

THE FUTURE OF COMPUTER CHARACTERS

Shi-Zhao Zhang

Meteorological Bureau of Shaanxi Province, Xi'an, 710015 China

Email: chancezoo@126.com; mzsgls@pub.xaonline.com

ABSTRACT

A new method for digitizing letters and Chinese characters is proposed.

Keywords: Language, Character, Unicode, Encoding, Storage, Computer character

The common people consider that language and characters are used for information exchange. In fact they have a more important function; they are very important tools of human thought. Since the birth of humanity, language and character have developed through 4 stages. While each stage greatly enhances the functions of the previous stage, the earlier stages are not replaced. All stages coexist and continue to expand. Table 1 shows the four stages of development.

Stage	Carrier	Characteristic	New Function
Body language	Body movement	Limited to sight Cannot be saved	Information exchange
Spoken language	Air vibration	Limited by hearing Cannot be saved	Tool of thought
Written character	Two-dimensional figure	Limited by sight, can be transmitted and saved	Potent Thought Promotion
Computer Character	Electromagnetism	Can be extensively duplicated and transmitted very quickly	Thought Mechanization

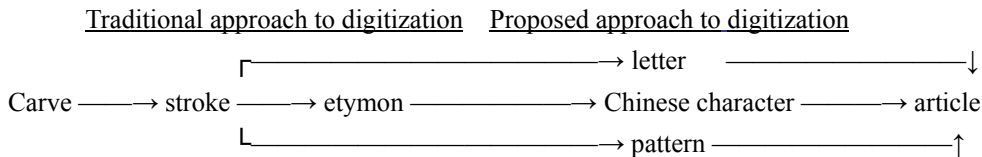
Table 1. The four stages in language and character development

The written character stage at first was the stage in which characters were written with a pen on paper. During the printing age, a large number of characters were printed on paper with type. This situation allowed the character to be stored and spread abroad, but it loses the function of individual signatures and calligraphy. Mainstream characters in the information age have become an invisible state of electromagnetism, and we must use monitors to read them. Computer characters are replications of printed characters; in other words, the method of "storage and code" is used, where different types correspond to different codes. The shortcomings of printed characters, however, remain as before. Furthermore, reading printing characters does not need a type, but reading computer characters must use whole types (character storage) and will continue into the future. For languages that assemble words by combining individual letters, printing does not require any understanding of the meaning of the printed characters. For such languages, computer storage of the word simply requires storage of the individual letters in correct sequence. However, in languages such as Chinese, the way in which the components of a word are associated affects the meaning of that word. Computer storage of words requires both the individual components as well as the "total" word by itself.

Written character: Pen —————→article —————→read
 Printed character: Type —————→article —————→read
 Computer-character: Storage —→ code —→article —————→read←—storage

If there was only one kind of character (such as English) in the world, there would be no problem. However, in Chinese, there are various types of characters. This has always been a problem, but it has become worse in the Internet age. Further, the necessity to use characters from different languages on the same page makes it necessary to use Unicode. Currently improved hardware can easily deal with increased code length and storage. The problem, however, is that there are too many characters, and more are continually developing. It is difficult to update changes and disseminate the new code, resulting in much confusion.

On the other hand, with the increase in the number of characters in storage, it becomes more difficult to choose the needed character. The Chinese language, when entirely put into digital form by a variety of methods (for example, “ten thousand yards gallop”) has characters that cannot be input individually and must simultaneously use many kinds of character, is even more difficult. There are about 100,000 Chinese type characters. Currently 80,000 characters are in storage, but because the quantity is too big, it is easy for the computer to make a mistake and retrieve the wrong character. We have found that the huge Unicode is not efficient and have, thus, begun to use the IDS and CDL methods. These methods construct new characters using old characters. In principle, they need only a few etymons or strokes to do this. However, the method needs additional structural characters and data components. If the length of the character string for a new character is too long, it can not be used in the code.



The system of Chinese without storage that I developed in 1984 only needs 240 etymons without additional characters (only a few Chinese characters need structural characters). The etymon string itself is the code of the Chinese character. In 2003, demonstration software on a PC for Chinese characters was developed; it only used 50 basic strokes to build all the Chinese characters included in any large dictionary. The solution for Chinese on the computer is to build characters with strokes. These same few elemental strokes can construct characters from all over the world. The first test, using 7 elemental strokes had been used to construct all the ASCII characters and many map charts. In fact, the 7 elemental strokes have been put into 7 subroutines. A character is not a picture; it is made up of single lines (strokes). By studying all the strokes of the character of all languages and using advanced software and hardware, we will be able to write all types of characters with a keyboard and not need an electronic pen. This is the future of computer characters.

50 genus strokes



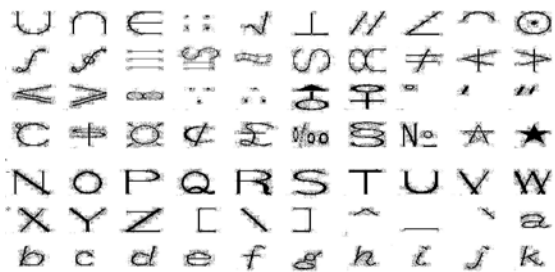
A character from strokes which is not included in any large dictionaries

Character "biang":

**biangbiang
noodle is
refection of
province
Shaanxi**



ASCII characters from 7 element strokes



REFERENCE

Retrieved from the WWW August 16, 2007: <http://www.chancezoo.net>