

L2 98-415

Date: 1 December 1998

Author: John H. Jenkins
Apple Computer, Inc.
2 Infinite Loop, M/S 302-2IS
Cupertino, CA 95014
Phone: +1-408-974-6276
Fax: +1-408-862-4566
Email: jenkins@apple.com

Comments on the CJK Radicals Supplement

Background

Radicals are ideographs or fragments of ideographs. Their primary uses are to provide a means to index dictionaries and to coin new ideographs. They originated as a set of determinatives which gave a clue as to the meaning of a character; homonyms would be written with the same phonetic and different radicals to distinguish meanings. Since not all ideographs have determinatives, arbitrary radicals have been assigned to a number of characters for purposes of dictionary indexing (e.g., assignation of the character 井 [well] to the two radical 二).

No single set of radicals has been used universally throughout East Asia. The first major Chinese dictionary, the *ShuoWen*, had a set of several hundred radicals. By the 17th century, when the definitive *KangXi* dictionary was published, that set had been reduced to 214. The 214 *KangXi* set remains the most common in modern use, particularly for traditional Chinese. Other sets are found, however.

Most radicals are themselves ideographs. For example, the common radical 水 —used in ideographs for liquids, rivers, lakes, and so on—is itself the ideograph meaning *water*. This is analogous to the fact that the English letters "A" and "I" serve double duty as English words.

Radicals generally alter their shapes when they are part of characters. The *water* radical, for example, is usually written as a set of three dots at the left of a character, 氵. This is not an invariable rule: in some characters, the water radical can appear at the top or bottom of the character as a vertically compressed 水 (e.g., 漿 U+6D2F), or in some other form (e.g., 永 U+6C38).

This is further complicated by the fact that several radicals can occur in numerous common graphic variants. The most notorious of these is the *grass* radical, 艹. When it occurs within a character, it is drawn on top of a character, e.g., as the 艹 in 艾 (U+872E, *artemesia*). In some writing styles, this form of the

grass radical is written with three strokes (艾) and in some with four strokes (艾).

Conceptually, there is a difference between a character's use as an isolated radical and as an ideograph—the character “A” has a different meaning in the sentence, “‘A’ is for ‘apple,’” and in the sentence “A Macintosh is a very nice computer.” Radicals *qua* radicals can occur in plain text and are used as headings in dictionaries (e.g., the Unicode book); in such cases, they do not carry meaning *per se* but describe the structure or graphic appearance of other characters.

The KangXi Radicals, Unicode, and the CJK Radicals Supplement

The 214 KangXi radicals are all in Unicode already. For the most part, the forms that the radicals actually take in characters are also encoded—e.g., the three-dot form of the *water* radical is found at U+6C35 separate from the full form at U+6C34.

There is a duplicate set of the 212 of the 214 in CNS 11643. All 214 are also being added to Unicode 3.0 in the guise of the KangXi radicals.

The CJK Radicals supplement is *not* derived from non-KangXi radical sets. It is, instead, *yet another* collection of KangXi radicals. Here, we have all the graphic forms that the KangXi radicals will take within characters—even to the point of having a three-stroke *grass* radical and a four-stroke one (P+2EBE and P+2EBF). This is analogous to encoding “a” and “ɑ;” it is a duplicate encoding of a single abstract shape. There are *three* graphic variations for the *walk* radical (P+2ECC, P+2ECD, and P+2ECE). Note that these are *stylistic* variations for the character. Any typeface would select one of the three and use it consistently for all characters with this radical; there is no evidence that any two of the three would be used in a single typeface design or in plain text.

The description for the use of Ideographic Description Characters which has come out of the IRG indicates that they are to be used to describe an “ideograph in the abstract form.” This makes this repeated duplication of abstract shapes within the CJK Radicals Supplement even more bizarre.

There is a case that can be made for including the 212 duplicated radicals from CNS.

There is a weaker case that can be made for the two not duplicated within CNS.

The CJK Radicals Supplement, however, should be rejected. It *does* contain characters/character fragments which are not found in any of the current ideographic blocks in Unicode 2.1 or proposed for Unicode 3.0; but it also contains a significant number of characters which are simply graphic variations of one another. (Note, for example, the presence of the PRC version of the *bone* radical—P+2EE3—despite the presence of its non-PRC version at P+1FBB, and despite the fact that the two versions of the character are explicitly unified as U+9AA8).

There is no sufficient rationale given for these characters. They are not needed for the ideographic description mechanism proposed by the IRG. There is no evidence provided that the distinctions they make between various graphic forms of radicals occurs in plain text. In particular, inasmuch as different (and undefined) criteria are used for unifying these characters than is the case

anywhere else in Unicode, and since these unification rules are contradictory to those used in Unicode, the sponsors of the proposal must provide strong evidence that this is the appropriate course to follow.

