Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1298
Date: 2007-5-25

Source:	Taichi Kawabata
Title:	Results of IDS check and comments.
Status :	
Actions required	Review by IRG Editors for discussion at IRG meeting No. 25.
Distribution:	IRG Members and Ideographic Experts
Medium :	Electronic

1. IDS check result

Attached archives include the report for possibly unifiable candidates. They are as follows:

- ids-china-070504.txt … IDS file submitted by China.
- ids-dprk-070504.txt … IDS file submitted by DPRK.
- ids-japan-070520.txt … IDS file submitted by Japan.
- ids-macao-070421.txt … IDS file submitted by Macao.
- ids-taiwan-070504.txt … IDS file submitted by Taiwan.
- ids-utc-070412.txt … IDS file submitted by UTC.
- ids-vietnam-070518.txt … IDS file submitted by Vietnam.
- report-china-070521.txt … List of possibly unifiable ideograph for China.
- report-dprk-070521.txt … List of possibly unifiable ideograph for DPRK.
- report-japan-070521.txt … List of possibly unifiable ideograph for Japan.
- report-macao-070521.txt … List of possibly unifiable ideograph for Macao.
- report-taiwan-070521.txt … List of possibly unifiable ideograph for Taiwan.
- report-utc-070521.txt … List of possibly unifiable ideograph for UTC.
- report-vietnam-070521.txt … List of possibly unifiable ideograph for Vietnam.
- error-070521.txt … List of problematic IDS.

Submitters, please check each IDS files for missing characters, report files s for unification candidates, and `error-07521.txt' file for problematic entries. This file contains a list of IDS such that:

- IDS contains wildcard ‘ ? ’ character. Though they are not the errors, IDS which include this character will be excluded from the IDS check processing. They are listed as `[Not-processable]’ in the error file.

- IDS contains non-ideographic, extra, or insufficient DC. They are incomplete IDS, and can not be processed at all. Such IDS are listed as `[errors]’ in the error file.

- IDS contains more than 3 IDCs. They are considered to be overly detailed, may not be processed accurately. They are also not recommended by IRG N 1183. They are listed as `[warnings]' in the error file.

- IDS contains overlapping IDC (⿻). Overlapping IDC is more ambiguous than other IDCs, and IDS that contains overlapping IDC may not be accurately compared with other IDS. Their use is strongly discouraged, as stated in IRG N 1183. They are also listed as `[warnings]' in the error file.

Here are the brief summaries of the errors, warnings and unification candidates, as of 2007/05/28.

	Total	Errors	Not-Processed	Warnings	Unification candidates	
China	6706	5	1	57	308	(4.6%)
DPRK	67	0	5	10	7	(10.4%)
Japan	438	0	22	0	8	(1.8%)
Macao	43	0	0	2	20	(46.5%)
Taiwan	8159	6	127	433	534	(6.5%)
UTC	472	0	7	12	64	(13.6%)
Vietnam	1327	0	0	40	52	(3.9%)

2. Comments on submitted IDS.

Some regions seem to neglect the instruction in the IRG N 1183. Followings are the comments on submitted IDS.

2.1. Please do not use more than three IDCs.

As stated by the IRG N1183, construct the minimal IDS. If you think you need more than three IDC for specific character, just give up and leave the IDS entry as ' ? '.

2.2. Please do not submit IDS by graphic images.

Some regions/countries submit IDS with embedded graphic images. IDS-checker can not handle such IDS for unification checking.

2.3. Please do not use non-ideographic characters.

Some regions/organizations heavily use non-ideographic characters, including PUA and Bopomofo characters for DC and IDC. IDS-checker can not check for the duplicate/unifiable characters if non-comparable characters are provided in the IDS.

2.3. Please do not use overlapping IDC (⧉).

Some regions/organizations heavily use overlapping IDC (⧉). They may not be accurately compared with other IDS, so avoid using them as possible as you can. For example, “夸” should be described as “⿱夫亏”, not “⧉夸丷”.

2.4. Please be faithful to visual shape.

Some regions/organizations neglect the actual shape of the character and adopted “original radical form” in the IDS. (For example, decomposing “迅” as “⿺辵卂”, not “⿺辶卂”, or “草” as “⿱艸早”, not “⿱艹早”.) Such IDS will not be correctly handled in the IDS-checking process. Please be faithful to the visual shape, not the original or archaic form, for the decomposition. This is also mentioned in IRG N 1183.

