Date: 1999-09-13

ISO/IEC JTC1/SC2/WG2
Coded Character Set
~~Secretariat: Japan (JISC)~~

Doc. Type: **Proposed disposition of comments**

Title: **Proposed disposition of comments on SC2 WG2 N2012R2 (ISO/IEC WD 10646-2)**

Source: **Michel Suignard (project editor)**
Project: JTC1 02.18.02
Status: For review by WG2
Date: 1999-09-10
Distribution: WG2
Reference: SC2 WG2 N2012R2, SC2 N 3332
Medium:

Comments were received from the Japanese, Swedish, British and US member bodies. The following document is proposing a disposition for those comments. The disposition is organized per country.

Sweden comments (page 4 and 5 of document SC2 N3332):

Technical comment:
The Swedish NB is of the opinion that the "tag characters" shall not be
standardised, at least not at present…

Partially accepted

  a. While it is true that some protocols or languages (like for example HTML or XML) might use higher level to express language tagging, many others use 'text' without any higher-level markup. It was such a need expressed explicitly by the IETF community concerning the IMAP protocol that drove these proposed characters.
  b. The plane 14 is not dedicated to tag characters; other characters may eventually be encoded in that plane. It should also be noted that syntax and usage of tag characters used as language mark is not normative. It was however perceived that the informative Annex C would be of great help for implementers of the tag characters for that purpose. And it would help interoperability between these implementer solutions.
  c. Most formal syntaxes, including those of the ISO standards for language names, have restricted themselves to ASCII only. Furthermore, extending the repertoire beyond ASCII does not offer any great advantage that offsets the foreseeable interoperability problems.

General comments about usefulness of tag characters: Clearly a need has been expressed by the IETF community for tag characters in 'plain' text, especially for the IMAP protocol. In fact, one of the issue has been the speed at which these characters could be accepted by SC2 /WG2 and consequently referenced by

IETF standards. Finally the format and scope of these characters was directly suggested by the requestor community, so it is felt that the current proposal has the best chance to satisfy them.

However as another member body (Japan) has also expressed technical comments concerning the tag characters, some modifications to the WD are suggested in the answer to the Japanese member body. For example additional text will be added to make sure that in most cases (like context using higher level protocol/syntax to express language tagging, e.g. HTML or XML) these tag characters must be filtered out. These modifications also aim at addressing Swedish concerns.

Japan comments (page 2 and 3 of document SC2 N3332):

1. Tagging character and conformance clause.
Accepted. The conformance clause of part-1 does not define 'functionality' for characters. All characters defined in part-2 are graphic characters and as such should conform to part-1 conformance clause. The definition of characters 'functionality' is beyond the scope of ISO 10646, except for few aspects like for example: bi-directionality, combining characters and printable graphical symbols. However the following changes are suggested to ease Japanese concerns:
<<
Clause 1 (definition of GPP, 6th paragraph):
Replace 'special purpose use characters' by 'special purpose use graphic characters'.

Clause 9.3 (General Purpose Plane).
Add the following sentence:
The characters specified in the TAGS block do not have printable graphic symbols, and are thus represented in the character code table by dotted boxes. A possible use of these characters is described in annex C.
>>
A new paragraph will be added to C.1 to describe that typical usage of 10646 (when usage of higher level protocol is used to express language information) should ignore these characters.

1. Note 1. Not relevant to 10646-2

1. Note 2. **Accepted**

2.1 **Accepted**.

2.2 **Accepted**. It does not seem necessary to add a description of tag identification characters in the main text (that is outside annex C). In fact, an earlier draft (WG2 N1717) containing such references was amended to remove them and produce the current version. Finally, an example of language tag identification character is already provided in C.5. However to address the clarification concerns the following changes in annex C are suggested:
   a) Move the second paragraph from C.2 (The tag identification character…to be used.) to become the last paragraph of C.1. This would clearly specify the tag identification character in the overview.
   b) In the C.1 second paragraph insert after "easily identified" the following phrase fragment: "by their coded value".