Conceivably, one might want to have the graphic forms the radicals take encoded in Unicode as this is useful for font design and producing literature which refers to the radicals in their fragment form (e.g., dictionary indices). Even here, however, the proposal is insufficiently justified. That is, one could argue that P+2F54 KANGXI RADICAL WATER is distinct from U+6C34—the latter is an ideograph and carries semantic meaning, the former isn't and doesn't. What, however, is the basis for distinguishing P+2EA1 CJK RADICAL WATER ONE from U+6C35? Neither is an ideograph proper; neither *means* anything. They look the same. They have the same function. How is a user to determine when to use one and when to use the other?

Recommendations

The United States should vote *no* on PDAM 15.

The CJK Radicals Supplement contains numerous characters which glyphic variations of other characters in ISO/IEC 10646, in the Vertical Extension A, or in the KangXi Radicals set, or of one another.

Character shapes which have been unified within ISO/IEC 10646-1 must not be deunified by the KangXi Radicals and CJK Radicals Supplement. For example P+2E63 CJK RADICAL BONE and P+2FBB KANGXI RADICAL BONE, have shapes unified as U+9AA8.

Graphic variations of radical shapes used within ideographs must not be allowed where it is not the case that both variations would be used within a single type design. For example, a single typeface would not use both P+2E3E CJK RADICAL GRASS ONE and P+2E3F CJK RADICAL GRASS TWO.

Characters which are already in ISO/IEC 10646-1 or the Vertical Extension A should not be duplicated in the CJK Radicals Supplement. For example, P+2E21 CJK RADICAL WATER ONE is a duplicate of U+6C35.

The United States will change its vote to *yes* if such duplicates are eliminated from the CJK Radicals Supplement.

P+2FA3 KANGXI RADICAL BIRD SIGN should be renamed KANGXI RADICAL WINE. The term "BIRD SIGN" was an error in the original set of names provided by the US to the IRG.

On a Related Note

PDAM 28 (Ideographic Description Characters), section D.3.1 says, "A DC may be any one of the following: a coded ideograph, a coded radical, a coded ideographic component, another IDS." There are currently no "coded ideographic componenets" in ISO/IEC 10646. This line should be removed.

Table 244 - Row 2E: CJK RADICALS SUPPLEMENT

	2E0	2E1	2E2	2E3	2E4	2E5	2E6	2E7
0	𠂇 000	尢 016	民 032	彡 048	廾 064	廾 080	个 096	龙 112
1	廾 001	允 017	彡 033	𠂇 049	虎 065	長 081	首 097	龜 113
2	𠂇 002	巳 018	氷 034	𠂇 050	衤 066	長 082	馬 098	龜 114
3	𠂇 003	彡 019	彡 035	𠂇 051	𠂇 067	長 083	骨 099	龜 115
4	𠂇 004	𠂇 020	𠂇 036	𠂇 052	𠂇 068	𠂇 084	鬼 100	
5	𠂇 005	𠂇 021	𠂇 037	𠂇 053	𠂇 069	𠂇 085	魚 101	
6	𠂇 006	𠂇 022	𠂇 038	𠂇 054	角 070	𠂇 086	鳥 102	
7	𠂇 007	𠂇 023	𠂇 039	𠂇 055	角 071	𠂇 087	𠂇 103	
8	𠂇 008	𠂇 024	𠂇 040	𠂇 056	𠂇 072	青 088	麥 104	
9	𠂇 009	𠂇 025	𠂇 041	𠂇 057	𠂇 073	𠂇 089	黃 105	
A	𠂇 010	无 026	正 042	𠂇 058	𠂇 074	𠂇 090	𠂇 106	
B	𠂇 011	𠂇 027	𠂇 043	𠂇 059	𠂇 075	𠂇 091	𠂇 107	
C	𠂇 012	𠂇 028	𠂇 044	𠂇 060	𠂇 076	𠂇 092	𠂇 108	
D	𠂇 013	𠂇 029	𠂇 045	𠂇 061	𠂇 077	𠂇 093	𠂇 109	
E	𠂇 014	𠂇 030	𠂇 046	𠂇 062	𠂇 078	𠂇 094	𠂇 110	
F	𠂇 015	𠂇 031	𠂇 047	𠂇 063	𠂇 079	𠂇 095	𠂇 111	

虎 (864E)

骨 (9AAS)

角 (89D2)

𠂇 (7F8B)

𠂇 (8288)

(Radical 197)

