**** Formulas ****

The compositional character formulas represent the core of the Jizura project. I became aware of the 漢字データベース project sometime around 2007 or 2008; at the time, being able to download thousands upon thousands of formulas under the GPL license was what really jumpstarted my project.

The first years I spent mainly with sorting out the data, trying to make sense of it, correcting it where I found it to be faulty or problematic, and writing first attempts at software that would help me to handle the data, compile collections of derivative data, make it searchable and so on.

```
u-cjk/4e00 一 ●
u-cjk/4e01 丁 ◓一亅
u-cjk/4e02 丂 ◓一㇆
u-cjk/4e03 七 ⊘一乚
u-cjk/4e04 丄 ◓丨一
u-cjk/4e05 丅 ◓一丨
u-cjk/4e06 丆 ◓一ノ
u-cjk/4e07 万 ◓一勹
u-cjk/4e08 丈 ⊘ナ㇀
u-cjk/4e09 三 ◓一二
u-cjk/4e0a 上 ◓卜一
u-cjk/4e0b 下 ◓一卜
u-cjk/4e0c 丌 ◓一川
u-cjk/4e0d 不 ◔𠂇卜
u-cjk/4e0e 与 ◔㇆一
u-cjk/4e0e 与 ◔◔㇆一一
u-cjk/4e0f 丏 ◔◔丁冂厶
u-cjk/4e10 丐 ◓下㇆
u-cjk/4e11 丑 ◔冂土
u-cjk/4e12 丒 ◓刃一
u-cjk/4e13 专 ⊘二◓㇆丶

[...]

u-cjk/4f9c 俜 ◐亻舟
u-cjk/4f9d 依 ◐亻衣
u-cjk/4f9e 俪 ◐亻如
u-cjk/4f9f 俬 ◐亻存
u-cjk/4fa0 俠 ◐亻夾
u-cjk/4fa1 倆 ◐亻面
u-cjk/4fa2 倈 ◐亻再
u-cjk/4fa3 倡 ◐亻吕
u-cjk/4fa4 倩 ◐亻考
u-cjk/4fa5 倹 ◐亻尧
u-cjk/4fa6 偵 ◐亻贞
u-cjk/4fa7 側 ◐亻则
u-cjk/4fa8 侨 ◐亻乔
u-cjk/4fa9 侩 ◐亻会
u-cjk/4faa 侪 ◐亻齐
u-cjk/4fab 倭 ◐亻妄
u-cjk/4fac 依 ◐亻农
u-cjk/4fad 侭 ◐亻尽
u-cjk/4fae 侮 ◐亻每
u-cjk/4faf 侯 ◐亻◓コ矢
u-cjk/4fb0 偌 ◐亻君
u-cjk/4fb1 偲 ◐亻呈
u-cjk/4fb2 偮 ◐亻辰

[...]

u-cjk/513d 儽 ◐亻纍
u-cjk/513e 儾 ◐亻囊
u-cjk/513f 儿 ◐ノ乚
u-cjk/5140 兀 ◓一儿
u-cjk/5141 允 ◓厶儿
u-cjk/5142 先 〓
u-cjk/5143 元 ◓一兀
u-cjk/5144 兄 ◓口儿

[...]

u-cjk/53ad 厭 ◔厂猒
u-cjk/53ae 厮 ◔厂斯
u-cjk/53af 歷 ◔麻㉿心
u-cjk/53af 歷 ◔麻&cdp#x8962;
u-cjk/53b0 厰 ◔厂敞
u-cjk/53b1 厱 ◔厂僉
u-cjk/53b2 厲 ◔厂萬
u-cjk/53b3 厳 ◔◓ッ厂敢
u-cjk/53b4 厴 ◔厭甲
u-cjk/53b5 厵 ◓原◐原原
u-cjk/53b6 厶 ◐厶丶
u-cjk/53b7 厷 ◔ナム

[...]

u-cjk-xb/20194 覕 ◐享兒
u-cjk-xb/20195 覢 ◐幷夜
u-cjk-xb/20196 覮 (◓亠◐方氏合)
u-cjk-xb/20197 覯 ◓流合
u-cjk-xb/20198 覲 (◓⺗八口一仌)
u-cjk-xb/20199 覴 ◐亭夜
u-cjk-xb/2019a 覺 (◓亠眙死)
u-cjk-xb/2019b 覾 ◓亠曋
u-cjk-xb/2019c 覿 ◓流妻
u-cjk-xb/2019d 覾 (◓亠◐日日八人戊)
u-cjk-xb/2019e 覾 ◐享亶
u-cjk-xb/2019f 豪 (◓亠興彐豕)
u-cjk-xb/201a0 貜 (◓亠⊘且同百灬)
u-cjk-xb/201a1 貜 (◓亠⊘且同杰)
u-cjk-xb/201a2 亽 ▿
u-cjk-xb/201a3 仌 ◓一人
u-cjk-xb/201a4 仐 ◓人亅
u-cjk-xb/201a5 仏 ⊘人ノ
u-cjk-xb/201a6 仚 ◔乚◓人一
u-cjk-xb/201a7 仗 ◐亻人
u-cjk-xb/201a8 仔 ◐亻了
u-cjk-xb/201a9 仉 ◐亻凡

[...]

u-cjk-xd/2b81b 齷 ◐齒星
u-cjk-xd/2b81c 齶 ◐齿兒
u-cjk-xd/2b81d 驁 ◓敞龜

u-cjk-sym/3005 々 ▿
u-cjk-sym/3006 〆 ●
u-cjk-sym/3007 〇 ●
u-cjk-sym/3021 〡 ▿
u-cjk-sym/3022 〢 ◐丨丨
u-cjk-sym/3023 〣 ◐ㅣ丨
u-cjk-sym/3024 〤 ▿
u-cjk-sym/3025 〥 ●
u-cjk-sym/3026 〦 ▿
u-cjk-sym/3027 〧 ◓亠一
u-cjk-sym/3027 〧 ◓、二
u-cjk-sym/3028 〨 ◓亠二
u-cjk-sym/3028 〨 ◓丶三
u-cjk-sym/3029 〩 ◐ノ又
u-cjk-sym/3038 〸 ▿
u-cjk-sym/3039 〹 ▿
u-cjk-sym/303a 〺 ▿
u-cjk-sym/303b 〻 ●
u-cjk-sym/303d 〽 ●

u-cjk-rad1/2f00 一 ▿
u-cjk-rad1/2f01 丨 ▿
u-cjk-rad1/2f02 丶 ▿

[...]

jzr/e100 申 ◓ノ甲
jzr/e100 申 ⊘白ノ
jzr/e100 申 ◓ノ由
jzr/e101 鬼 ◓ノ鬼
jzr/e102 艮 ◐㠯丶

[...]

hzk1/b3d0 &hzk1#xb3d0; ◔厂◐◓土方夕
hzk1/d044 &hzk1#xd044; ◐亻◓隹乃
hzk1/d143 &hzk1#xd143; ◐亻◓廾◔勹用
hzk1/d2ea &hzk1#xd2ea; ◐亻黃
```

```
[...]
```

