

(12) **UK Patent Application** (19) **GB** (11) **2 166 573 A**

(43) Application published 8 May 1986

(21) Application No **8526483**
 (22) Date of filing **28 Oct 1985**
 (30) Priority data
 (31) **8427281** (32) **29 Oct 1984** (33) **GB**

(71) Applicant
Wong Kam-Fu,
22nd Floor, Wu Sang House, 655 Nathan Road, Kowloon,
Hong Kong

(72) Inventor
Wong Kam-Fu

(74) Agent and/or Address for Service
Marks & Clerk,
57-60 Lincoln's Inn Fields, London WC2A 3LS

(51) INT CL⁴
H03M 11/00 B41J 5/10

(52) Domestic classification
G4H KJ
B6F CB

(56) Documents cited
GB A 2120977 **GB A 2089730** **GB A 2030335**
GB A 2118749 **GB A 2071018** **US 4327421**
GB A 2100899

(58) Field of search
G4H
B6F
Selected US specifications from IPC sub-classes B41J
G06F G08C

(54) **Encoding chinese characters**

(57) A chinese character to be keyed in is divided into a "leading part" and a "body part". The leading part is keyed in as a whole. The body part is keyed in in terms of its constituent strokes. The keyboard has function keys, the leading part keys and the body part stroke keys. The body part stroke keys are in the positions which are used most frequently. The invention can also be used for indexing a chinese dictionary.

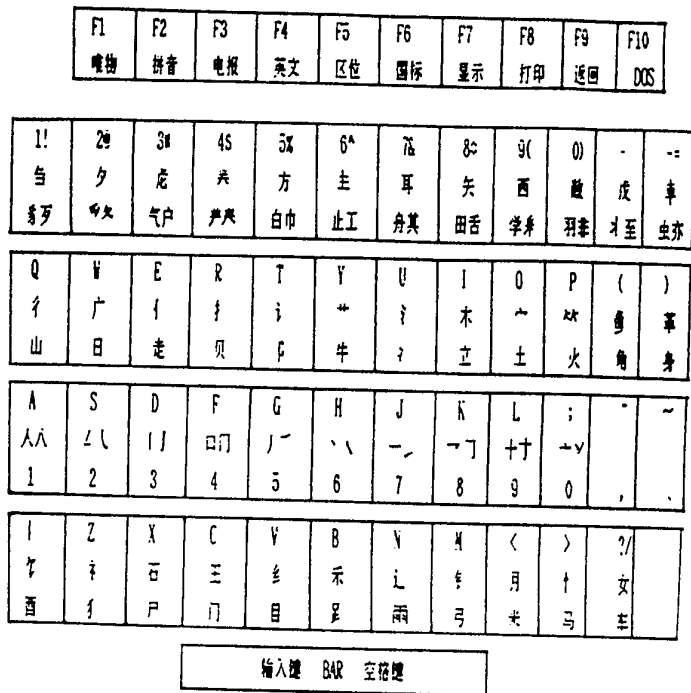


FIG. 1

GB 2 166 573 A

The drawing(s) originally filed was/were informal and the print here reproduced is taken from a later filed formal copy. This print takes account of replacement documents submitted after the date of filing to enable the application to comply with the formal requirements of the Patents Rules 1982.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
唯物	拼音	电报	英文	区位	国标	显示	打印	返回	DOS

1!	2@	3#	4\$	5%	6^	7&	8:	9(0)	-	=
与	夕	虎	夫	方	主	耳	矢	西	被	戊	丰
豸	夕	气	产	白	止	舟	田	学	羽	才	虫

Q	W	E	R	T	Y	U	I	O	P	()
彳	广	彳	扌	讠	艹	彳	木	宀	火	鱼	革
山	日	走	贝	冫	牛	彳	立	土	火	角	身

A	S	D	F	G	H	J	K	L	;	'	~
人	厶	冂	口	冂	丶	一	冂	十	彳	,	.
1	2	3	4	5	6	7	8	9	0	,	.