2.5. Please avoid the use of surrounding IDC (⧈) for `split and put’ patterns.

Some regions/organizations use surrounding character for split-and-put patterns. (For example, decomposing “弼”as “⧈弜刀”, not “⿰弓百弓”, or “辦” as “⧈辡刀”, not “⿰辛刀辛”.) Please avoid such use of surrounding IDC, and try use ⿰ or ⿱ instead, as possible as you can.

3. Some concerns on IDS checking process

3.1. Concern on IDS quality.

Some regions/organizations submit IDS data without reviewing by themselves. I’m afraid that such data may also contain many wrong IDS entries. Due to such quality concerns, I propose that at least TWO regions/countries should construct IDS of the SAME proposed character sets, and the third party should compare them line by line and choose the better IDS.

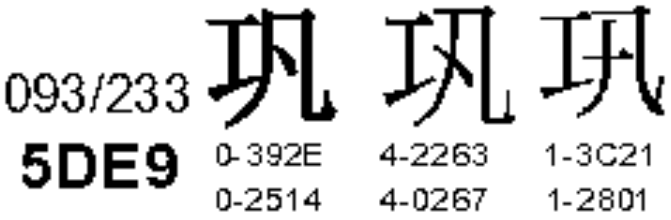
3.2. Concern on Duplicate Characters.

Characters submitted by some regions/organizations found to contain many duplicate characters. Please KEEP IN MIND of 5% RULE, as stated in IRG N 880. If your submission is found to contain more than 5% of duplicate characters, you might be asked to withdraw your entire submission. More characters found to be duplicate by the IDS, so the higher possibility of duplicate character that cannot be found by the IDS. If you think your organization has a problem on quality management, you may withdraw submission voluntarily and re-submit later for a better quality of the standard.

End of document.

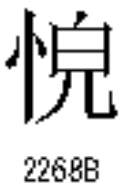
number of lines=8

JK-65081 ⿰土八

U+5DE9 (巩) ⿰工凡[G]/⿰工孔[TJ] 

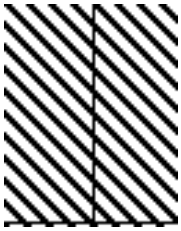
(unification:26a,61a,69)

JK-65170 ⿰忄臾

U+2268B(悦) ⿰忄兑 

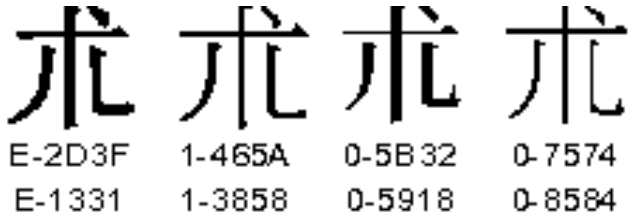
(unification:26a)

JK-65180 ⿰力戈

U+2AB61(extC) ⿰戈力 

JK-65220 ⿰木一

U+672E (朮) ⿺朮丶 103/046 672E



U+672F (术) ⿺木丶 103/047 672F

(unification:30a,丶)

JK-65265 ⿰木勇

U+2ACE6(extC) ⿰木勇

JK-65328 ⿰木畫

U+236DB(橅) ⿰木畫

(unification:177)

JK-65836 ⿰金荘

U+28AF2(錘) ⿰釒 荘

(unification:162a,釒)

JK-66058 ⿰亻 鳩

U+2A006(鴆) ⿰仕鳥

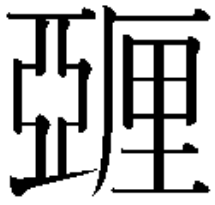
(unification:1a,61a)

number of lines=16

JK-60285 ⿰扌 ⿱艹亡

V04-4921 ⿰扌 ⿱艹亡 (no image)

JK-65014 ⿰亞厘



UTC00485 ⿰亞⿰厂里 (no image)

JK-65073 ⿰口宿

UTC00559 ⿰口宿 (no image)

JK-65079 ⿰土⿱亠合

TD-2367 ⿰土⿱亠合 (no image)

JK-65081 ⿰土八

CYY00197 ⿰土八 (no image)

TC-234E ⿰土八 (no image)

WZ100441 ⿰工几 (no image)

(unification:26a,61a,69)

