

214573

M&C FOLIO: 230P49182X

WANGDOC: 0841R

ENCODING CHINESE CHARACTERS AND KEYBOARD THEREFOR

The invention relates to a system for encoding Chinese characters by means of which Chinese characters can be fed into an electronic computer, "Telex" (RTM) or teleprinter machine, or the like, ~~and the system can also be used to index a Chinese directory.~~

Up to now, there are more than 400 methods for encoding Chinese characters, including the method of encoding Chinese characters in numerical notation described and claimed in United Kingdom Patent No. 2 100 899, but only a few are workable. Generally speaking, there are five categories of systems for encoding and inputting Chinese characters.

1. Phonetic alphabet system

This kind of system is only applicable to professional computer operators who have a good command of Chinese alphabetic spelling. Only when the alphabetic spelling is 100 percent correct, can the character be identified. Since there are many homonyms in Chinese characters, it is too hard or even impossible for a person who does not know Chinese alphabetic spelling or only has scant knowledge thereof.

2. Large keyboard system

The whole character is fed into computer by only one

key. Only those characters which are on the keyboard can be fed into the computer. It needs a very large keyboard which occupies a large space. Furthermore, it takes a long time to get familiar with the keyboard. A person has no interest in learning if he is not specifically required to do so and nobody can use the keyboard to type or feed into the computer without looking at the keyboard.

3. Parts system

The Chinese character is divided into several parts to one part of which a code (or digit) will be given and then can be fed into computer. The advantage of this system is that the Chinese character which can be easily divided can be fed into computer quickly, for example, "日+月=明" and "王+王=狂". Such feeding depends on a 100 percent correct division of the character, otherwise the character cannot be identified. However, it is almost impossible to divide some characters, for example "翰.蒂.高", into parts, particularly for those who are not conversant with Chinese characters; furthermore, as there is no strict rule for such division and the division of such a character varies from person to person, it is not easy to make the correct division.

4. Number code system

This is a better system which is relatively easy to learn. But it usually involves a long code and some

characters may have the same codes. The procedure of encoding under this system is first to divide a character into several parts and then give each part a digit. Since there is no strict rule in dividing a character into parts and a stroke may be given a different digit in different characters, people often hesitate when encoding a character.

5. Word, phrase and sentence system

Under this system Chinese characters compose 200,000 phrases and sentences, and the whole phrase or whole sentence is fed into a computer. This system is only suitable for a computer which has good performance and a large memory.

To sum up, any one of the above five systems has its own defect. It is too difficult to divide Chinese characters into parts and encode them. In addition, it is very inconvenient to encode a Chinese character from stroke to stroke, and the code is quite long. Most Chinese characters consist of a leading part and of a body part or strokes. How to divide them is very difficult.

It is desirable to provide a simple encoding system for Chinese characters and in particularly, a system whereby the characters can be encoded in a form suitable for simple and quick entry into an electronic computer.

The present invention provides methods as set forth in Claims 1 and ¹⁶~~17~~, keyboards as set forth in Claims 10

TR

¹⁷
and ~~18~~, ~~indexes for Chinese dictionaries as set forth in~~
~~Claims 16 and 19~~, and a typewriter, computer, "Telex"
(RTM), or telecopier, as set forth in Claim ¹⁸~~20~~. The
remaining Claims set forth optional features of the
invention.

Thus each Chinese character is considered as a plain figure which can be divided into two parts, a leading part and a body part. The leading part can be represented by one code element and the body part can be represented by up to four code elements (digits). The code element for the leading part is first fed into the computer, and then the code elements for body part, so that the whole Chinese character can be fed into the computer.

Since the encoding system of the present invention is not based on the phonetic alphabet or on the order of writing the strokes by hand, it is especially suitable for those who are not familiar with Chinese characters. By means of the method of the present invention, anyone can feed Chinese characters into ^acomputer without difficulties, no matter whether he or she knows the phonetic alphabet and the hand writing rules for Chinese characters or not.

The keyboard of the present invention can be used for Chinese characters and English, and it can be applicable anywhere in the world. Besides, one can use it to typewrite without looking at the keyboard.

TJ

5

Brief description of the attached drawings is following:

Figure 1 shows a keyboard arranged by the encoding system of the present invention.

5 Figure 2 shows the relation between 10 basic strokes and the corresponding keys on the keyboard of the encoding system of this invention.

Figure 3 shows a index example of the Chinese dictionary compiled upon the encoding system of this 10 invention.

