ISO CD 10646 Part-2 vote
- A proposal to move JIS X 0213 Kanji characters on Extension-B into BMP -

IBM Corporation and Microsoft Product Development Ltd.
January 21 2000

Both IBM Corporation and Microsoft Product Development Ltd. s trongly recommend that the 302 JIS X 0213 ideographs included in the CJK Extension B for encoding as part of ISO CD 10646 Part-2 (ISO/JTC 1/SC2 /N3393) not be encoded in Plane 2, but be encoded in the BMP. We recommend disapproval of ISO CD 10646 Part-2 (ISO/JTC 1/SC2 /N3393) unless this change is made.

Based on the expectation that the JIS X 0213 character set will be the primary character set standard in Japan, and that it will be promoted through Unicode, we propose that the 302 JIS X 0213 Kanji characters in CJK Unified Ideographs Extension-B should be relocated to the BMP. To achieve this goal, we propose splitting the CJK Unified Ideographs Extension-B into two components, B1 and B2, where B1 contains all of the ideographs in Extension B except for th e 302 JIS X 0213 Kanji, and B2 contains the 302 JIS X 0213 Kanji Ideographs.
B1 ideographs should be included as planned in 10646 -2 as the CJK Unified Ideographs Extension -B. B2 JIS X 0213 ideographs should be encoded on the BMP.
This proposal documents the detailed reasons for such a measure.

1) Overview of JIS X 0213

JIS X 0213 is new JIS character set standard announced on 20th January 2000, its objective is to develop JIS standard for the characters that should have been defined on JIS X 0208, Coded Character Set to encode contemporary Japanese. That is, the need for JIS X 0213 character is not new and these two JIS X 0213 character sets have almost equal requirements for the following use.

- Japanese publications such as textbooks for elementary and high school students, newspapers, (NTT) phone directories, zip-code directories, and bibliographies.
- Japanese legal directives for compilation of family registers.

Another aspect of JIS X 0213 standard is that it has the reasonable size, total about 10,000, so it fits in the extension of the legacy encodings in Japan such as new Shift_JIS and new EUC code, which have been widely used today on the Internet[*1]. However, it was found that the specification of these two encoding schemes have the compatibility problems with their de-facto encoding, such as Windows Shift_JIS code. Both Shift_JIS extension and EUC extension in JIS X 0213 conflict in the way it assigns the characters on the code points that Windows Shift_JIS and some of major EUC codes reserve for User Definable Characters and assign vendor selected characters. We anticipate corrupted data after any data transfer from the current. They must introduce the garbage characters after the data transfer from current Shift_JIS/EUC to new Shift_JIS/EUC. For example, the code point x'FBE8' on new JIS X 0213 Shift_JIS has the character corresponding to U+995F, while Windows Shift_JIS has the character corresponding to U+9592.
With above reason, when JISC (JIS Committee) reviewed the proposed draft of JIS X 0213 by JC S committee, they modified the specification of Shift_JIS extension and EUC extension from normative standard to informative standard during the review. At that time, *JISC also had the consensus to recommend JIS X 0213 implementation by Unicode, not by either of Shift_JIS or EUC.*

2) Why should JIS X 0213 be encoded in BMP?

---

[1] * Draft RFC (informational) to use X0213 in the MIME context is available at

http://www.ietf.org/internet-drafts/draft-ohta -jcs-JIS X 0213-00.txt

 * ISO-IR announcement: http://anubis.dkuug.dk/jtc1/sc2/open/02i071.txt

Firstly, missing JIS X 0213 Kanji in BMP would be a lost clause of 2.1.A, "*The Basic Multilingual Plane should contain all contemporary characters in common use*" (ISO/IEC JTC 1/SC 2/WG 2 *N2002*).
A Kanji character is not necessarily included in JIS X 0213 if it just appears in one of the sources, for instance, in the largest Kanji dictionary in Japan "Dai Kanwa".
JIS X 0213 sources were collected to cover good number of contemporary Japanese publications and the repertoire was made after through identification of each 'character' (not 'glyph' – see the 2nd criterion –) that has passed the following basic criteria:
 1) Usages are found in the context of contemporary Japanese.
 2) It's being used with a stable shape.
 3) The usages are found on printed materials.
This is the way that JIS X 0213 achieved its design goal - build the repertoire good enough to encode contemporary Japanese.
In other words, from Unicode point of view, the collected characters in JIS X 0213 are the repertoire for those in the world who need to know and/or use contemporary Japanese.

Secondly, as described above, JIS X 0213 character set is expected to be the primary character set standard in Japan and the key IT companies in Japan would like to support JIS X 0213 repertoire through Unicode Standard.
Especially, for the following industry, implementation of JIS X 0213 will immediately be required based on the existing technologies and products for Unicode BMP. The readiness for implementation will affect a lot of business opportunity.
We agree that we must promote UTF-16 and UTF-8, but the industry cannot ignore the existing software and data built with legacy encodings. It will be an extremely difficult and tediously time consuming project to implement JIS X 0213 with non-BMP Unicode[2].