**** Strokeorders (札字五筆法) ****

```
u-cjk/4e00      一 <1>
jzr/e167    ㇒  <1>
u-cjk/4e8c      二 <11>
u-cjk-xb/2011f 𠀟  <11>
u-cjk-xb/20120 𠄠  <11>
u-cjk/4e09      三 <111>
u-cjk/4e96      三 <1111>
jzr/e15a    丰  <11112>
u-cjk/8a00      言 <1111251>
u-cjk-xb/229d4 𢧔  <1111342511534>
u-cjk/5f0e      弍 <111154>
u-cjk/5f10      弐 <111154>
u-cjk/4e30      丰 <1112>
jzr/e13f    王  <1112>
jzr/e14d    ⺹  <1112>
jzr/e225    丵  <11121112>
u-cjk-xb/234f5 𣓵  <111211121234>
u-cjk-xb/27bee 𧯮  <111211121251431>
u-cjk-xb/277f7 𧟷  <11121112125351>
u-cjk-xa/3ece  爽  <11121112134>
u-cjk-xb/28af9 𨫹  <1112111213434112431>
u-cjk-xb/24acd 𤲍  <11121112134341211243l>
u-cjk-xb/24ad2 𤲒  <1112111213445412121>
u-cjk-xb/24ad2 𤲒  <1112111213445412121>
u-cjk-xb/21689 𡚉  <1112111213445434>
```

