

## Re: Inverted 2 and inverted 3 in Unicode?

*Source:* <http://sci.tech-archive.net/Archive/sci.lang/2004-09/2708.html>

---

*From:* Tak To ([takto\\_at\\_alum.mit.edu](mailto:takto_at_alum.mit.edu).-)

*Date:* 09/17/04

Date: Fri, 17 Sep 2004 13:20:26 -0400

Tak To wrote:

TT.1> Likewise for hangul blocks. It is not too far fetched to imagine that  
TT.1> computers equipments in the future can compose "on the fly" Chinese  
TT.1> "characters" (with a font of "radicals"),

Harlan Messinger wrote:

HM.2> A table that gave all the rules to combine radicals (including  
HM.2> positioning information, sizing information, modification information,  
HM.2> etc., for each radical) would be entirely font-specific, not  
HM.2> necessarily generalizable from character to character, and therefore  
HM.2> at least as complicated as a table of all the characters, so why would  
HM.2> you do that?

TT.3> If by "rules" you mean detailed inter-radical spacing information then

HM.4> yes;

TT.3> if you mean the actual composition of radicals and rough relative position  
TT.3> then no. For example, the character <lin2> æ — u+6797 ("forest") could be  
TT.3> encoded as <wood radical><right><wood radical> regardless of which  
TT.3> font is being used. Stroke order (and thus radical order) is actually  
TT.3> standardized and taught to school children in PRC and Taiwan alike.

HM.4> What method of defining all characters in terms of their components, no  
HM.4> matter how much sizing and positioning information were to be provided,  
HM.4> would suffice for the character for Mandarin hei(4) ("black") to be rendered  
HM.4> correctly in all the different fonts shown in Figure 12-4 at  
HM.4> [http://www.newworldcider.com/COMM\\_590/article1.htm](http://www.newworldcider.com/COMM_590/article1.htm)  
HM.4> ?

I am not sure what you are looking for. Laying a Chinese character is similar to laying out a web page, so I imagine a reasonable setup would be that there will be a description of the character representable by some sort of markup language, a default "stylesheet" defining a set of reasonably aesthetic proportions (e.g., that the "four\_dots" radical should be half as tall as the component above it), and a collection ("font") of (scalable) radicals and strokes. These three components are more or less independent of each other. In addition, there should probably be

## sci.lang: Re: Inverted 2 and inverted 3 in Unicode?

a provision to override some of the parameters at the generic radical, generic character, or individual (occurrence of) character level (e.g., via additional "stylesheets").

A layout of the <hei> character will be something like:

```
<obj name=#upper
  <obj name=#box1
    <obj name=#box
      <name=#b1 stroke=down>
      <name=#b2 stroke=ne_corner beg=top(#b1) ratio=.5 >
      < stroke=cross beg=bot(#b1) end=bot(#b2) >
    </obj>
    <stroke=\_dot loc=left( in(#box)) >
    <stroke=/_dot loc=right(in(#box)) >
  </obj>
  <name=#v0 stroke=down beg=mid(top(#box1) len=2*ht(#box) >
  <name=#h1 stroke=cross mid=3/4*ht(#v0) wd=narrower(#h2) >
  <name=#h2 stroke=cross mid=bot(#v0) wd=wider(#h1) >
</obj>
<stroke=four_dots loc=below(#upper) >
```

Btw, this is the stroke order that kids are taught to write in.

In practice, the shape formed by the #box1 is probably common enough to be a radical by itself (box\_with\_two\_dots); so are the two horizontal strokes (forming the character for "two"). These two will be abstracted out as radicals[\*]. The layout will be something like:

```
<obj name=#upper>
  <name=box radical=box_with_two_dots >
  <name=two radical=two loc=below(#box) >
  <stroke=down beg=mid_top(#box) end=mid_bot(#two) >
</obj>
<stroke=four_dots loc=below(#upper) >
```

This would describe the layout of a generic <hei>.

[\*] Actually, <hei> is traditionally its own radical.

Just as HTML is not the same as free-format (e.g., it does not allow laying out text along an arbitrary curve), this setup does not support all possible shapes for each character, but will be sufficient for most practical usage. Most of the samples of <hei> in your reference are too stylized to be practical fonts. They are more like dingbats. Still, most of them can be constructed by the setup I described above.

Tak

--

-----+-----

sci.lang: Re: Inverted 2 and inverted 3 in Unicode?

Tak To

takto@alum.mit.eduxx

-----<sup>^^</sup>  
[taode takto ~{LU5B~}] NB: trim the xx to get my real email addr