3.1 **Accepted**. It is clear that the equivalent of part-1 clause 27 should be provided in part-2, probably in clause 9.2 or a new section following it. A mapping to the reference sources will be provided but not in the chart tables themselves but in a different table. Further discussion with ITTF will determine whether these reference tables are made available on paper, on line, both or only on line.

3.2 **Accepted**. Such an example will be provided in the CD.

3.3 **Accepted**. The following text is suggested for Annex B.
<<
The usage of combining characters enumerated in this annex is specified in clause 24 of ISO/IEC 10646-1.

B.1 List of all combining characters

The following characters are combining characters.
<list>

B.2 List of combining characters not allowed in implementation level 2
The following characters are not allowed.

NOTE – This list is a subset of the list in clause B.1.
<list>
>>
The following amendment to part-1 are proposed:
In clause 14, replace 'annex B' by 'annex B of each part of ISO/IEC 10646', same replacement for B.1 and B.2.
In clause 24, same replacement for B.1 and B.2.

3.4 **Accepted**. A new annex equivalent to part-1 annex M will be created. It is really up to the repertoire submitters to provide that information.

3.5 **Accepted**. It is true that the Annex T (btw Annex T should read annex R as now appearing in the current draft of the 2nd edition of 10646-1) cannot be fully referenced by part2 as it has sections (like R.16) that are relevant only to part1. The editor will prepare a partial reference to annex R with minor additions when annex R doesn't apply (like reference sources).

3.6. **Accepted** (as already covered). Clause 12.2 of part-1 is already covering that issue. We just have to make sure that each new part of 10646 has a similar annex A describing its collections.

3-7. **Accepted**. Replace in clause 3 'Part 1' by 'Part 1 of this international standard'. The editor will apply this change to all relevant sections.

3.8. **Accepted** Replace in clause 8 'These characters may not have a visual representation and may not have printable graphic symbols' by 'Some characters do not have a visual representation and do not have printable graphic symbols'.

3.9. **Accepted.** Will add the text

3.10 **Accepted**. All characters specified in ISO/IEC 10646 have semantics, however the standard way to reference this is through their name. No further explanation in annex is necessary.

3.11 **Accepted.**

4.1. **Accepted**. We have been through too much iteration on this.

4.2 **Accepted.** Same as 3.7

4.3 **Accepted**. Replace 'a special plane' by 'another plane'.

4.4 **Accepted**. Use collections Do not see the need.

5.1 **Accepted in principle**, but only for annex A and B, others won't follow that principle.

5.2 **Accepted**. Will add normative text to clause 4 to explain use of UTF16 and UTF8 in the context of Part1.

U.K. comments (page 6 of document SC2 N3332):

Technical comment

1. **Accepted**. The editor will add an annotation to the names (will be: (u), (m) or (l)), with a small explanation in clause 9.1.

2. **Accepted**. The editor will contact the originator to get the information.

3. **Accepted**. The editor will contact the document originator and work with BSI to come up with a satisfactory solution.
4. **Accepted.** See 3.

5. **Accepted**. A list of combining characters will be added to annex B, corresponding to original input document and suggestions from BSI.

Editorial comments: **all accepted**.

U.S.A. comments (page 6 of document SC2 N3332):

1. Reserve the range 0E0000-0E1000 for alternative format characters:
Accepted: A new zone, similar to Part 1 zones will be specified. A clause similar to Part 1 clause 8 will be drafted to define the content and type of behavior of characters (alternative format character) to be included in that zone. The zone range will be 0E000-0E0FFF.

2. Unassigned code points in these ranges should be ignore…
Accepted. Covered by 1

3. Reserve 2060-2069 (BMP) for future format characters (part 1 clause 8)
Accepted. A new collection (2060-206F) will be created and be introduced in the Part1 amendment covering the introduction of Part2. A note covering this request will be proposed in the amendment.

4. Unassigned code points in these ranges should be ignored in normal processing and display.
Accepted. The editor will propose an appropriate text to describe this.
 [end]