**** Factors ****

```
1 一 1             一 1                          一
1 一 1             一 1                          一v1            ㇀
1 一 1             一 1                          一v2   jzr-e167  ㇒

1 一 11            二 11                         二
1 一 11            二 11                         二
1 一 11            二 11                         二
1 一 11            二 111                        三
1 一 11            二 1111                       三
1 一 11            二 1111111111111111   jzr-e24d 𠄠
1 一 11            二 11112              jzr-e15a 丰
1 一 11            二 111112225211       jzr-e2b3 𠦂
1 一 11            二 1112               jzr-e14d ⺹
1 一 11            二 1112               jzr-e13f 王
1 一 11            二 1112                       丰
1 一 11            二 1112                       丰x2h  jzr-e225  丵

[...]

1 一 12  jzr-e21a  亅 12                  jzr-e21a  亅
1 一 12  jzr-e21a  亅 125111152                     𠂤

1 一 12  jzr-e1a6  𠂉 12                  jzr-e1a6  𠂉
1 一 12  jzr-e1a6  𠂉 1215                          弓

1 一 12            丁 12                         丁
1 一 12            丁 12                         丁v1            丅
1 一 12            丁 121                        工
1 一 12            丁 121                        工x4q           𠫓
# 1 一 12           丁 1212             jzr-e2d6  𠂢
1 一 12            丁 12121                      正
1 一 12            丁 12121                      正xlr           玉
1 一 12            丁 12122111           jzr-e21c  耳

[...]

2 ｜ 25            冂 25                         冂
2 ｜ 25            冂 25                         冂v1   jzr-e2e3  冂
2 ｜ 25            冂 25                         冂x2c           冏
2 ｜ 25            冂 251                        冏
2 ｜ 25            冂 251                jzr-e352  回
2 ｜ 25            冂 251                jzr-e365  冏
2 ｜ 25            冂 251                        口
2 ｜ 25            冂 251                        口x2c           回
2 ｜ 25            冂 251                        口x2h           𠙵
2 ｜ 25            冂 251                        口x2v           吕
2 ｜ 25            冂 251                        口x3h           𠙵
2 ｜ 25            冂 251                        口x3p           品
2 ｜ 25            冂 251                        口x4q           𠙵
2 ｜ 25            冂 251                jzr-e2a9  弓
2 ｜ 25            冂 251                jzr-e2aa  冃
2 ｜ 25            冂 251                jzr-e2dd  冊
2 ｜ 25            冂 2511                       冃
2 ｜ 25            冂 2511                       冃
```

2	|	25	冂	2511			冉	
2	|	25	冂	2511			日	
2	|	25	冂	2511			日x2h	昭
2	|	25	冂	2511			日x2v	昌
2	|	25	冂	2511			日x3p	晶
2	|	25	冂	2511			日x4q	曐
2	|	25	冂	2511			曰	

[...]

