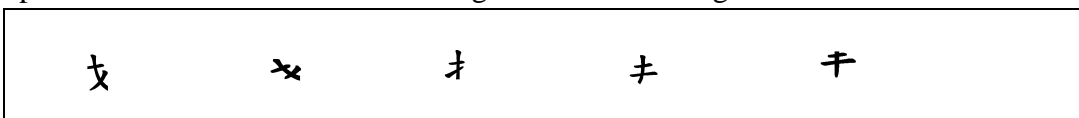"2" represents a vertical stroke or a left slanting stroke.

"3" represents a dot or a right-falling stroke.

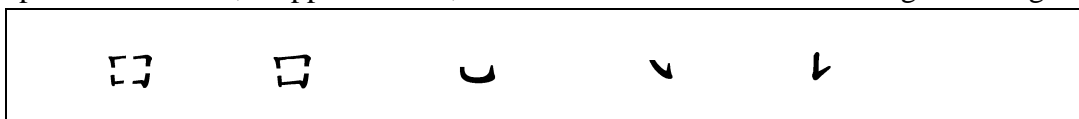"4" represents a cross or the radical of "Grass".

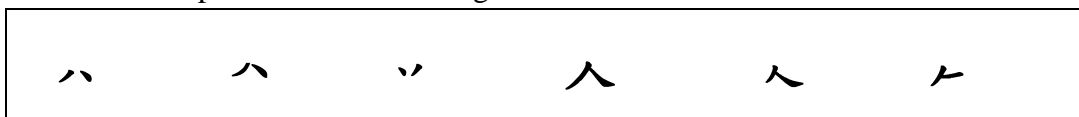"5" represents the combination of a falling stroke runs through 2 horizontal strokes.

"6" represents a square.

"7" represents a corner, 2 upper corners, 2 lower corners or a stroke with right-turning.
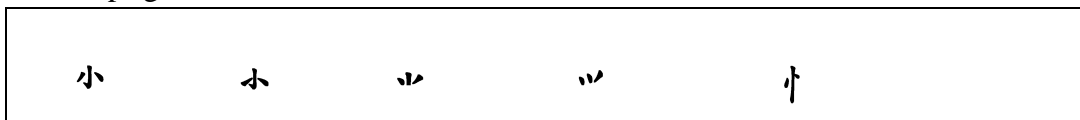

"8" refers to the shape of the numeral "Eight" in Chinese or its variants.

"9" represents the shape of the Chinese character which means "Small", its variants and the

radical of "Upright Heart".

|				
小	小	ㅛ	ㅛ	忄

For easy memorization, users can use the following rhyme to remember the relationship between character strokes and their respective codes:

一横二竖三点捺，
叉四插五方块六，
七角八八九是小，
撇与左钩都是零。

The main steps of composing a CKC code to form a single Chinese character are:

1. Take the stroke of the upper left hand corner as the 1st code.
2. Take the stroke of the upper right hand corner as the 2nd code.
3. Take the stroke of the lower left hand corner as the 3rd code.
4. Take the stroke of the lower right hand corner as the 4th code.
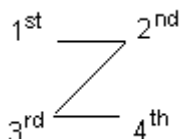
For example, the character of "教", the 1st code is "4", the 2nd code is "8", the 3rd code is "0", the 4th code is "4". The CKC code is "4804".

Sometimes, the CKC code may not necessarily be 4 digits. For example, the character of "大", the 1st code is "4". Since the stroke "cross" represents both the upper left and upper right corners so it does not need the 2nd code. This rule also applies to the lower left and lower right corners. The stroke "Eight" represents them both, thus only 1 digit "8" is needed. For this character the CKC code is therefore "48".

CKC Online Dictionary

The CKC Online Dictionary is designed for users who are not familiar with Chinese. It provides English definitions so users will not be hindered by the Chinese definitions which they are not so familiar with. The Putonghua and Cantonese pronunciation demonstrations can help users learn how to pronounce. The information on spelling, radical and stroke index can help users to look up more details from traditional dictionaries.

