

COMPUTERISED CODES FOR CHINESE CHARACTERS

There are three Chinese character sets codes for information interchange which so far as we know, have been issued by authorities in Peking, Taipei and Tokyo respectively. The Peking code listed 6,763 Chinese characters, the Taipei code 4,807 and the Tokyo code 6,349.

A. The Peking code entitled "Code of Chinese Graphic Character Set for Information Interchange (Primary set)" was drafted by the North China Institute of Computing Technology and adopted as the National Standard No. GB2312-80 by the State Bureau of standardization and published by the Technical Standards Press in May 1981. Both Chinese and English language editions of this standard are available from Chinese Publications Centre, P.O. Box 399, Beijing, China.

The Standard totals 7,445 graphic characters, including 262 general characters, (the space, punctuation marks, arithmetic signs, unit marks, and tabulation marks), 60 ordinal numbers (20 of them are numbered from 1 to 20, another 20 from (1) to (2), 10 from (1) to (10), and another 10 from (-) to (+), 22 numerical characters (10 of them are numbered from 0 - 9, 12 of them are numbered from I to XII), 52 Latin letters (26 capital letters and 26 small letters), 169 Japanese Kana (82 hiragana and 87 katakana), 48 Greek letters (24 capital letters and 24 small letters), 66 Russian letter (33 Capital letters and 33 small letters) 26 Chinese phonetic symbols, 37 Chinese phonetic annotated letters and 6763 Chinese Characters.

The Standard adopts 2-byte representation for any graphic characters and both the bytes use 7-bit representation in GB1988-80 (For information processing interchange) and GB2311-80 (For code extension techniques).

The code table of this standard is composed of 94 sections and 94 positions. The sections are numbered from 1 to 94 and each of them is identified by the first byte. The positions are also numbered from 1 to 94, and each of them is identified by the second byte. The location of any graphic character in the Code Table is identified

by the combination of its section number and position number.

The 6763 Chinese characters are divided into 2 levels. Level I includes 3755 Chinese characters which are set in sections 16 to 55. Level II includes 3008 Chinese characters which are set in Sections 56 to 65. Sections 1-6, are allocated to all other characters.

The Chinese characters in Level I are arranged in the order of Chinese phonetic letters. Homonyms are subdivided by strokes of Chinese characters in the following order: horizontal stroke (-), vertical stroke (I), stroke down to the left (/), dot (.), and twist (Z). For the characters with the same first stroke, the stroke is used. Those in Level II are arranged according to the radicals. The radicals adopted are basically similar to those in the ordinary Chinese Dictionaries with minor variations. Characters with the same radical are arranged according to stroke - comma.

5. The Taipei code entitled Chinese Character Code for Information Interchange, Vol. 1. was compiled and published by the Chinese Library Association of Taiwan in April 1980. It is available in both Chinese and English editions.

This code is basically a 7-bit code set of 150646. Provision for Chinese characters is made by three 7-bit code vectors which form a finite 3-dimensional $94 \times 94 \times 94$ coding space. Each of 94 planes has 94 sections and each section has 94 positions. It is designed to provide coding rooms for more than 50 thousand characters. This is the code of the first plan which includes 4,807 most frequently used characters, 35 punctuation marks, 214 radicals, 41 Chinese numerical characters, 37 Chinese phonetic symbols and 4 tone marks.

This publication consists of the code table and section tables for the above mentioned characters. In plan 1, the first two sections are punctuation marks and special symbols. Section III is for ISO 646. Section XI is assigned for Chinese punctuation marks. Sections XII-XIV are assigned for the set of 214 radicals, Section XV is allocated for the set of Chinese phonetic codes

and Chinese numerical characters. Section XVI through to section LXVII are assigned for the first of 4807 Chinese characters.

The arrangement of Chinese characters is according to the established radical and stroke-count sequence. For those characters which belong to the same radical and the same stroke-count, they are sub-arranged by stroke-order. The order of strokes used is as follows; (1) a dot, (2) a horizontal stroke, (3) a vertical stroke, (4) a stroke down to the left, (5) a stroke down to the right. The radical set adopted is the same as those in the Kang-Hsi dictionary.