JK-65280 ⿰木苗

TD-3C55 ⿰木苗 (no image)

JK-65294 ⿰木荷

CYY00776 ⿰木荷 (no image)

TD-4E30 ⿰木荷 (no image)

JK-65332 ⿰木夢

TE-2D21 ⿰木夢 (no image)

JK-65486 ⿱竹乃

CYY01138 ⿱𥫱乃 (no image)

(unification:𥫱)

JK-65612 ⿱艹寬

TE-3030 ⿱廾⿴𠂉⿱廾見 (no image)

(unification:82,艹)

JK-65678 ⿰虫任

TD-3228 ⿰虫任 (no image)

JK-65699 ⿰虫莆

V04-5372 ⿰虫⿱艹雨 (no image)

JK-65710 ⿰虫夢

V04-536E ⿰虫夢 (no image)

JK-65941 ⿰魚卜

TD-4772 ⿰魚卜 (no image)

JK-66055 ⿰車鳥 UTC00841 ⿰車鳥 (no image)
JK-66064 ⿰沙鳥 WZ102737 ⿰沙鳥 (no image)

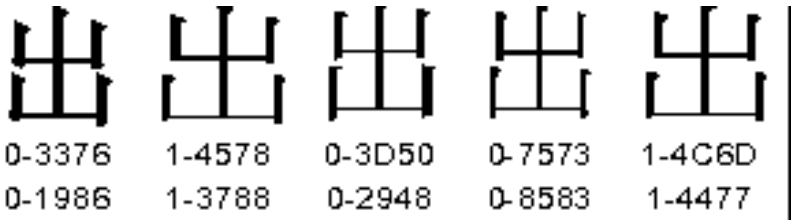
number of lines=64

UTC00090 ⿴凵屮

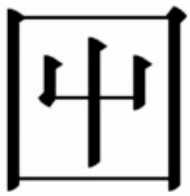
U+51FA(出) ⿱屮凵

081/250

51FA



UTC00091 ⿴口屮



U+211AF(𑋯) ⿴口山

211AF

(unification:62-5)

UTC00101 ⿱𠂉口皿

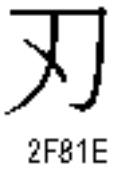
U+2504F(𥁏) ⿱召皿

2504F

(unification:31a,81)

UTC00129 ⿴刀一

U+2F81E(刃) ⿻刀丶



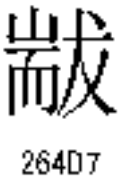
(unification:丶)

UTC00166 ⿰歹粲

U+6BA9(殩) ⿰歹粲

UTC00188 ⿰嵬支

U+264D7(𦓗) ⿰嵬攴



(unification:152-5,24)

UTC00194 ⿰虫郤

U+27355(蝃) ⿰虫卻

(unification:郤)

UTC00221 ⿰⿱古⿱勹缶皿 U+250A9(𥫩) ⿰⿱古匋皿	𥫩 250A9
UTC00224 ⿰⿱大二日 U+21610(𡘐) ⿰日大亠日	𡘐 21610
UTC00224 ⿰⿱大二日 U+21610(𡘐) ⿰日大亠日 (unification:124a,丶)	𡘐 21610
UTC00244 ⿰丷 ⿱灬大	𤇾

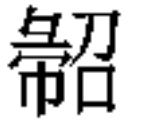
U+452A (蒌)	日艹爽 [T]/日艹爽 [G]	069/042 452A	3-6A56 3-7454	4-4C6F 4-4479

(unification:152-12,82,艹)

UTC00260	⿰刃刂	刂
U+206C2(刂)	⿰刃刂	刂 206C2

(unification:129-7,170)

UTC00266	⿰帚召	韶
U+22466(韶)	⿰目彑冖巾召	韶 22466



(unification:167)

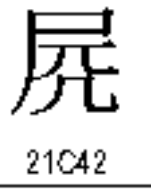
UTC00269	⿰忄瑩	㦤
U+228D9(㦤)	⿰忄日𤇾王	㦤 228D9

(unification:129-6)

UTC00274	⿱小兒	覍
U+3778(覍)	⿱小兒[GT]/⿱小⿱日儿[J]	055/120 3778 覍 5-4078 5-3288 覍 4-2C4C 4-1244 覍 A-2235 A-0221
U+21B6A(覍)	⿱小⿱日儿	覍 21B6A

(unification:71-3)