I	Z	X	C	V	B	N	M	<	>	?	/
彳	彳	石	王	彳	禾	辶	彳	月	彳	文	车
酉	彳	尸	冂	目	足	雨	弓	米	马	车	

输入键 BAR 空格键

FIG. 1

2166573

2/3

stroke	^	L		□	'	\	-	→	+	÷
	八	∪	↓	▣	/	∪	↙	↘	+	∪
corresponding digit	1	2	3	4	5	6	7	8	9	0
key in the keyboard	A	S	D	F	G	H	J	K	L	;

FIG.2

leading part, page	leading part, page	leading part, page	leading part, page	leading part, page
人八 1-	山 11-	衤 21-	鱼角 31-	矢田舌 41-
ㄥ 2-	广日 12-	石尸 22-	革身 32-	西学糸 42-
丨 3-	走 13-	王门 23-	酉 33-	敝羽非 43-
口 4-	貝 14-	纟目 24-	自豸歹 34-	戊才至 44-
ノ 5-	讠 15-	禾足 25-	夕 35-	幸虫亦 45-
、 6-	**牛 16-	辶雨 26-	虎气户 36-	
- 7-	讠 17-	钅弓 27-	关 37-	
ㄗ 8-	木立 18-	月米 28-	方白巾 38-	
十 9-	土 19-	竹马 29-	圭止工 39-	
ㄥ 0-	**火 20-	女车 30-	耳舟其 40-	

FIG. 3

SPECIFICATION

Materialistic system for encoding Chinese characters and the feeding keyboard thereof

- 5 The invention relates to a system for encoding Chinese characters by means of which Chinese characters can be fed into an electronic computer, telex or teleprinter machine or the like, and the system can also be used to index a Chinese directory. 5
- 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. 10
1. Phonetically Alphabetic 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. 15
2. Large keyboard system
- Under this system, the whole character should be 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 to learn if he is not intended to and nobody can use it to type or feed into computer without looking at it. 20
3. Parts system
- Under this 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, 王+王=狂". Such feeding depends on 100 per cent correct division of the character, otherwise the character can not be identified. As for some characters, for example "输. 莽. 高", it is even almost impossible to divide them into parts, particularly for those who are not conversant with Chinese characters. And more difficult thing is that there is strict rule for such division and the division of a character varies from person to person, it is not easy to make a correct division for a character. 25
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 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, so it makes people often hesitate in taking code of a character. 30
5. Word, phrase and sentence system
- This encoding system has high logic of Chinese character. Under this system Chinese characters compose 200,000 phrase and sentence and the whole phrase or whole sentence is fed into computer. This system is only suitable to a computer which has good performance and large store space. 35
- To sum up, we can see that any one of system among above mentioned 5 systems has its own proper 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. The Chinese character consists of leading part and body part or strokes. How to divide them is very difficult. 40
- It is therefore an object of the present invention to provide a simple encoding system for Chinese characters and in particularly, a system whereby the character can be encoded in a form suitable for simple and quick entry into a electronic computer. 45
- According to the present invention, 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 is represented by one code element and the body part is represented by four code elements 50
- 55

(digits). The code element for the leading part is first fed into computer, and then the code elements for body part, so that the whole Chinese character can be fed into computer.

5 Since the encoding system of the present invention is not based on Phonetical Alphabet or stroke order of hand writing, it is especially suitable for those who are not familiar with Chinese characters. By means of the encoding method of the present invention anyone can feed Chinese characters into computer without difficulties no matter whether he or she knows phonetical Alphabet and the Chinese hand writing rules for Chinese character or not. 5

The keyboard of the present invention can be used for Chinese character and English, it can be applicable anywhere in the world. Besides, one can use it to typewrite without looking at it. 10 Brief description of the attached drawings is following: 10

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

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 invention. 15 of this invention. 15

According to the present invention there are 82 leading parts which are arranged on 35 keys. On each key in second line 3 leading parts are arranged (see Fig. 1) and on each key in third and fifth line 2 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 so as to decrease hesitation during the division of the leading part and the body part, decrease memory, decrease consideration pressure and increase encoding velocity. 20

For the body part there are 10 stroke shapes 丿, 丨, 丨, 口, 丨, 丨, 丨, 丨, 丨, 丨 which are represented respectively by 1, 2, 3, 4, 5, 6, 7, 8, 9, 0 (i.e. key A, S, D, F, G, H, J, K, L, +). Since for most of Chinese character 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 velocity can be increased. 25

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 leading part is fed into computer. For example, if someone wants to feed the Chinese character "軀" into 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 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 "部" has a part "阝" on the right. Although the "阝" part has a key on the keyboard representing the same as a leading part, the "阝" part of the character can not be inputted by that key. The character "部" thus can only be treated as a connected character. Another example is "塑". The "塑" part can not be inputted by the same leading part key. This character "塑" can only be inputted as a connected character. 30 35

40 After the leading part is fed into computer, you can turn to encode the body part of the character. The times for extracting stroke 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 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 7337. 40

45 Another example, the Chinese character "订" can be divided into the leading part "讠" and the body part "丁". The strokes of this body part are "一" and "丨" corresponding digit 7 and 3, so the code for "订" is 73. 45

But there are some of Chinese characters whose body part can be divided into left part and right part. 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 "对" 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. 50

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 "法" which has the 55

body part “**厶**”, its outside part is “**厶**”, inside part is “**厶**”, the first four strokes of the body part are “**厶**”, “**厶**”, “**厶**” and “**厶**”.