~~According to the present invention~~ There are 82 leading parts which are arranged on 35 keys. ^{Three leading parts are arranged on} ~~On each~~ key in ^{the} second line ~~3 leading parts are arranged~~ (see Figure 1) and ^{two leading parts} ~~on each key in~~ ^{the} third and fifth lines, ~~2~~ ¹⁵ ~~leading parts.~~ With the help of ~~leading part~~ keys the operator can easily divide a Chinese character into leading part and body part and does not need to think how to divide them ^{which will} ~~so as to~~ decrease hesitation during the division of the leading part and the body part.
 20 ~~decrease~~ memory. decrease consideration pressure and increase encoding velocity.

For the body part there are 10 stroke shapes 八, 乚, 丩, ㄥ, ㄣ, ㄤ, ㄨ, ㄩ, ㄚ, ㄛ which are represented respective by 1. 2. 3. 4. 5. 6. 7. 8. 9. 0 (i.e. key A. S. D. F. G. H. J. K. 25 L. +). Since for most ~~of~~ Chinese characters it is

necessary to push 3 or 4 times body-part-keys. therefore
it is proper to arrange the body-part-keys in the
positions which are frequently used so that the moving
frequency of the fingers can be decreased and encoding
5 velocity can be increased.

During encoding, at first one can divide a Chinese
character into a leading part and a body part. find the
corresponding leading-part-key on the keyboard and press
it. then the code representing ^{the} leading part is fed into
10 computer. For example, if someone wants to feed the
Chinese character "軀" into ^a computer. at first divide
"軀" into "身" and "區". The "身" is the leading
part. On the keyboard he can find the leading-part-key
for leading part "身" and then press it. the leading
15 part "身" is now fed into computer. The lower
part and the right part of a character, however, can not
be called the leading part. thus these parts can not be
fed in by leading-part-keys. By way of example, the
character "郅" has a part "卩" on the right. Although
20 the "卩" part has a key on the keyboard representing the
same as a leading part. the "卩" part of the character
can not be inputted by that key. The character "郅"
thus can only be treated as a connected ^(see below) character.
Another example is "型". The "型" part can not

7

be inputted by the same leading part key. This character "型" can only be inputted as a connected character (see below).

After the leading part is fed into computer, you
 5 can turn to encode the body part of the character. A ~~The~~ ^{is used to input} ~~maximum of four~~ ~~times for extracting strokes from~~ the body part, ~~is at~~
~~most 4.~~ The basic rule for taking strokes is upper
 level first, lower second. For example, the character "証"
 " its leading part is "讠" and first four
 10 strokes of body part are "一", "丨", "丨" and "一"
 corresponding code element (digit) 7, 3, 3 and 7, so the
 code of the character "証" is 讠7337. Another example,
 the Chinese character "订" can be divided into the
 leading part "讠" and the body part "丁". The
 15 strokes of this body part are "一" and "丨"
 corresponding digit 7 and 3, so the code for "订" is
 讠73.

But there are some of Chinese characters whose
 body part can be divided into left part and right part.
 20 For those characters there is another rule for taking
 strokes, left first, right second. By way of example,
 the body part "对" of the character "树" can be
 divided into "又" and "寸". In encoding it is

first to take the strokes of left part "又" and second the strokes of right part "寸" with the rule upper first, lower second.

For some of the characters which can be divided into outside part and inside part the rule for taking strokes is outside first, inside second. Of course for the inside sub-part or outside sub-part it is still the rule upper first, lower second. For example, the character "函" which has the body part "函".
 10 its outside part is "凵", inside part is "承". the first four strokes of the body part are "L", "I", "7" and "J".

The connected characters are those characters which have no leading parts. So the code for such characters should be taken ^{in a number of separate} ~~by stroke~~. The times
 15 of strokes taken from connected characters is at most five. The rule for ^{encoding} ~~taking code for~~ connected characters is just the same as the rule for the body part of characters. For example, the code for character "义" is 656, the code for the character "有" is 07547.
 20

The keyboard of the present invention can be the normal English keyboard. see figure 1. The first line

of keys serves as the function-key by means of which feeding mode can be selected. For example. F1 represents the feeding mode of the encoding system according to the present invention. F2 represents the feeding mode of the ^{phonetic alphabet} ~~Phonetically Al~~phabetic system, and so on.

The second, third and fifth line in the keyboard serve as "leading-part-key". 82 basic leading parts are arranged on 35 keys.