If users are not familiar with the CKC code, they can copy the characters from the screen into the text box of the Online Dictionary. The CKC code will then be shown so that the users can learn them. Users can also use the "wild card" key ("Num Lock" on the numeric pad) to represent the uncertain code and then select the desired character. This also makes the lookup possible.
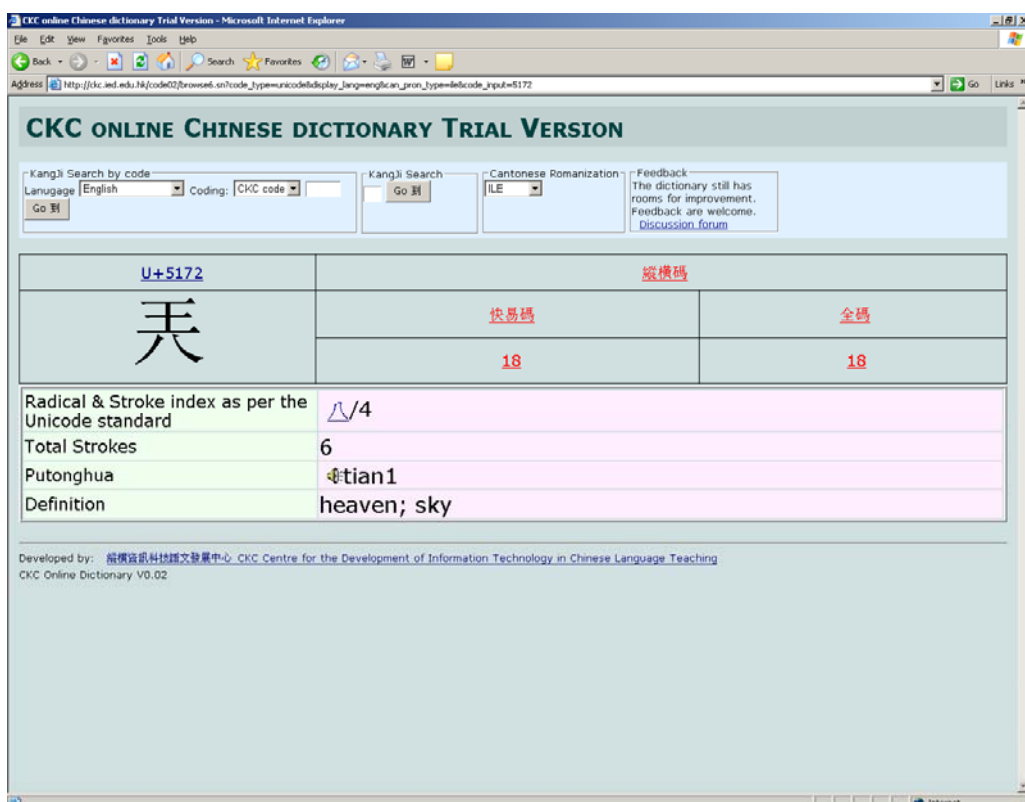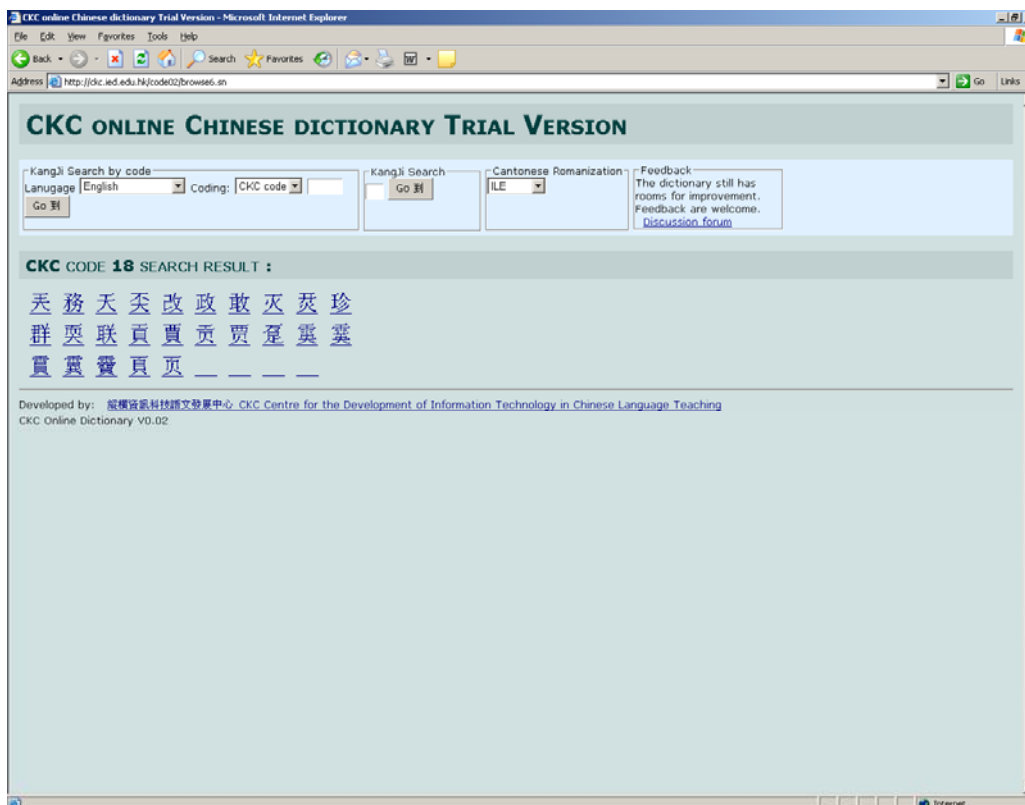
CKC Phrase Encoding Rules