𠂇

𠂇

G = 00

P = 00

辭海合訂本部首索引

一畫	一	乙	二	二畫	人	入	八	一	几	刀	力	勺	匕	匚	匚	十	卜	匚	同		
總頁數	一	三	七	八	八〇	七五	五八	一三一	一四一	一四一	一四一	一四一	一四一	一四一	一四一	一四一	一四一	一四一	一四一	一四一	
又	厶	厂	附	口	口	口	口	士	士	夕	夕	夕	夕	夕	夕	夕	夕	夕	夕	夕	夕
三畫	巴	同	同	同	同	同	同	同	同	同	同	同	同	同	同	同	同	同	同	同	同
斗	文	支	支	手	戶	戈	心	小	小	小	小	小	小	小	小	小	小	小	小	小	小
六二二	六〇八	五九九	五九八	五五七	五五五	五五〇	五二二	四九二	四八八	四八八	四八八	四八八	四八八	四八八	四八八	四八八	四八八	四八八	四八八	四八八	四八八
无	方	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤	斤
六二九	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六	六二六
立	穴	禾	内	帝	石	矢	矛	目	血	皮	白	夂	疒	疒	疒	疒	疒	疒	疒	疒	疒
一〇〇二	九九七	九九六	九九六	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五	九九五
舟	舛	舌	白	至	自	臣	肉	聿	耳	耒	而	老	羽	羊	羊	羊	羊	羊	羊	羊	羊
一一七	一一六	一一五	一一二	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一	一一一
采	酉	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑	邑
一三五六	一三四八	一三三五	一三三三	一三三一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一	一三一一
高	骨	馬	十	附	香	首	食	飛	風	頁	音	韭	韋	革	面	非	青	雨	佳	隶	阜
一五一四	一五一二	一四九九	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二	一四九二
俞	龜	龍	齒	齊	鼻	鼠	鼓	鼎	聾	黑	黍	黃	麻	麥	鹿	鹵	鳥	魚	鬼	扇	鬯
一五七七	一五七二	一五七二	一五七一	一五六九	一五六八	一五六七	一五六六	一五六五	一五六四	一五六〇	一五五九	一五五三	一五五二	一五五二	一五四八	一五四六	一五四三	一五三四	一五三一	一五三一	一五三一

部首索引		《	152	支	225	牛	331	
一畫	十	60	工	153	支	226	犬	334
	卜	61	己	154	文	230	女	同支
	1	62	巾	155	斗	230	一	同火
	1	64	干	159	斤	231	一	同爪
	5	65	幺	160	方	232	允	同七
	6	66	又	161	尤	234	手	同玉
	7	同入	乚	165	日	235	目	同网
	8	同刀	卩	165	冂	242	冂	同网
二畫	巳	同卩	弋	166	月	244	冂	同网
	9	三畫	弓	166	木	246	冂	同肉
	10	69	ヨ	169	欠	274	冂	同岫
	12	98	彳	170	止	276	冂	同走
	35	101	兀	同尤	夕	278	五畫	
	38	114	互	同ヨ	爰	281	女	341
	39	115	亅	同心	母	281	玉	341
	40	115	扌	同手	比	282	瓜	348
	41	117	彳	同水	毛	283	瓦	348
	42	121	彳	同犬	氏	284	甘	349
44	131	阝	右邑	气	287	生	350	
45	133	阝	左阜	水	318	用	351	
45	140	四畫	爪	328	田	疋	351	
54	142	心	175	父	329	疋	355	
57	143	心	190	爻	329	疋	356	
59	144	尸	193	片	330	八	363	
59	147	尸	193	片	330	白	364	
60	147	手	194	牙	331	皮	365	

五	366	耒	440	豆	533	九畫	鹿	660
目	368	耳	442	豕	533		麥	661
子	375	聿	444	豕	534		麻	662
彳	375	肉	444	貝	535		面	619
彳	376	臣	456	赤	542		革	619
彳	385	自	457	走	543		韋	622
彳	388	至	458	足	545		韭	622
彳	389	白	458	身	552		音	622
彳	395	舌	459	車	553		頁	623
彳	398	舛	460	辛	558		風	628
同目		舟	460	辰	558	飛	629	
同水		艮	462	走	559	食	629	
同网		色	462	邑	559	首	634	
同夕		艸	463	酉	570	香	634	
同疋		虎	491	采	574	十畫		
同疋		虫	492	里	579	馬	637	
同母		血	504	里	579	骨	641	
六畫		行	505	八畫		高	643	
竹	401	衣	506	金	581	髟	643	
彳	411	西	513	長	600	鬥	645	
彳	415	羊	同羊	門	601	鬯	645	
彳	432	西	同西	阜	606	鬲	646	
河	433	七畫		隹	611	鬼	646	
彳	435	見	515	雨	612	十一畫		
彳	437	角	517	青	614	魚	647	
老	439	言	518	非	618	鳥	653	
彳	440	谷	532	非	618	鹵	660	
十二畫								
黃	663							
黍	663							
黑	663							
脊	665							
十三畫								
龍	665							
鼎	665							
鼓	666							
鼠	666							
十四畫								
鼻	667							
齊	667							
十五畫								
齒	668							
十六畫								
龍	669							
龜	670							
十七畫								
龠	670							