10 ^{the} In ^{the}fourth line, first 10 keys from the left serves as "body-part-key" each of which corresponds two similar strokes, see figure 2.

15 The lowest bar-key on the keyboard serves as feeding key which is used after encoding. Of course the key is also the space key.

Furthermore the last two lines on the bottom of the screen of the computer ~~are acted as a~~ "note-window" which show 22 Chinese characters and their codes in two lines. The computer operator can see Chinese characters and their codes on the screen so that he can learn the code through looking at the "note-window" and he need not look for the code of some hard-to-encode characters in the code hand-book.

25 By way of example, we want to feed ^{into a computer} the character "客" ~~into computer~~ whose leading part is "宀".

First: press leading part-key "宀 土", 22 Chinese characters which have leading part "宀" or "土" are shown on the "note-window. most of them have a code composed of 4 digits.

5 案2759 安275 宝7967 宾5373 系3866 宅527 穿1732 窗1545 定7357 富1007 输入键
埃2657 坝275 埠5337 场8835 城2675 赤531 堤4773 地328 堵9574 堆3309 宀土-

Since the body part of "客" is "各" the first stroke of which is "丿" corresponding^{to} digit 5. so the second step is pressing the body-part-key "5". ★ Some of the characters, which have^{the} first stroke "丿" in^{the} body part ~~among the characters which~~ ^{and the} have^{the} leading part "宀" or "土", are shown on the "note-window. These characters have left at most 3 digit code.

15 宾373 系866 宠672 宏726 客864 蜜562 宿375 宛868 宅27 允72 输入码
埠837 赤31 堆309 垛729 垢574 郝318 赫319 均867 坎81 志266 5-5

Since the second stroke of body part "各" is "乚" corresponding^{to} digit 8. the third step is to press the body part-key "8". A number of characters which have^{the} second stroke "乚" in its body part are shows on 20 note-window. These characters have left at most 2 digits.

案66	客64	宛68		输入码
埠37	均67	坎1	坞7	5-58

25 Since the third stroke of the body part "各" is "㇏", corresponding^{to} digit 6. the fourth step is to press body-part-key "6". A number of characters which have^{the} third stroke "㇏" in body part are shown on the "note-window".

素6	客4	宛8	输入码
均7			3-586

The fourth stroke of the body part, "口", is corresponding to digit 4.

- 5 Fifth step: press the body part key "4". The character "客" is the one that remains on the "note-window"; by pressing the bar key, then one can feed the character "客" into the computer.

客	输入码
	3-5864

In the same way, the Chinese character can be fed into the computer by means of phonetically alphabetic system.

For example, feeding the character "肥" which has phonetic alphabetic "Fei".

First step: pressing the key F, then a number of characters which have the phonetical Alphabet F will be shown on the "note-window". The area from the far right shows the Alphabet fed into computer.

20	发a	罚a	伐a	乏a	法a	反an	返an	范an	帆an	翻an	拼音	输入码
	菲ei	肥ei	翡ei	废ei	费ei	丰eng	封eng	锋eng	风eng	冯eng	f-	

Second step: pressing the key "e". The "note-window" then will show many characters.

25	菲i	非i	啡i	飞i	肥i	匪i	诽i	吠i	肺i	废i	拼音	输入码
	费n	芬n	酚n	吩n	氛n	分n	纷n	坟n	焚n	汾n	f-e	

Third step: pressing the key "i". All characters which have the same phonetical Alphabet "Fei" and their notes are shown on the "note-window". You can now select "肥" from them.

5

-	-	-	-	/	✓	✓	-	-	声调 拼音 输入码
菲	非	啡	飞	肥	匪	诽	肺	废	请选择 f-ei

Another advantage of the present invention is that there ^{can be} ~~are~~ two or more codes for one character. This advantage brings much convenience to the operator. If one character has only one code, it will depend on 100% correct encoding and feeding. Otherwise one cannot find the character. If there are two or more codes for one character, it will make the encoding much ^{easier} ~~more easy~~. For example, the character "磨" has two codes, Y0591 and YW1 through which the character "磨" can be fed into the computer, so that the feeding ^{velocity} ~~velocity~~ and efficiency will be improved.

This system for encoding Chinese characters can also be used for indexing a Chinese dictionary. Under this system, characters of the same code are put on the same page of the dictionary. This is a new development of indexing a Chinese character. Under this system the leading part code and body part code are used as the

13
page number in a Chinese dictionary. ~~the~~ ^{and} ^{ing} ^{for} ~~search~~ ~~of~~ ~~words~~
in such ^a dictionary becomes ~~more~~ simpler

For example, we want to search ^{for} the character
" 漁 ". The procedures for search ^{for} the character
5 " 漁 " are:

(1) to determine the leading part code first.
From figure 3 you can find the code for " 艹 " is 17
which corresponds the page 17 in ^{the} dictionary.