3	丿	31	𠂉	31			𠂉	
3	丿	31	𠂉	31115			𠂉	
3	丿	31	𠂉	3112			午	
3	丿	31	𠂉	3112			牛	
3	丿	31	𠂉	3112			牛x2v	隼
3	丿	31	𠂉	3112			牛x4q	犇
3	丿	31	𠂉	31121			生	
3	丿	31	𠂉	31121			生x2h	甡
3	丿	31	𠂉	311212			年	
3	丿	31	𠂉	3112121			𠂤	
3	丿	31	𠂉	31122221	jzr-e18d		無	
3	丿	31	𠂉	311234			朱	
3	丿	31	𠂉	31125	jzr-e353		用	
3	丿	31	𠂉	311252	jzr-e175		朱	
3	丿	31	𠂉	311252			缶	
3	丿	31	𠂉	31134			失	
3	丿	31	𠂉	31134			矢	
3	丿	31	𠂉	3115			气	
3	丿	31	𠂉	312	jzr-e13d		𠂉	
3	丿	31	𠂉	312	jzr-e13e		𠂉	
3	丿	31	𠂉	312	jzr-e13d		𠂉x2h	竹
3	丿	31	𠂉	312	jzr-e13e		𠂉x3p	笁
3	丿	31	𠂉	312	jzr-e13e		𠂉x6p	篇
3	丿	31	𠂉	312	jzr-e13e		𠂉x8q	篇
3	丿	31	𠂉	3121			生	
3	丿	31	𠂉	3121			牛	
3	丿	31	𠂉	31211			乍	

[...]

4	丶	41	亠	41			亠	
4	丶	41	亠	4111251			言	
4	丶	41	亠	4111251			言x2h	詰
4	丶	41	亠	4111251			言x2v	喜
4	丶	41	亠	4111251			言x4q	讠
4	丶	41	亠	41121			主	
4	丶	41	亠	412234	jzr-e11a		亦	
4	丶	41	亠	4124			卞	
4	丶	41	亠	4125125251			高	
4	丶	41	亠	413			广	

[...]

5(12)		𠃌	5		𠃌	5			𠃌	
5(12)		𠃌	51	jzr-e346	𠃍	51		jzr-e346	𠃍	
5(12)		𠃌	51	jzr-e346	𠃍	5121			𠃍	
5(12)		𠃌	51	jzr-e350	𠃊	51		jzr-e350	𠃊	
5(12)		𠃌	51	jzr-e350	𠃊	5121121211		jzr-e31e	罡	
5(12)		𠃌	51	jzr-e1c2	𠃎	51		jzr-e1c2	𠃎	
5(12)		𠃌	51	jzr-e1c2	𠃎	5122111		jzr-e21b	耳	
5(12)		𠃌	51	jzr-e1c2	𠃎	5134			央	

5(12)		𠃌	51	jzr-e341	𠃋	51		jzr-e341	𠃋	
5(12)		𠃌	51	jzr-e341	𠃋	511			彐	
5(12)		𠃌	51	jzr-e341	𠃋	511			彐x2v	jzr-e26d 彗
5(12)		𠃌	51	jzr-e341	𠃋	511112			聿	
5(12)		𠃌	51	jzr-e341	𠃋	51112			聿	
5(12)		𠃌	51	jzr-e341	𠃋	511121		jzr-e186	聿	
5(12)		𠃌	51	jzr-e341	𠃋	5112		jzr-e114	尹	
5(12)		𠃌	51	jzr-e341	𠃋	5112			尹	
5(12)		𠃌	51	jzr-e341	𠃋	51121			尹	
5(12)		𠃌	51	jzr-e341	𠃋	5112234		jzr-e115	隶	
5(12)		𠃌	51	jzr-e341	𠃋	51123134			隶	

[...]