UTC00276	⿸尸兂	屍
U+21C42(屍)	⿸尸⿱兀几	屍 21C42

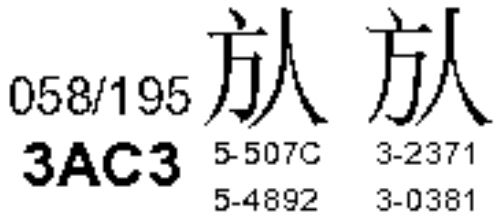


UTC00277	⿱山六	岦
U+21D06(岦)	⿱屮六	岦 21D06

(unification:62-5)

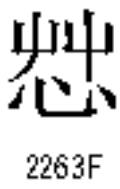
UTC00282 ⿰方人

U+3AC3(旀) ⿰方人



UTC00283 ⿱屾心

U+2263F(𢘏) ⿱艸心



(unification:62-5)

UTC00318 东

U+4E1C(东) 东

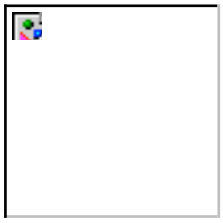
UTC00325 ⿰久龜

U+2A6AA(龜) ⿰夕龜

2A6AA

(unification:62-2)

UTC00331 ⿰亻皮



U+4F4A(彼) ⿰亻皮

079/074
4F4A

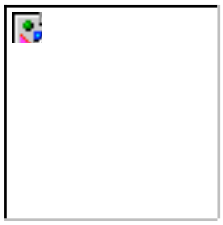
3-3076
3-1686

3-2444
3-0436

2-217B
2-0191

0-3029
0-1609

UTC00354 ⿰口虐



U+5611(嘘) ⿰口虐

086/017
5611

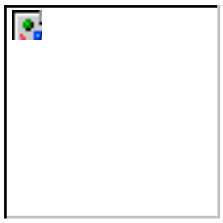
E-2472
E-0482

3-4634
3-3820

1-3631
1-2217

2-284D
2-0845

UTC00368 ⿰礻阝

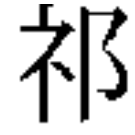


U+7941 (祁) ⿰礻阝[GTJ] /⿰示阝[K]

121/065
7941

0-466E
0-3878

1-4F25
1-4705


0-3737
0-2323

0-5135
0-4921

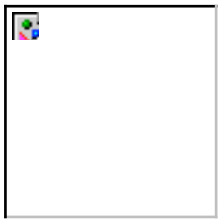
(unification:161)

UTC00372 ⿰口賴

U+210E4(𡃤) ⿰口賴

210E4

(unification:151)

UTC00411 ⿰石扇 

U+25531(碱) ⿰石扇

25531

UTC00421 ⿰糸弔

U+25FAF(紼) ⿰糸弔

25FAF

(unification:糸)

UTC00434 ⿱⺮咼

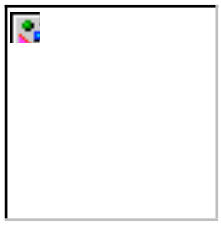
U+4211(篙) ⿱竹⿱⿰口冂⿱⿰口冂 066/017 **4211** 篙 3-4D61 3-4565

(unification:⺮)

UTC00445 ⿰亻⿱⿰口冂⿱⿰口冂先先日 僭

U+50ED (僭) ⿰亻⿱⿰口冂⿱⿰口冂 080/237 **50ED** 僭 0-5954 0-5752 僭 1-6776 1-7186 僭 0-5128 0-4908 僭 0-7350 0-8348 僭 1-4C2C 1-4412

UTC00461 ⿰口戸



U+20BD6(⿰口戸) ⿰口戸 ⿰口戸 20BD6

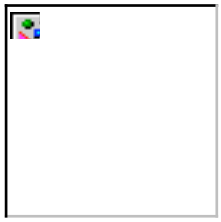
(unification:5)

UTC00471 ⿰口感 嘁

U+20FD1(嘁) ⿰口感 嘁 20FD1

UTC00489 ⿰犀刂

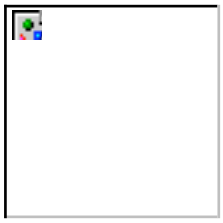
U+34FE(剾) ⿰犀刂 052/254 剾 剾
34FE 3-3331 4-4241
3-1917 4-3433