The connected characters are those characters which have no leading parts. So the code for such characters should be taken stroke by stroke. The times of strokes taken from connected characters is at most five. The rule for 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 “**着**” is 07547.

The keyboard of the present invention can be the normal English keyboard. see Fig. 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 Phonetically Alphabetic 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.

In fourth line, first 10 keys from the left serves as “body-part-key” each of which corresponds two similar strokes, see Fig. 2.

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 “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.

By way of example, we want to feed the character “**客**” into computer whose leading part is “**厶**”.

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

30

案2739 安275 室7967 宾5373 系5866 宅527 穿1732 窗1545 定7357 富1007 输入键
埃2637 坝275 埠5837 场6855 坡2675 赤531 堤4773 地328 堵9574 堆5309 厶土

 30

Since the body part of “**客**” is “**各**” the first stroke of which is “**厶**” corresponding digit 5, so the second step is pressing the body-part-key “5”. A some of the characters, which have first stroke “**厶**” in body part among the characters which have leading part “**厶**” or “**厶**”, are shown on the “note-window. These characters have left at most 3 digit code.

40

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

 40

Since the second stroke of body part “**各**” is “**厶**” corresponding digit 8, the third step is to press the body part-key “8”. A number of characters which have second stroke “**厶**” in its body part are shows on note-window. These characters have left at most 2 digits.

45

系66 客64 宛68	输入码
埠37 均67 坎1 坞7	厶-8

 45

50 Since the third stroke of the body part “**各**” is “**厶**” corresponding digit 6, the forth step is to press body-part-key “6”. A number of characters which have third stroke “**厶**” in body part are shown on the “note-window”.

索6 客4 苑8	输入码
均7	5-386

5

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

5

10

客	输入码
	5-3864

10

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

15

For example, feeding the character "肥" which has phonetical 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 霏ei 废ei 费ei 丰eng 封eng 缝eng 风eng 冯eng	f-

20

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 诹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

30

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

30

35

· · · · ·	声调 拼音输入码
菲 非 啡 飞 肥 匪 诹 吠 肺 废	请选择 f-ei

35

Another advantage of the present invention is that there 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 more easy. For example, the character "麈" has two codes. Y0591 and YW1 through which the character "麈" can be fed into computer, so that the feeding volocity and efficiency will be improved.

40

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 page number in Chinese dictionary, the search of words in such dictionary becomes more simple.

45

For example, we want to search the character "麈". The procedures for search the character "麈" are:

50

(1) to determine the leading part code first. From Fig. 3 you can find the code for "彡" is 17 which corresponds the page 17 in dictionary.

(2) to determine the body part code. In this case the body part is "麈". The first stroke of "麈" is "ノ" corresponding digit 5.

55

characters into a machine having a screen, wherein a "note-window" is provided in two lines on the bottom of the screen.

9. The method, as claimed in any preceding claim, wherein connected characters are input by taking and encoding at most five strokes from the character in the same rule as to the body part.

10. The method as claimed in any preceding claim, wherein a character can be encoded in many different ways.

11. A dictionary of Chinese characters in which the characters have been encoded by a method as claimed in any preceding claim and the order of the codes in the dictionary is such that all those starting with the same digit have been grouped together and sub-divided.

12. A keyboard through which the encoded Chinese character can be input, characterized in that the keyboard is based on an ordinary Roman letter typewriter keyboard.

13. The keyboard, as claimed in Claim 12, in which thirty five keys in three lines of the keyboard represent the eighty two basic leading parts, two or three leading parts will be input by pressing one leading part key.

14. The keyboard, as claimed in Claim 13, wherein a row of function keys are provided on the keyboard for selecting different input modes.

15. The keyboard, as claimed in Claim 14, wherein a row of body part keys are placed in the positions which are in the most frequent use on the keyboard, i.e. where the hands of the operator are usually rested.

16. The keyboard, as claimed in Claim 15, wherein the Roman alphabet along with the leading parts or strokes actuated by the same key can be selected by pressing different function keys.

17. The keyboard, as claimed in Claim 16, in which after encoding, the Chinese character can automatically be fed into computer by pressing the space key one time.

18. The keyboard, as claimed in Claim 17, wherein the keyboard is formed as shown in Fig. 1.