

Ideographic Variation Sequences

Implementation Details

Ken Lunde — lunde@adobe.com

Senior Computer Scientist, CJKV Type Development

Adobe Systems Incorporated

October 17, 2007

What Is An Ideographic Variation Sequence?

- Base Character + Variation Selector = Glyph
 - Base Character (BC) — Any “CJK Unified Ideograph”
 - Variation Selector (VS) — U+E0100 through U+E01EF (240 VSes in total)
 - Designated VS17 through VS256
 - *Note that U+FE00 through U+FE0F (VS1 through VS16) are not for IVS use!*
- An IVS Resolves To A Glyph
- An IVS Is Registered & Unique
 - There is a registration process in place
- The Power & Safety Of “Plain Text” Representation
 - The ability to survive or endure in more environments

IVS Registration Procedures

- UTS #37: Ideographic Variation Database
 - Authored by Hideki Hiura & Eric Muller
- Ninety-Day Public Review Period
- Variant Forms Shall Adhere To “Annex S” Principles
 - If they do not, or when in doubt, submit as new characters
- Register All Ideographs In Glyph Collection
- For More Information...
 - See: <http://www.unicode.org/reports/tr37/>

“Adobe-Japan1” IVS Registration Details

- The Adobe-Japan1-6 Character Collection
 - 23,058 glyphs — CIDs (Character IDs) 0 through 23057
 - 14,665 of these glyphs are classified as ideographs
- Adobe-Japan1 IVS Registration Submission
 - Data prepared on 11/30/2006 — 14,658 draft IVSes
 - Available as “PRI 98” two weeks later
 - See: <http://www.unicode.org/ivd/pri/pri98/>
- The Ninety-Day Public Review Window
 - T. Kawabata, J. Knightley & T. Suzuki were most helpful

“Adobe-Japan1” IVS Registration Details (Cont’d)

- Semi-Final Form Declared On 09/26/2007
 - 14,637 registered IVSes
 - Currently available as “PRI 108” (through 11/25/2007)
 - See: <http://www.unicode.org/ivd/pri/pri108/>

OpenType Support

- New 'cmap' Subtable — Format 14
 - General-purpose, for Unicode Variation Sequences (UVSes)
 - Differentiates default and non-default UVSes
 - Maps two Unicode code points to a GID (Glyph ID)
 - Base Character + Variation Selector = GID
- Expected To Work With Existing 'cmap' Subtables
 - For example, Format 12 for UTF-32 encoding
- Default UVSes Are Those In Subtable Format 12

Default Versus Non-Default UVSes

- Determined/Distinguished By Format 12 Mappings
- Default UVS
 - The BC resolves to the same GID with or without the VS
 - 13,269 default IVSes for Adobe-Japan1-6 fonts
- Non-Default UVS
 - The UVS is required to represent the GID in plain text
 - 1,368 non-default IVSes for Adobe-Japan1-6 fonts

Default & Non-Default UVS Distribution

- Adobe-Japan1 IVSes for U+8FBB
 - 8FBB E0100; Adobe-Japan1; CID+3056 — JIS90 form
 - 8FBB E0101; Adobe-Japan1; CID+8267 — JIS2004 form
- UniJIS-UTF32-H CMap: U+8FBB → CID+3056 — 辻
 - Default UVS: U+8FBB + U+E0100 = CID+3056
 - Non-Default UVS: U+8FBB + U+E0101 = CID+8267
- UniJIS2004-UTF32-H CMap: U+8FBB → CID+8267 — 辻
 - Default UVS: U+8FBB + U+E0101 = CID+8267
 - Non-Default UVS: U+8FBB + U+E0100 = CID+3056
- 167 Distribution Differences

Base Characters With Many IVSes

- Eight Registered IVSes For Adobe-Japan1-6 U+908A

908A E0100; Adobe-Japan1; CID+6929 — Default IVS

908A E0101; Adobe-Japan1; CID+14235

908A E0102; Adobe-Japan1; CID+14236

908A E0103; Adobe-Japan1; CID+14237

908A E0104; Adobe-Japan1; CID+14238

908A E0105; Adobe-Japan1; CID+14239

908A E0106; Adobe-Japan1; CID+14240

908A E0107; Adobe-Japan1; CID+20234

- Visually:

邊 → 邊邊邊邊邊邊邊

Building IVS-Enabled OpenType Fonts

- AFDKO Version 2.1 — *Available At No Charge*
 - MakeOTF — new “-ci” option to specify UVS definition file
 - Modest increase in ‘cmap’ table size — 26K for all Adobe-Japan1-6 IVSes
 - Spot
- Very First IVS-Enabled Fonts Built On 08/01/2007
 - Kozuka Mincho Pr6N, six weights, Adobe-Japan1-6
 - Version 6.003 (head.FontRevision=6.003)
 - Kozuka Gothic Pr6N, six weights, Adobe-Japan1-6
 - Version 6.003 (head.FontRevision=6.003)

IVS Registration Experiences

- Originally Submitted IVSes — 14,658
 - 14,664 ideographs
 - Seven were previously submitted as new characters
 - One ideograph has two IVSes — CID+19071
- Semi-Final (09/26/2007) IVSes — 14,637
 - One IVS was added — CID+12869
 - Fourteen IVSes were changed
 - Eleven IVSes were re-sequenced
 - Twenty-two IVSes were removed

IVS Registration Experiences (Cont'd)

- Twenty-Nine Glyphs Remain Without IVSes

Seven were previously submitted as new characters

壻煥燭萑薺誦軫

One was discovered in Extension C — CID+14145

寨

Twenty-one to be submitted as new characters

互座勢蓑齧恅嘸囀撈旺樅殲涼瘞盈聃葉鎗艱鯁鹵
【工座勢蓑齧協囀囀撈曜樅殲涼瘞盈職葉鎗養鯁鹵】
【 蓑 殲 鹵】

IVS Registration Learning Experiences

- Adhering To “Annex S” Principles Is Critical
 - UTS #37 should make this point explicit
- To Unify Or Not To Unify...
 - Postpone the registration of borderline cases
 - Submit/propose as new characters
- Reviewer Comments Are Especially Important
- Prepare IDSeS For Future Registrations
 - Ideographic Description Sequences

IVS Infrastructure & Support

- Fonts — *Done!*
 - Via Format 14 'cmap' subtable
- IMEs
 - More intuitive input of variant forms
 - IVSes are “plain text”
- OSes & Libraries
 - Build APIs for applications and IMEs

IVS Infrastructure & Support (Cont'd)

- Applications

- Must gracefully handle IVSes
 - Whether defined in the selected font or not
- *Unsupported IVSes must display as Base Character!!!*
 - This is considered “Step One” for IVS support



Thank You!



Revolutionizing
how the world engages
with ideas and information