Section IV to Section X are allocated for users expansion. Section LXVIII through to section XCIV are reserved for the second series of Chinese characters. The second series of Chinese characters will also occupy plan 2. Plan 9 to Plan 64 are for variations and simplified forms of each Chinese character.

C. The Tokyo code entitled "Code of 16 Japanese Graphic Character Set for Information Interchange" was drafted by the Japan Information Processing Development Centre, Ministry of International Trade and Industry and adopted as the Japanese Industrial standard No JIS C6226 by the Agency of Industry, Science and Technology in 1978. It was published by the Japanese Standards Association in the same year and only the Japanese language edition is available.

This code uses 2 7-bit byte representation in accordance with Japan Standards JIS C6220 and JIS C6228. There are 94 sections and 94 positions in the code table. Each character is represented by the section number and position number. The section number is the first byte and the position number is the second byte.

The code covers the following graphic characters:

Special Characters (including the Space, punctuation marks, tabulation marks, unit marks, and symbols)	108
Arabic numerical characters	10
Roman alphabet (26 capital letters and 26 small letters)	52

(continued)

Hiragana	83
Katakana	56
Greek alphabet	48
Russian alphabet	66
Chinese characters	6,349
Total	6,802

The 6,349 Chinese characters are divided into two groups according to the degree of the frequency of use. The first group totalling 2,965 characters has a high frequency and the second group includes the rest, 3,384 characters which are considered to be less used.

The Chinese characters in the first group are arranged in the order of Japanese syllabary according to their on-yomi (reading taken from Chinese) or kun-yomi (native Japanese reading), whichever reading is designated for each Chinese character, the "on" reading is given preference. Where several Chinese characters have the same "on" or "kun" reading, they are arranged according to the order of their radicals and the number of their strokes.

Those in the second group are subgrouped according to the order of the 214 radicals in the Kang-hsi Dictionary. Within each subgroup of characters having the same radical, they are arranged by stroke-count. Where there are characters have the same number of strokes with that subgroup, they are arranged by their Japanese "on" or "kun" reading.

General characters, ordinal numbers, numerical characters, the Latin alphabet, the Japanese kana, the Greek alphabet, the Russian alphabet, Chinese phonetic symbols, and Chinese phonetic - annotated letters occupied from section 1 to section 10.

Section 10 to section 15 are reserved for necessary expansion.

(S. W. Wang & K. Hosaka
National Library of Australia)