(2) to determine the body part code. In this case
10 the body part is " 魚 ". The first stroke of " 魚 " is
" 丿 ", corresponding ^{to} digit 5.

(3) So character " 漁 " should be on page "17-5"
(see the following chart).

15	艹 17-1	浴沦浹淡淫泔液浹治淪淪汰改潜淫漆塗涂淑淪淪淪
	艹 17-2	溜潍泗涵浹浹治浹浹浹浹浹浹浹浹浹浹浹浹浹浹浹
	艹 17-3	淌消沙染染染染染染染染染染染染染染染染染染染染
	艹 17-4	泗漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚漚
20	丿 17-5	沃塗漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆漆
	丿 17-6	洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞洞
	一 17-7	汇沅汛渠汀沔漂漂漂漂漂漂漂漂漂漂漂漂漂漂漂漂漂
25	冫 17-8	汜泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃泃
	十 17-9	汁漬泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔泔
	一 17-0	滂滴泣洁漳漳漳漳漳漳漳漳漳漳漳漳漳漳漳漳漳漳漳

15

Claims

1. A method of encoding and inputting Chinese characters which have just a leading part or are divisible into a leading part and a body part, using inputting means with a glossary of different leading parts at least some of which have more than one stroke, in which method, when the glossary holds the leading part of a character, the leading part is input as a whole and the body part, if any, is input as one or more separate strokes, and, when the glossary does not hold the leading part of a character, the leading part and the body part, if any, are input as separate strokes.

2. The method of Claim 1, in which the leading part is input before the body part.

3. The method of Claim 1 or 2, in which the leading part is input in one keying operation and the body part is input by up to four keying operations.

4. The method of Claim 3, in which the glossary holds eighty-two leading parts, which are basic leading parts extracted from all Chinese characters.

5. The method of any of the preceding Claims, wherein there is a glossary of ten basic strokes for encoding



15

and inputting the body part.

6. The method of Claim 5, wherein the strokes of the body part are taken in the order required by the following rules: upper first, lower second; if the body part can be divided into a left-hand part and a right-hand part, or an outside part and an inside part, left first, right second or outside first, inside second; and if a part of the body part can be divided into sub-parts, in each sub-part upper first, lower second.

7. The method of any of the preceding Claims, when used with a machine having a screen, a note-window being provided on two lines on the bottom of the screen.

8. The method of any of the preceding Claims, wherein^{when} the glossary does not hold the leading part of a character, ~~and~~ the character is input by encoding up to five strokes of the character using the same rule as for encoding the body part of a character whose leading part is held by the glossary.

9. The method of any of the preceding Claims, wherein a character can be encoded in at least two different ways.

10. A keyboard for inputting Chinese characters by the



* C L N S *

TD

method of any of the preceding Claims, the keyboard being ~~based on~~ a Roman-alphabet typewriter keyboard and having a number of keys marked with leading parts of Chinese characters, at least some of which leading parts have more than one stroke, and a number of keys marked with individual strokes of Chinese characters.

11. The keyboard of Claim 10, in which thirty-five keys in three lines of the keyboard represent eighty-two basic leading parts, two or three leading parts being input by pressing the same leading part key.

12. The keyboard of Claim 11, and having a row of function keys for selecting different input modes.

13. The keyboard of Claim 12, wherein the same key can be used to input a Roman character or a leading part or a stroke, by operation of a function key.

14. The keyboard of any of Claims 10 to 13, wherein there is a row of body part stroke keys in a position where the hands of the operator usually rest.

15. The keyboard of any of Claims 10 to 14, wherein, after encoding, the Chinese character is automatically fed into a machine connected to the keyboard by actuating a spacer key once.



FF

M&C FOLIO: 230P49182X

WANGDOC: 0886R

16. A method of encoding and inputting Chinese characters, substantially as herein described with reference to Figures 1 and 2 of the accompanying drawings.

17. A keyboard, substantially as herein described with reference to, and as shown in, Figure 1 of the accompanying drawings.

18. A typewriter, computer, "Telex" (RTM), or telecopier, arranged to be input with Chinese characters by the method of any of Claims 1 to 9 and 16.