UTC00496 ⿱執目 

U+2528D(𥴉) ⿱執目 𥴉
2528D

UTC00505 ⿰冫今

U+2A79A(extC) ⿰冫今

UTC00509 ⿰各刂 

U+34E2(剢) ⿰各刂 052/226 剢 剢
34E2 3-327B 4-254C
3-1891 4-0544

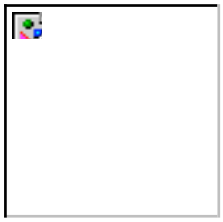
UTC00512 ⿰善刂

U+207E4(剹) ⿰善刂

剹

207E4

UTC00518 ⿰口兀



U+3564(吰) ⿰口九

053/100

3564

吰

5-3764

5-2368

吰

3-2230

3-0216

U+3571(吱) ⿰口尤

053/113

3571

吱

3-246F

3-0479

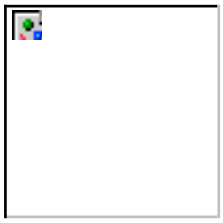
U+20BBE(吰) ⿰口兀

吰

20BBE

(unification:129-4,62-8,65a)

UTC00518 ⿰口兀



U+53FD(叽) ⿰口几

083/253
53FD

0-5F34
0-6320

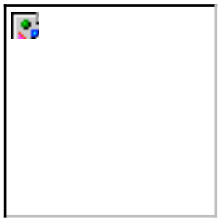
4-216A
4-0174

U+20BBE(𠮾) ⿰口兀

20BBE

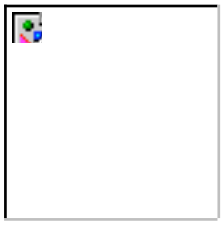
(unification:123-1)

UTC00541 ⿰口耗



U+2A875(extC) ⿰口耗

UTC00543 ⿺辶⿱日隹



U+66B9 (暲) ⿺辶⿱日隹

102/185
66B9

0-655F
0-6963

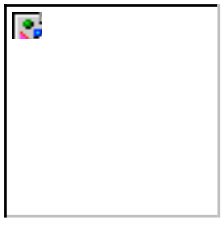
1-706F
1-8079

0-5A7B
0-5891

0-6067
0-6471

1-586C
1-5676

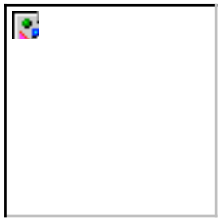
UTC00551 ⿺辶⿱口窒



U+35E7(噧) ⿰口窒 053/231 **35E7** 噧 3-372E 3-2314 噧 4-424A 4-3442

UTC00556 ⿰口敏

U+20F4A(𠽊) ⿰口敏 𠽊 20F4A

UTC00570 ⿰口微 

U+2100E(𡈎) ⿰口微 𡈎 2100E

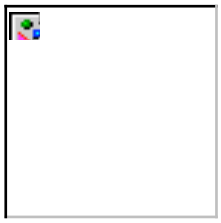
UTC00578 ⿰口確

U+210DC(𡓜) ⿰口確 𡓜 210DC

UTC00594 ⿱臼工

U+573C(⿱日土) 087/060 573C 5-494C 5-4144 3-2531 3-0517

(unification:61a,71-3)

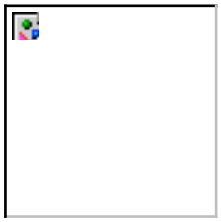
UTC00600 ⿰口聽 

U+2A8AB(extC) ⿰口聽 2B 2F

UTC00610 ⿰土步

U+57D7(⿰土步) 087/215 57D7 5-362A 5-2210 F-2E5C F-1460

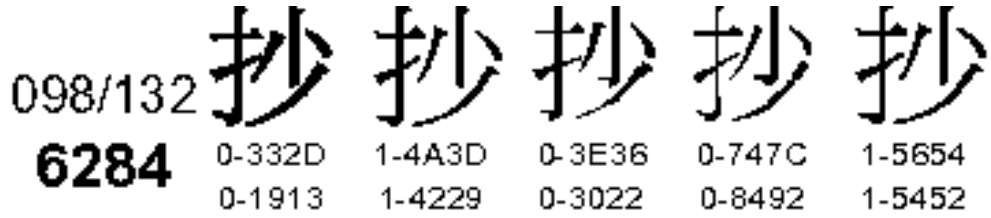
(unification:136)