SECOND BYTE 第二卦

Section 段
Position 位號

FIRST BYTE 第一卦

b ₇	b ₆	b ₅	b ₄	b ₃	b ₂	b ₁		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	1	
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	
0	1	0	0	0	0	1												ㄟ	片	音	丨	一	休	倒	內	劍	又	哀	喚	址	鈔	塊	密			
0	1	0	0	0	1	0												ㄨ	一	牙	頁	丨	丁	伏	施	全	劉	友	咨	哈	均	資	婦	寇		
0	1	0	0	0	1	1												ㄣ	丨	牛	風	川	七	伐	倘	兩	劑	反	咳	喻	冊	夢	婁	貨		
0	1	0	0	1	0	0												ㄥ	、	犬	飛	又	三	仲	俱	八	力	及	咬	喃	汲	夥	娶	寄		
0	1	0	0	1	0	1												ㄍ	丿		食	彡	上	件	倡	六	功	取	咩	嘛	坎	大	姿	寂		
0	1	0	0	1	1	0												ㄎ	乙	玄	首	丨	下	任	個	兮	加	叔	咪	嗷	坐	太	娟	宿		
0	1	0	0	1	1	1												ㄏ	丨	玉	香	丨	丈	仰	候	公	劣	受	喫	噴	立	天	姿	寒		
0	1	0	1	0	0	0												ㄏ		瓜		丨	丑	份	俳	共	劫	叛	腮	嗽	坪	夫	姐	富		
0	1	0	1	0	0	1												ㄏ	二	瓦	馬	女	弓	企	倭	兵	助	叟	哇	嘉	珂	天	婁	寅		
0	1	0	1	0	1	0												ㄏ	丨	甘	竹	十	不	住	修	其	幼	曼	晒	嗽	坩	央	婢	陳		
0	1	0	1	0	1	1												ㄏ	人	生	商	卅	丙	位	倪	具	勞	叢	戰	囉	坡	失	婚	寔		
0	1	0	1	1	0	0												ㄏ	丨	儿	用	彭	卅	且	佇	典	幼	口	哄	唯	坦	夷	婢	寧		
0	1	0	1	1	0	1												ㄏ	丨	入	田	門		世	佗	儻	兼	勇	叮	咸	囉	坤	夸	媚	寧	
0	1	0	1	1	1	0												ㄏ	一	八	疋	豎	丘	佞	倫	冀	勃	可	咽	喚	垂	夾	婦	奈		
0	1	0	1	1	1	1												ㄏ	ㄥ	門	疋	豎	丕	伴	停	冉	勁	司	品	嘈	疥	奉	媒	葬		
0	1	1	0	0	0	0												ㄏ	ㄥ	一	天	鬼	〇	丟	佛	偽	册	勉	叩	咯	咬	垠	奈	媛	實	
0	1	1	0	0	0	1												ㄏ	ㄥ	ㄥ	白		一	丞	何	假	再	務	古	咱	咎	型	奇	嫌	股	
0	1	1	0	0	1	0												ㄏ	ㄥ	几	皮	魚	二	並	何	健	冒	勳	右	咻	喫	垣	奔	嫁	諧	
0	1	1	0	0	1	1												。	ㄥ	口	皿	鳥	三	丫	估	偉	胃	勒	召	哈	嘍	垢	奄	嫉	察	
0	1	1	0	1	0	0												。	刀	目	齒	四	中	佐	借	冕	動	叨	唱	囉	城	奕	媽	察		
0	1	1	0	1	0	1												、	力	才	屮	屮	五	串	佑	偵	最	勞	叨	啞	啞	垮	契	婦	寔	
0	1	1	0	1	1	0												：	ㄥ	ㄥ	矢	麥	六	丸	伽	偃	冗	助	另	喃	噎	噎	奏	媿	寔	
0	1	1	1	0	0	0												、	ㄥ	示		八	主	但	側	冢	募	台	卿	嘶	埋	與	嫂	龍		
0	1	1	1	0	0	1												；	ㄥ	內	黃	九	丹	仲	假	冥	勞	句	哪	嘲	埃	蚤	媿	寶		
0	1	1	1	0	1	0												：	十	禾	黍	十	乃	佣	偶	寬	勁	叱	哈	寮	執	裝	嬌	寸		
0	1	1	1	0	1	1												？	卜	穴	黑	百	久	佃	倏	冬	勳	只	哥	嘿	培	奚	嬌	寺		
0	1	1	1	1	0	0												丨	口	立	齋	千	么	佔	偏	冰	勳	叭	呼	噎	噎	噎	噎	噎	噎	
0	1	1	1	1	0	1												！	厂			万	之	似	偷	治	勳	史	哲	嘶	噎	噎	噎	噎	噎	
0	1	1	1	1	1	0												丨	ㄥ	竹	龜		尹	作	傍	冷	勻	呼	喃	嘆	噎	噎	噎	噎	噎	
0	1	1	1	1	1	1												丨	又	米	鼎		乏	尙	傢	冽	勾	吋	哨	啤	噎	噎	噎	噎	噎	
1	0	0	0	0	0	0														系	鼓	零	乍	你	傅	凍	勿	吉	貝	殘	噎	噎	噎	噎	噎	
1	0	0	0	0	0	1														口	岳	風	壹	乎	伯	備	凌	包	吐	哩	噎	堂	奮	媿	尊	
1	0	0	0	0	1	0														口	网		貳	兵	低	傑	湖	奴	更	哭	噎	噎	噎	噎	噎	
1	0	0	0	0	1	1														土	羊	鼻	叁	兵	份	傀	淮	河	同	喚	噎	噎	噎	噎	噎	
1	0	0	0	1	0	0														土	羽	齊	肆	乖	余	偷	涼	徇	吊	峻	噎	噎	噎	噎	噎	
1	0	0	0	1	0	1														久	老		伍	乘	依	傘	凝	幻	吃	噎	噎	噎	噎	噎	噎	
1	0	0	0	1	1	0														久	而		齒	陸	乙	伴	備	兒	飽	名	商	噎	噎	噎	噎	

Common Core Character Set