CKC Input System provides the function of phrase input. Indeed Chinese input is based on phrase as an entity. This is like typing a word in English. In English, we think of a word as an entity and we type it on the keyboard without breaking it down to alphabets. For example, for the word "today", we type the 5 alphabets naturally without thinking of the individual alphabets that forms the word. Though some Chinese input methods provide the association

characters, by typing the first character, a list of association characters will appear for selection. However, this kind of method is inconvenient and inefficient. While CKC provides this function, it also allows the input of phrases to enhance the efficiency. For users who are not familiar with Chinese, this function can consolidate their memorization of phrases. To input an arbitrary phrase, we only need to specify up to 6 digits. The principle of encoding a phrase is to iteratively apply the encoding rule for each individual character that forms the phrase, from the 1st till the 5th character of the phrase. The rest of the phrase do not contribute nor affect the formation of its CKC code and therefore can be ignored. This limits the number of keystrokes required to input any arbitrary phrase to 6, saves lots of keystrokes and time. There are 4 variations of rule that can be applied to derive the corresponding CKC code:

2-character phrase

Get the first 3 digits at most of each character. This is known as the "3 + 3" rule. The minimum number of digit of this phrase is 2 and the maximum is 6, it is because a minimum of 1 digit can represent a single character. For example the phrase "一日" requires only 2 digits , that is "16". After entering the CKC code, users have to complete the input with the key "asterisk". If a phrase has the maximum of 6 digits, the completion key can be saved.

^{1}一 ^{6}日

3-character phrase

Get the first 2 digits at most of each character. This is known as the "2 + 2 + 2" rule. The minimum number of digit of this phrase is 3 and the maximum is 6. For the phrase "加拿大" the code is "468048".

$^{4}_{0}$加^{6} $^{8}_{0}$拿^{8} $_{8}^{4}$大

4-character phrase

Get the first 2 digits at most of the 1st character, then the first digit of the 2nd and the 3rd characters. Finally get the first 2 digits at most of the 4th character. This is known as the "2 + 1 + 1 + 2" rule. The minimum number of digit of this phrase is 4 and the maximum is 6. For the phrase "身体健康", the code is "070030".

$^{0}_{0}$身$^{7}_{0}$ $^{0}_{2}$体$^{4}_{4}$ $^{0}_{2}$健$^{5}_{4}$ $^{3}_{0}$康$_{9}$

Multi-character Phrase

Any phrase which is composed of more than 4 characters is considered as a multi-character phrase. Get the first 2 digits at most of the 1st character, then the 1st digit of the 2nd, 3rd, 4th and 5th characters, the rest of the phrase are ignored. This is known as the "2 + 1 + 1 + 1 + 1" rule. The minimum number of digit of this phrase is 5 and the maximum is 6. For the phrase "纵横汉字输入法" the code is "284334".

纵横汉字输入法

With the help of CKC Input System and its Online Dictionary, we hope learning and teaching Chinese will become easier.