5(34)		𠂆	5		𠂆	5			𠂆	
5(34)		𠂆	53	jzr-e1cb	𠂇	53		jzr-e1cb	𠂇	
5(34)		𠂆	53	jzr-e1cb	𠂇	531			女	
5(34)		𠂆	53	jzr-e1cb	𠂇	531			女x2h	奴

```
5(34)                 ㄑ  53  jzr-e1cb  ㄡ  531                        女x2v              姦
5(34)                 ㄑ  53  jzr-e1cb  ㄡ  531                        女x3p              姦
5(34)                 ㄑ  55            ㄍ  55                          ㄍ
5(34)                 ㄑ  55            ㄍ  555                         ㄍ
5(34)                 ㄑ  55            ㄍ  555                         ㄍx6q              ㄍ
5(34)                 ㄑ  55            ㄍ  5551                        ㄍ
5(37)                 丿  5             丿  5                           丿
5(47)                 ㇁  5             ㇁  5                           ㇁
5(47)                 ㇁  5             ㇁  5                           ㇁
5(47)                 ㇁  53  jzr-e27c  ㇁  53          jzr-e27c      ㇁
5(47)                 ㇁  53  jzr-e27c  ㇁  534         jzr-e27a      ㇁
5(47)                 ㇂  5             ㇂  5                           ㇂
5(8)                  〓  5             〓  5                           〓
# ideographic space:
5(8)                      5                 5
5(8)                  ⍁  5             ⍁  5                           ⍁
5(8)                  ⍀  5             ⍀  5                           ⍀
5(8)                  ⍉  5             ⍉  5                           ⍉
5(8)                  ⍋  5             ⍋  5                           ⍋
5(8)                  ⍬  5             ⍬  5                           ⍬
5(8)                  §  5             §  5                        §
5(8)                  ◯  5             ◯  5                           ◯
```

(over 4,000 lines)

垂重熏
鳳鳳
朱羔
炙灸炙
杰乎
堇重焦煮黒黑魚君鳥黒烏厩熏墨鳳嵐喃簟裳贏
体体伾
侀倆
牀牀
侀侀抓抓
坋扮扮松
村材
荧栄荣荥
莞莞
茎荥
芧茎
包包
晋晋普審晉普晋晋竟韶普普普普韶晋普晉韶晋
晋嗇
丧嗇嗇嗇
尚亩面卣茴
茵商胃商
宵宵審
痟疳
音峉音育章兢童童竟音
林粖
涔涔
澄消
茁茁
扙找我牧
我栽戕義蓑戕巃蓑巃簔
宗宗
辰辰辰晨辰
更曳
臭臭
赴赴
垔堊
圇圇
豳鵋鵋
函
詛器器器器器器囂瞴
奧罕型垔垔罟罟罣帚垔器墾墾
蠹

**** Figural Themes ****

龘龘
另另另足疋亞至戾
足⻊克克兇兇
舀宮宮
圩打扑扯扛圩扯任玨玷扦圩扦牡
妅妣
江汢
吖吐
抎扜
址扯
地扡
垚圯
屯圯
屯垔
扒抨
圩扦
扙圩圾扙
圬扬
圩扦扦
亰啻音享辜亶嘗辜富享章京亮塞言冑禦暮
憂啅吊高塋鬻棄嘾嘉奠亹帟豪薨亹蠡蠹
鬲嘾辜壟稟壼褒奠薨骨驀臺蠢膏高毫稟
亨毫毫壼冑章驀豪薨骨驀臺蠢膏高毫稟
毫鬳亮嚞享乏戶戶扃亨享
爾爾
電雷電雷
麃麃
眾衆
阣阣
耋耆鬳鬻鬳鬳
耆耆
耇耇耄
耋耋
斷斷斷
匱匱
滿滿
褻褻
王王玊壬壬
开开
勻刁习勻勺勾句
褰褰
毯麃麛氅羉麩麩纇
箄箄

**** Shape Identity Mappings (SIMs) ****

(22,000 lines)

To keep u-cjk-xb/2011e 二 'archaic form of 下 上' as distinct from u-cjk/4e8c 二 'two', which has the exact same shape, may be justified by the《康熙字典》, which emulates the《説文解字》 in this respect; however, doing so would be incompatible with our 'shapes-only' policy. I do not know what motivated the IRG to encode u-cjk-xb/201a2 𠆢'(人-like, somewhat flattened shape that appears as top component)' as distinct from u-cjk/4eba 人 'man; human'; anyway, u-cjk-xb/201a2 𠆢 does not seem to appear in 《康熙》 and is likewise incompatible with the 'shapes-only' policy.