UTC00621 ⿰口秃 

U+20CF6(哦) ⿰口禿

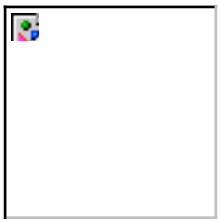
UTC00635 ⿰扌小

U+6284(抄) ⿰扌少



(unification:10a,136)

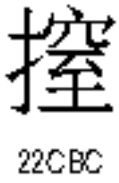
UTC00646 ⿰扌朗



U+2ABB8(extC) ⿰扌朗

UTC00650 ⿰扌窒

U+22CBC(揆) ⿰扌窒



UTC00685 ⿰氵壬

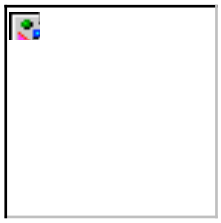
U+6C6A(汪) ⿰氵王

108/106
6C6A

0-4D74 1-4A6D 0-5D6A 0-685C
0-4584 1-4277 0-6174 0-7260

(unification:129-6)

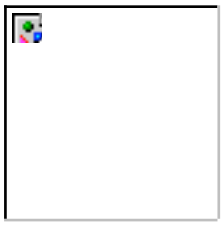
UTC00691 ⿰氵頃



U+23EF3(澒) ⿰氵頃

23EF3

UTC00718 ⿰禾皮



U+79DB(秡) ⿰禾皮

121/219
79DB

3-5B67 3-3279
3-5971 3-1889

UTC00725 ⿰缶呈

U+2B16C(extC) □缶呈

UTC00732 □月敦

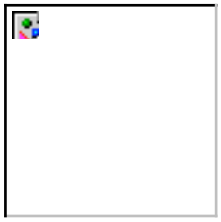
U+3B3F(⿰月敦) □月敦

059/063
3B3F

⿰月敦
3-5125
3-4905

⿰月敦
3-2761
3-0765

UTC00734 □月業



U+26882(⿰月業) □月業

⿰月業
26882

UTC00735 □月廉

U+81C1 (⿰月廉) □月廉 [GJ] /□?廉 [T]

129/193
81C1

⿰月廉
0-6C21
0-7601

⿰月廉
4-5827
4-5607

(unification:152-13)

UTC00743 ⿰扎虫

U+27252(蚑) ⿰扎虫

蚑

27252

UTC00763 ⿰天氺

U+6C97 (汞) ⿰天氺 [GJ] /⿰天水[T]

108/151 6C97

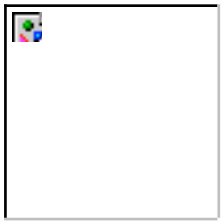
汞 E-3028 E-1608

汞 3-2963 3-0967

汞 1-466F 1-3879

(unification:17a)

UTC00779 ⿰匚⿰厂虎



U+3538 (匴) ⿰匚虎 [G]/⿰匚虎 [T]

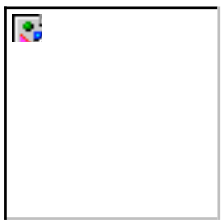
053/056 3538

匴 3-304F 3-1647

匴 4-3650 4-2248

(unification:118,厂)

UTC00780 ⿰匚贛

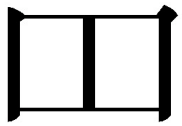


U+3536 (⿰匚贛)	⿰匚贛 [G]/⿰匚 ⿰匚 贛[T]	053/054 3536	3-3052 3-1650	4-6D53 4-7751

(unification:贛)

UTC00784	⿰口术	味
U+20C32(⿰口朮)	⿰口朮	20C32

(unification:30a)



UTC00834	?	皿
U+2053F(⿰冂人)	⿰冂人	2053F
U+20573(⿱冂入)	⿱冂入	20573

(unification:79,冖,⿱冂入,⿰冂人)

number of lines=47

UTC00109 ⿰⻊夆

TD-564D ⿰足夆 (no image)
(unification: ⻊)

UTC00124 ⿰爿刂

UTC00127 ⿰爿刂

UTC00127 ⿰爿刂

UTC00124 ⿰爿刂

UTC00131 ⿰木烋

TD-3D3B ⿲木火禾 (no image)

UTC00224 ⿱日⿴大二日

CYY00485 ⿱日大一日 (no image)

