

JIS X0213 (IBM CCSID 1394) to UCS MAPPING TABLEs

SOME NOTES ABOUT THE TABLES:

IBM tables are consistent with other existing mapping tables for Japanese, such as Host, EUC, 932/942, so on.

For existing ShiftJIS (CCSID 943) Unicode mapping, IBM has different mappings from MS.

ShiftJIS		UCS by IBM table	UCS by MS table
=====		=====	=====
815C	<====>	2014	2015
815C	<====	2015*	
8160	<====>	301C	FF5E
8160	<====	FF5E*	
8161	<====>	2016	2225
8161	<====	2225*	
817C	<====>	2212	FF0D
817C	<====	FF0D*	

IBM added one way mapping from UCS to ShiftJIS (marked *) for the connectivity with MS application, although IBM's position is above MS mappings are error. IBM mapping follows JIS X0221 standard.

Now, for JISX0213 ShiftJIS (CCSID 1394) Unicode mapping, above U+FF5E, U+2225 and U+FF0D map new characters on JISX0213 as below. Only U+2015 still has two-to-one mapping.

ShiftJIS		UCS by IBM table	UCS by MS table
=====		=====	=====
815C	<====>	2014	2015
815C	<====	2015	
8160	<====>	301C	FF5E
81B0	<====>	FF5E	xxxx
8161	<====>	2016	2225
81D2	<====>	2225	xxxx
817C	<====>	2212	FF0D
81AF	<====>	FF0D	xxxx

What is MS mapping for xxxx ?

=====

Other points of interest can be:

SUB characters:

Single-byte SUB:

Other vendors may have used different SBCS SUB from IBM -- IBM table has x'7F' for SUB. ISO 646/ISO 6429 SUB is x'1A'. (E.g. Microsoft may have used x'3F' question mark.) What do other vendors do?

Double-byte SUB is x'EAA5'.

(SUB characters are used to indicate no mapping to JIS X0213. U+0000 to U+00FF have SBCS equivalents; rest are considered to be DBCS equivalents for purposes of SUB usage.)

IRV and Non-IRV options:

non IRV (Glyph map) option: x'FD' to U+005C and x'FE' to 007E mapping is IBM unique.

IRV (Syntax map) option: x'5C' maps to U+005C and x'7E' maps to U+007E.

Tables (in the zip package):

IBM1394toUCS4-IRV.txt

UCS4toIBM1394-IRV.txt

IBM1394toUCS4-NIRV.txt

UCS4toIBM1394-NIRV.txt