u-cjk-xb/2011f 二 'archaic form of 下' is in fact (if minimally) distinguishable from both u-cjk-xb/20120 二'archaic from of 二 "two" and u-cjk/4e8c 二 'two', so we keep these codepoints separate—except when writing out character formulas, where we always use u-cjk/4e8c 二 in an attempt to prevent becoming overly specific.

```
u-cjk/4eba 人 u-cjk-rad1/2f08 ⼈
u-cjk/4eba 人 u-cjk-xb/201a2 𠆢
u-cjk/4e8c 二 u-cjk-rad1/2f06 ⼆
u-cjk/4e8c 二 u-cjk-xb/2011e 𠄞
u-cjk/4e8c 二 u-cjk-xb/2011f 𠄟!components
u-cjk/4e8c 二 u-cjk-xb/20120 𠄠!components
```

The amount of duplication in Unicode is certainly surprising to the layman; some of this is due to Unicode's 'roundtrip encoding compatibility with legacy encodings', some is deliberate, and some is due to oversight.

As far as CJK Ideographs are concerned, the two biggest single sources for duplicate codepoints are Compatibility Codepoints and the (I believe misguided) attempt to treat well-known 'radicals' (字典部首) and some characters explicitly labeled as 'symbols' apart from the 'ideographs proper'. To do so, the Unicode Blocks u-cjk-rad1/2f00 … u-cjk-rad1/2fdf *Kangxi Radicals*, u-cjk-rad2/2e80 … u-cjk-rad2/2eff *CJK Radicals Supplement* and u-cjk-sym/3000 … u-cjk-sym/303f *CJK Symbols and Punctuation* have been set up; the latter contains, inter alia, u-cjk-sym/3007 〇 'zero' (to me undoubtely a 'character', just as 一, 二, 三, 萬 are characters, not symbols), duplicated as u-cjk-strk/31e3 ㇣ '(circular stroke, used for some ideographs created in Korea)'.

```
u-cjk-xb/2967f 𩙿 u-cjk-rad2/2ede ⻞
u-cjk/98df 食 u-cjk-rad1/2fb7 ⾷
u-cjk/98df 食 u-cjk-rad2/2edd ⻝
u-cjk/98e0 飠 u-cjk-rad2/2edf ⻟
u-cjk/9963 饣 u-cjk-rad2/2ee0 ⻠
u-cjk/98e0 飠 u-cjk-xb/2967f 𩙿!components
u-cjk/91d1 金 u-cjk-cmpi1/f90a 金
u-cjk/91d1 金 u-cjk-rad1/2fa6 ⾦
u-cjk/91d1 金 u-cjk/91d2 釒!components
u-cjk/8fb6 辶 u-cjk-cmpi1/fa66 辶
u-cjk/8fb6 辶 u-cjk-rad2/2ecc ⻌