2-1) Mobile applications

Excerpt from Mobile Computing Promotion Consortium (MCPC) web site (http://www.mcpc-jp.org/):
"In the computer-system arena, dramatic advances in technology have enabled client/server systems for downsizing and the Internet and intranet for open systems.

Computer systems typically provide economic and efficient ways to communicate, including electronic mail. Recently, communications technology - especially mobile communications technology - has evolved and prevailed, and the Internet market has expanded dramatically.

Notebook computers become smaller in size and lighter in weight, and hand-held personal computers (HPCs) have emerged. The technological infrastructure for achieving full-scale "mobile computing" is about to be established by which local information-processing can be taken advantage of without limitations of time and location.

These mobile computing systems, effectively integration information-processing and communications technologies, are expected to provide the means for driving the advanced information society of the future, commensurate with users' expanding needs. For mobile computing to evolve full-scale and soundly, it is important for entitled such as communications carriers, computer-hardware manufactures, software developers, and systems integrators to cooperate beyond the existing framework. All elements must address the technical issues, standards promotion, institutional issues, and operational issues."

In Japan, Internet based mobile communication terminals are exploding soon to be number one in the world in terms of the number of units and sophistication of applications. Key words in this area include i-Mode phone of DoCoMo, Windows CE, Zaurus, WorkPad and others. Even if applications in servers is ready to handle UTF-16, mobiles will not follow due to heavy burden of UTF-16 on those small terminals. This implies that if the entire set of JIS X 0213 is not coded with 16-bit fixed length Unicode the majority

---

[2] Some vendors even think that they would use PUA area to support non-BMP X0213 characters rather than to support UTF-16 just only for 302 characters.

of systems in Japan will remain in the Shift JIS scheme for fairly long time in the future leaving Unicode unused on the shelf.

2-2) Several projects planned in the government offices

- IT master plan of general affairs office

  The general affairs office plans to expand its service to open administrative information to public. Access to administrative information will be allowed via web browsers. Employment of JIS X 0213 will be mandated as one of key basic architectures that support the infrastructure of the whole project.

  - Electronic Documents

    As shown in the following excerpt, on the master plan for promoting government-wide use of information technology (http://www.somucho.go.jp/gyoukan/kanri/b_01e.htm), employment of the new JIS X 0213 has been positioned as a solution for digitizing administrative documents. Please note that JIS X 0213 covers the third and fourth levels of JIS.

    "Problems of non-standardized Chinese characters shall be resolved by employing the third and fourth levels of Japan Industrial Standards (JIS), which are now under standardization. "

  - GIS (Geographical Information Systems)

    The master plan also mentions the standardization of data format used in GIS. Since one of major objectives for JIS X 0213 is to define the Kanji characters used in the place name, there will be higher usage for GIS application.

- Resident registration system

  A new resident registration law has been issued last year. A new resident registration network system is being deployed in all local government offices through Japan and the system will be cut over within 3 years. JIS X 0213 will be used as a common character set in the system, since it has been extended to include more Kanji characters of the person name.

- New economic revival plan

  The government announced last November the new economic revival plans with 1800 million dollars to be budgeted for the total projects. They will cover:

  1) IT in educational fields
     All public schools: elementary schools, junior high and high schools, will be connected each other by internet by the year 2001. Computers and internet will be utilized in all classes in the public schools by the year 2005.
  2) IT in local autonomies
  3) e-government
     All administrative paper procedures will be migrated to internet paper free procedures by the year 2003.

  JIS X 0213 will be employed as a common character set on these projects.

2-3) Others

- KyoDo News co.

  KyoDo News company has provided almost all newspaper companies in Japan with common specifications for computers and networks used in the newspaper community. JIS X 0213 will be

adopted as a new character set for future systems.

For the encoding of JIS X 0213 Kanji characters on ISO 10646-1/Unicode, Japan NB of ISO/SC2/WG2 proposed their addition into CJK Extension-B at IRG #13 HongKong meeting after they received the request from JCS Committee. This is based on the assumption that Extension-B is the nearest possibility as the code assignment. However, the primary requirement from JCS Committee is still to encode into BMP. And, recently some of computer makers in Japan (Fujitsu, NEC, IBM Japan, Microsoft Product Development Ltd., and Oki) and JEIDA (Japan Electronic Industry Development Association) officially requested the government standard office (Agency of Industrial Science and Technology, MITI) to support the addition of all JIS X 0213 Kanji characters to be encoded into Unicode BMP as the requirement from the Japanese IT industry.

With JIS X 0213 Kanji characters fully encoded in the BMP, use of Unicode will be expedited in the Japanese IT industry to benefit the user in deployment of e-business/e-commerce in a global scale.

Note:
JIS X 0213 Kanji characters to be added into ISO 10646/Unicode CJK Unified Ideographs are 314 characters according to L2/99-310 (Addition of 314 Kanji from JIS X 0213, by JCS Committee, 1999-08-23) document, however it should be amended to 302 characters based on the recent review.