ISO/IEC JTC 1/SC 2/WG 3
7-bit and 8-bit codes and their extension
SECRETARIAT : ELOT

DOC TYPE :	Definitions of collections
TITLE :	Re-visiting definitions of collections in COR.2 of 10646
SOURCE :	USA (ANSI) and Canada (SCC)
PROJECT:	JTC 1.02.18.01.00.02
STATUS :	US and Canada - position
ACTION ID :	For the consideration of WG2 and for information to WG3
DUE DATE :	---
DISTRIBUTION :	P, O and L Members of ISO/IEC JTC 1/SC 2 WG Conveners, Secretariats WG 3 Members ISO/IEC JTC 1 Secretariat ISO/IEC ITTF
MEDIUM :	P
NO OF PAGES :	4

Introduction

Corrigendum No. 2 to ISO/IEC 10646:1- 1983 (E) was recently balloted on in JTC 1/SC 2, and most likely has been approved by SC 2. This contribution describes a deficiency in the standard for not being able to define sub-repertoires of 10646 which are *fixed over time*, and proposes that either the current definition of *"dense collection"* be modified or a new *"fixed collection"* be defined.

Reference

(Note : the reference below is the version of DCOR.2 that was circulated to SC 2/WG 2 experts, which was further processed as COR.2 by an SC 2 ballot.)

Ref. ISO/IEC JTC 1/SC 2/WG 2 N1664 : Draft Technical Corrigendum No. 2 to ISO/IEC 10646-1 : 1 1983 (E)

> *4.11 : collection : A set of coded characters which is numbered and named and which consists of those coded characters whose code positions lie within one or more identified ranges.*
>
> > *Note – If any of the identified ranges include code positions to which no character is allocated, the repertoire of the collection will change if an additional character is assigned to any of those positions at a future amendment of this International Standard.*
>
> *4.17 dense collection : A collection in which every code position within the identified range (s) has a character allocated to it.*
>
> > *Note – The repertoire of a dense collection can not be extended at a future amendment of this International Standard unless one or more of the identified ranges of code positions is extended.*

Annex A of 10646 lists a number of sub-repertoires of 10646. The DCOR.2 has marked some of the listed collections (with an "asterisk") as "dense collections" based on the definitions cited above.

The Problem

Among other things, the original request in SC 2/WG 2 N1512, for clarification of the term "collection" in the standard was to remove the ambiguity as to whether the repertoire / sub-repertoire identified by the collection identifier was "fixed" or "variable" over time. Since some of the ranges of code position used to identify some of the collections unassigned code positions, it was ambiguous as to whether a new collection has to be issued when one or more of the unassigned code positions was assigned a character in the future.

SC 2/WG 2 N 1512 proposed the following definitions :

> *4.11 Collection : A set of coded characters which is numbered and named and which consists of those coded characters whose code positions lie within one or more identified ranges.*

Note – If the identified range includes code position that are unassigned, the repertoire of the collection will change if additional characters are assigned to one or more of those positions at a future amendment of this standard.

> *4.12 Specified collection : A collection which contains no code positions reserved for future coding*

The notion of a "specified collection" was to state that the repertoire identified as fixed collections cannot be "changed" – expanded or contracted – over time, whether the expansion is by adding characters to unassigned position within identified ranges or by extending the range of code positions. It identifies a fixed sub-repertoire of 10646.

The notion of a not-specified (or variable) collection was to give a convenient way of referencing a set of related characters with the possibility of being able to add to that repertoire without having to issue a new collection identifier and without any consequences to the users of the collection identifier. The proposal was to re-define "collection" in the standard to permit variability over time.

The redefinition of term "collection" in COR.2 basically defines all collections to be potentially variable. The "dense collection", as is defined, states that within the ranges of code positions identified by the collection identifier, if there are currently unassigned positions, they cannot be included in the identified collection in the future – and to that extent it is fixed over time. However, it leaves the possibility of "adding bye extending the range" and hence it is potentially a "variable" collection. It does not fully meet the requirement for identifying sub-repertoires of 10646 that are fixed over time.

It was unfortunate that we did not catch the nuances of the definitions proposed in document SC 2/WG 2 N1556 at the WG 2 meeting in Crete, and in subsequent ballot on the Corrigendum No.2.

Proposal

Two options are proposed for a further corrigendum to the standard. Option 1 is preferred over option 2, since the current definition of "dense collection" is believed to be really not needed.

Also, note that the words "extended at a future amendment" should be changed to "extended by a future amendment" in both clauses 4.11 and 4.17 in COR.2 (quoted above).

Option 1 :

Redefine the "dense collection" in COR.2 as follows :

> *4.17 : dense collection : A collection in which every code position within the identified range (s) has a character allocated to it*
>
> > *Note – The repertoire of a dense collection can not be extended by a future amendment of this International Standard.*

By deleting the following from the end of the "Note" in 4.17 of COR.2 :

"unless one or more of the identified ranges of code positions is extended"

Option 2 :

Define another type of collection called "fixed collection" as follows:

4.1x: fixed collection : A collection in which every code position within the identified range (s) has a character allocated to it.

Note – The repertoire of a fixed collection can not be changed.