u-cjk/8fb6 辶 u-cjk-rad2/2ecd ⻍!components
u-cjk/8fb6 辶 u-cjk-rad2/2ece ⻎!components
u-cjk/5ef4 廴 u-cjk-rad1/2f35 ⼵
u-cjk-sym/3023 〣 jzr-fig/e177 
u-cjk/4e28 丨 u-cjk-sym/3021 〡
```

```
u-cjk/4e28 丨 u-cjk-rad1/2f01 ⼁
u-cjk/4e28 丨 u-cjk-strk/31d1 ㇑
u-cjk/4e85 亅 u-cjk-rad1/2f05 ⼅
u-cjk/4e85 亅 u-cjk-strk/31da ㇚
u-cjk/4e28 丨 u-cjk/4e85 亅!components/search
u-cjk/4ea0 亠 u-cjk-rad1/2f07 ⼇
u-cjk/4ea0 亠 u-cjk-sym/3026 〦
u-cjk/5341 十 u-cjk-rad1/2f17 ⼗
u-cjk/5341 十 u-cjk-sym/3038 〸
u-cjk/535d 卝 u-cjk-cmpi1/fa5d 卄
u-cjk/535d 卝 u-cjk-rad2/2ec0 ⻀
u-cjk-rad2/2ebf ⺿ u-cjk-cmpi1/fa5e 艹
u-cjk/5344 卄 u-cjk/8279 艹
u-cjk/5344 卄 u-cjk-rad2/2ebe ⺾
u-cjk/5344 卄 u-cjk-sym/3039 〹
u-cjk-xb/25ad7 𥫗 u-cjk-rad2/2eae ⺮
u-cjk-xb/2099d 𠦝 u-cjk/9fba 龺
u-cjk-xd/2b740 𫝀 jzr/e14f NO USE
u-cjk-xc/2b1e6 𫇦 jzr/e182 NO USE
u-cjk/6075 恵 u-cjk-cmpi1/fa6b 恵
u-cjk/8218 舘 u-cjk-cmpi1/fa6d 舘
u-cjk/6ed1 滑 u-cjk-cmpi1/f904 滑
u-cjk/5951 契 u-cjk-cmpi1/f909 契
u-cjk/61f6 懶 u-cjk-cmpi1/f90d 懶
u-cjk/7669 癩 u-cjk-cmpi1/f90e 癩
u-cjk/908f 邏 u-cjk-cmpi1/f913 邏
u-cjk/4e82 亂 u-cjk-cmpi1/f91b 亂
u-cjk/6feb 濫 u-cjk-cmpi1/f922 濫
u-cjk/85cd 藍 u-cjk-cmpi1/f923 藍
u-cjk/8964 襤 u-cjk-cmpi1/f924 襤
u-cjk/5eca 廊 u-cjk-cmpi1/f928 廊
u-cjk-xb/20628 𠘨 u-cjk-rad2/2e87 ⺇
u-cjk-xb/206a4 𠚤 u-cjk-sym/3005 々
u-cjk-xb/208de 𠣞 u-cjk-cmpi2/2f9dd 𠣞
u-cjk-xb/21018 𡀘 u-cjk-xb/2103c 𡀼
u-cjk-xb/21d0b 𡴋 u-cjk-cmpi2/2f8f8 𡴋
u-cjk-xb/21f37 𡼷 u-cjk-xb/2439a 𤎚
u-cjk-xb/22331 𢌱 u-cjk-cmpi2/2f891 𢌱
u-cjk-xb/232ad 𣊭 u-cjk-xb/232ab 𣊫
u-cjk-xb/23d1e 𣴞 u-cjk-cmpi2/2f906 𣴞
u-cjk-xb/23f41 𣽁 u-cjk-xb/23f9e 𣾞
u-cjk-xb/23f5e 𣽞 u-cjk-cmpi2/2f910 𣽞
u-cjk-xb/243ab 𤎫 u-cjk-cmpi2/2f91f 𤎫
u-cjk-xb/24425 𤐥 u-cjk-xb/2444b 𤑋
u-cjk-xb/249bc 𤦼 u-cjk-xb/249e9 𤧩
u-cjk-xb/24c36 𤰶 u-cjk-cmpi2/2f935 𤰶
u-cjk-xb/24d14 𤴔 u-cjk-rad2/2eaa ⺪
u-cjk-xb/24fb8 𤾸 u-cjk-cmpi2/2f93c 𤾸
u-cjk-xb/25249 𥉉 u-cjk-cmpi1/fad5 𥉉
u-cjk/9091 邑 u-cjk-rad1/2fa2 ⾢
u-cjk/9149 酉 u-cjk-rad1/2fa3 ⾣
```

