ISO/IEC JTC 1/SC 2/WG 3 N 468

Date : 1998-09-28

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

DOC TYPE :	Officer's Contribution
TITLE :	Alternative to 32-graphic characters in columns 8 & 9 of ISO/IEC 2022
SOURCE :	Bruce Paterson
PROJECT:	JTC 1 02.02
STATUS :	Personal contribution
ACTION ID :	For review by JTC1/SC 2/WG 3
DUE DATE :	---
DISTRIBUTION :	Project editor 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, Def
NO OF PAGES :	4

Contact 1: Secretariat ISO/IEC JTC 1/SC 2/WG 3 ELOT Mrs K.Velli (acting)
Acharnon 313, 111 45 Kato Patissia, ATHENS – GREECE
Tel: +30 1 22 80 001 Fax : +30 1 22 86 219 E-mail : kkb@elot.gr

Contact 2 : Convenor ISO/IEC JTC 1/SC 2/WG 3 Mr E.Melagrakis
Acharnon 313, 111 45 Kato Patissia, ATHENS – GREECE
Tel: +30 1 22 80 001 Fax : +30 1 22 86 219 E-mail: eem@elot.gr

1. Introduction

The code table of a Part of ISO/IEC 8859 shows two empty columns for control functions at the left side of the table, and a further two empty columns near the centre of the table (columns 08 and 09). Could the empty columns 08 and 09 be put to better use, to contain a set of 32 graphic characters? It would increase the repertoire of graphic characters available for information interchange from 191 to 223 characters in a single-byte code, corresponding to the repertoire size of a single code page for PCs.

The most recent proposal on this subject was SC2/WG3 N341, from ELOT; the arguments against it were given in SC2/WG3 N432, from the Canadian National Body. These were discussed by WG3 at the Redmond meeting in March 1998, but no Resolution was taken.

This paper explains that the *interchange* of a 223-character repertoire can already be achieved using the method of ISO/IEC 2022. However the capability for *processing* a 223-character repertoire which exists in PCs is often not provided in other 'processing environments'. Thus, the standardized *interchange* of a 223-character repertoire may not be very useful unless the intended recilients have the capability of *processing* such a repertoire within their applications. This requirement should be considered first, before any changes to ISO standards are proposed.

2.The Interchange code

Standard interchange of characters from a PC code page can be achieved already. No requirement for a single-byte coding in the ISO *interchange environment* has been clearly expressed or justified. So ISO/IEC 2022 can already support a 223-graphic character repertoire in the 'interchange environment', where a single-byte code is not needed (the requirement for a single-byte code applies only to the 'processing environment', and generally the coding there is different from the coding used in the 'interchange environment' – see Annex).

Such a repertoire can use G0, G1 & G2 character sets, and the control single-shift-two (SS2) for the G2 set in accordance with 2022 (or 4873) Level 2. No changes would be required to these standards.

When mapping from a PC code page to a Part of 8859, the G0 set is 'ASCII', but only 96 of the 128 characters in columns 8-F can be mapped intothe G1 set. The other 32 characters are mapped into an additional 96-graphic-character set (G2) in which 64 code positions are unused. These 32-in-96 sets can be thought of as 'supplements' to the 191 characters of a Part of 8859. By convention, the middle two columns of the-6-column graphic character sets would be used for the 32 graphic characters, leaving the other four columns entirely 'reserved'.

One such 32-in-96 character set has already been registered several years ago – ISO-IR 155 Basic Box Drawing Set. This set is intended to supplement the graphic character sets which comprise 8859 Parts 1 or 2, when converting from PC code pages 850 or 852 respectively. Additional sets could be registered as supplements to appropriate Parts of 8859 for use when mapping from other code pages.

3.No Changes to ISO/IEC 2022

Section 2 above shows that no new technical features are required to permit conforming interchange at Level 2 of ISO/IEC 2022. However it would be possible to describe the selection of features needed in this case by defining a new intermediate level-call it level 2R – as follows.

Level 2R in an 8-bit code shall consist of:
- C0 (*32 controls in columns 0 & 1).*
- G0 (*95 graphic characters in columns 2 – 7, typically ASCII).*
- Characters SPACE and DELETE.
- No C1 set (*32 controls in columns 8 & 9).*
- G1 (*96 graphic characters in columns A to F).*
- G2 (96-character set in which only columns 4 & 5 contain graphic characters).
- The shift function SS2 *(to introduce a single character from G2 into columns 4 or 5).*

This new Level could be added to 2022 by means of a short Amendment. This would give confidence to implementors who might consider that the demand in the market would justify provision of support for such a level. Conformity to this level could then be readily identified and claimed for particular implementations.

Annex. Distinction between processing and interchange environments

Single-byte code pages for PCs typically comprise ASCII plus controls in columns 0-7, and a further 128 graphic characters in columns (hex) 8-F. This provides a repertoire of 223 graphic characters in the proprietary single-byte coding used by applications in PCs. The character data is mapped to a different coding for transmission in information interchange conforming to an ISO standard, since the coding structure of columns 8 to F does not conform to ISO/IEC 2022, and the arrangement of graphic characters differs from the arrangement in the corresponding Part of ISO/IEC 8859.

Similarly, when information from a large processor, using the EBCDIC character code for example, is transmitted in ISO standard information interchange, the character data is mapped to the coding of the ISO standard, since the EBCDIC code that is used for processing is qauite different from the ISO code.Applications in this processing environment can usually support a repertoire of only 191 graphic characters if designed for single-byte coding. For this reason the repertoire of the Parts of 8859 was deliberately restricted to 191 graphic characters.

Even for those processors that use ISO coding in their internal processing environment, the protocols between the computer and its associated display devices (local or remote) usually require that columns 8 and 9, as well as columns 0 and 1, are reserved for control functions. Thus again the repertoire of graphic characters is restricted to 191 if the application is designed for single-byte coding.

Clearly, the principal problem of interchange for a repertoire of 223 graphic characters is the capabillity of the ‘processing environments’ and of the protocols for associated display devices. For open interchange of a 223-character repertoire, the following would be needed:

- conversion facilities between ISO coding (see 3 above) and the internal coding of the various ‘processing environmnets’,
- Redesign of selected applications for the processing environments of some types of larger processors, to support a repertoire larger than 191 graphic characters.
- Extension of protocols for associated display devices.

Such changes may be rather substantial and costly to achieve.