(unification:124a,丶)

UTC00273 ⿰日艹暪

GH101446 ⿰日艹暪 (no image)

UTC00302 ⿰土建

CYY00362 ⿰土建 (no image)

UTC00327 ⿰韦华

CH402292 ⿰韦华 (no image)

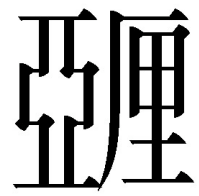
UTC00338 ⿰火乚

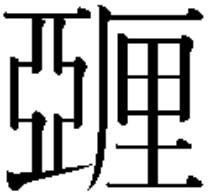
ZFY01026 ⿰火乚 (no image)

UTC00345 ⿰月枕

UTC00443 ⿰月枕

UTC00390 ⿰口夆	
UTC00562 ⿰口夆	
UTC00418 ⿰扌頓	
ZFY00594 ⿰扌頓 (no image)	
UTC00440 ⿰氵蕊	
TE-355F ⿰氵蕊 (no image)	
UTC00443 ⿰月枕	
UTC00345 ⿰月枕	
UTC00470 ⿰扌篤	
ZFY00678 ⿰扌篤 (no image)	

UTC00485 ⿰亞⿰厂里 

JK-65014 ⿰亞厘 

UTC00491 ⿰扌 ⿱巛一夕

TC-4927 ⿰扌 ⿱巛歹 (no image)

UTC00506 ⿰冫艮

CYY00128 ⿰冫皂 (no image)

TC-3567 ⿰冫艮 (no image)

(unification:153,78)

UTC00506 ⿰冫艮

TC-3567 ⿰冫艮 (no image)

UTC00519 ⿰口之

WZ100178 ⿰口之 (no image)

UTC00544 ⿰口狽

ZFY00779 ⿰口狽 (no image)

UTC00545 ⿰口庵

TD-493D ⿰口庵 (no image)

UTC00549 ⿰口猜

V04-4368 ⿰口⿰犭青 (no image)

UTC00559 ⿰口宿

JK-65073 ⿰口宿

UTC00562 ⿰口夆

UTC00390 ⿰口夆

UTC00564 ⿰口琼

ZFY00701 ⿰口琼 (no image)

UTC00594 ⿱臼工

TC-2C2F ⿱日工 (no image)

(unification:61a,71-3)

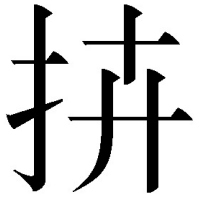
UTC00606 ⿰土夆

WZ100261 ⿰土夆 (no image)

UTC00607 ⿰月关

ZJW00436 ⿰月? (no image)

(unification:15a)

UTC00638 ⿰扌卉 

TC-323B ⿰扌卉 (no image)

UTC00647 ⿰扌庵

CYY00707 ⿰扌庵 (no image)

UTC00655 ⿰扌落

V04-495A ⿰扌⿱艹洛 (no image)

UTC00659 ⿰扌厲

TE-2C3F ⿰扌厲 (no image)

UTC00685 ⿰氵壬

TC-2C57 ⿰氵壬 (no image)

TC-337B ⿰氵玉 (no image)

(unification:129-6)

UTC00698 ⿰火興

TE-3C2C ⿰火興 (no image)

UTC00700 ⿰王扇

TD-506C ⿰王扇 (no image)

UTC00715 ⿰石肯

CYY01046 ⿰石肯 (no image)

MAC00111 ⿰石肯 (no image)

UTC00719 ⿱穴弗

CYY01124 ⿱穴弗 (no image)

UTC00724 ⿰絇糸

UTC00788 ⿰絇糸

UTC00730 ⿰月吞

CYY00723 ⿰月吞 (no image)
(unification:6)

UTC00749 ⿰⻊某

ZFY00761 ⿰⻊某 (no image)

UTC00776 ⿰黑卢

BK102945 ⿰黑卢 (no image)

UTC00788 ⿰絢糸	
UTC00724 ⿰絢糸	
UTC00790 ⿰扌履	
BK100597 ⿰扌履 (no image)	
UTC00835 ⿰火魚	
TD-5F32 ⿰火魚 (no image)	
UTC00836 ⿰土睿	
TE-2159 ⿰土睿 (no image)	
UTC00841 ⿰車鳥	
JK-66055 ⿰車鳥	