漢字データベース

```
jzr/e1ef 馬 cdp/896a &cdp#x896a;
jzr/e1f0 甘 cdp/88c8 &cdp#x88c8;
jzr/e1f1 亞 cdp/85fd &cdp#x85fd;
jzr/e1f8 爾 cdp/8569 &cdp#x8569;
u-cjk/5e80 庀 c2/215a &c2#x215a;
u-cjk/68b5 梵 c1/5b31 &c1#x5b31;
u-cjk/904b 運 c1/672a &c1#x672a;
u-cjk/9db3 鶳 c2/6d34 &c2#x6d34;
u-cjk-xb/26ddf 𦷟 cb/15d5 &cb#x15d5;
u-cjk/5e97 店 c1/4d33 &c1#x4d33;
u-cjk-xa/387b 㡻 c3/2866 &c3#x2866;
u-cjk/7385 玅 c2/2c28 &c2#x2c28;
u-cjk/9d19 鴙 c2/5c39 &c2#x5c39;
u-cjk/6b63 正 c1/465f &c1#x465f;
u-cjk/9117 鄗 c1/673a &c1#x673a;
u-cjk/7380 玀 c1/7b66 &c1#x7b66;
```

yadda yadda

yadda yadda

**** Variants and Usage ****

This file contains around 8,400 records that connect 'variant characters' and 'regions of usage'.

The regions as used here are `C` for the PRC, `J` for Japan, `T` for Taiwan, `H` for Hong Kong, `M` for Macau, and `K` for North and South Korea; this reflects both how these countries and territories are represented in the Unicode Consortium, and, in particular, in the data presented as a result of the Unicode IICore (International Ideographs Core, 國際表意文字核心, 東アジアの諸国で一般に使用される漢字集合)[1] effort.

The term 'variant character' is understood as an umbrella term for what is variously called 俗字, 古字, 略字, 異體字, 簡/繁體字, 新字体 and so on in the traditional literature; no effort has been made to differentiate beyond making simple statements like 'glyph A [which is used in regions X...] is a variant of glyph B [which is used in regions Y...]'. In the software, variant relationships are modeled as both symmetric and transitive[2]; further, there is no temporal aspect whatsoever encoded. All of these assumptions make the display and the handling of the data simpler, but they also oversimplify somewhat.

Here are some samples:

```
01 也CJKTHM  艹  辶  芒
02 鷗JKTHm  鸥C  鴎J
03 飢JKTHm  饑KTHm  饥C
04 個JKTHM  箇JKTm  个CJ  𠂆  仝
05 團JKTHM  糰T  团C  団J  糰
06 紆  紆
07 龜KTHM  龟C  龜J  龝  皺  皸  龜  龜
  龜  龜  龝  皺  龜  龝
08 亞JKTHM  亚C  亜J
09 畝JKTHM  亩C  畝  畆  晦  畝  畝
10 偎cTHm  隈cJt  溳
11 台CJKTHM  颱JKTHM  臺KThM  檯TM  柗th  儓  傘
  彶  允  臺  壇  薹  垥  傘  臺  臺  塏  臺  臺
12 元CJKTHM  圓JKTHM  円JK  圆C  圓
```

Line `01` tells us that 艹, 辶and 芒are variants of 也. Of these, only 也 has a 'usagecode'—`CJKTHM` in this case—attached to it from which we can immediately see that the IICore team did not include any of 艹, 辶, 芒in their listing of important characters; in other words, these glyphs are presumably of minor importance for everyday communication. Conversely, `CJKTHM` implies that the glyph 也 is used in all six regions under consideration.

In line `02`, 鸥 is marked as used in the PRC and 鴎 as used in Japan; the other regions use 鷗, as it is marked `JKTHm`. Both 鴎 and 鷗 show up with `J`, the implication being that both are current in Japan, possibly for different usages (I guess 鷗 is used in names).

[1] https://zh.wikipedia.org/wiki/國際表意文字核心

[2] that is, when 'A is a variant of B' holds, then 'B is a variant of A' also holds, and when additionally 'B is a variant of C' is true, then 'A is a variant of C' is also true.