ISO/IEC JTC 1/SC 2 N 3342/WG 2 N 2063

Date: 1999-07-26

Supersedes SC 2 N 3273

ISO/IEC JTC 1/SC 2

CODED CHARACTER SETS

SECRETARIAT: JAPAN (JISC)

DOC TYPE: Other document

TITLE: SC 2/WG 2 Charter

SOURCE: ISO/IEC JTC1/SC 2

PROJECT: --

STATUS: In accordance with Resolution M09.02, SC 2 formed an ad hoc group to revise document N 3273 and N 3278 for submittal to JTC 1. The ad hoc group submitted its final draft to the SC 2 members for review by 1999-07-15 via e-mail and comments received have been incorporated in this document. This document is submitted to the JTC 1 Seoul plenary meeting, 1999-11-01/05 in support of the SC 2 Business Plan to be prepared by the Chairman.

ACTION ID: FYI

DUE DATE:

DISTRIBUTION: P, O and L Members of ISO/IEC JTC 1/SC 2
WG Conveners and Secretariats
Secretariat, ISO/IEC JTC 1
ISO/IEC ITTF

NO. OF PAGES: 4

ACCESS LEVEL: Defined

WEB ISSUE #: 058

Contact: Secretariat ISO/IEC JTC 1/SC 2 - Toshiko KIMURA
IPSJ/ITSCJ (Information Processing Society of Japan/Information Technology Standards Commission of Japan)*
Room 308-3, Kikai-Shinko-Kaikan Bldg., 3-5-8, Shiba-Koen, Minato-ku, Tokyo 105-0011 JAPAN
Tel: +81 3 3431 2808; Fax: +81 3 3431 6493; E-mail: kimura@itscj.ipsj.or.jp
*A Standard Organization accredited by JISC

SC2/WG2 CHARTER

Introduction

Coded character sets are the foundation onto which all the modern text-based protocols on the Internet and web are built, and form the essential basis for many other applications of information technology from databases to electronic commerce and publishing. For these reasons they are also one of the fundamental building blocks of the cultural and linguistic adaptability strategy of JTC1.

This document tries to explain the importance of coded character set standards, in particular the Universal Coded Character Set, ISO/IEC 10646, by answering the following questions about it and about the program of work of the Working Group responsible for it - JTC1/SC2/ WG2:

- Who benefits from ISO/IEC 10646?
- What are their requirements for ISO/IEC 10646?
- How can the market relevance of ISO/IEC 10646 be measured?
- How far has ISO/IEC 10646 progressed, and what is left to be done?
- What level of effort will be required to continue the work, and when will the work be completed?
- What development and maintenance procedures are most likely to be successful?

Beneficiaries of ISO/IEC 10646

It is useful to distinguish between the direct users of the standard and its ultimate beneficiaries.
The ultimate beneficiaries of the standard will be end users able to access and exchange textual data from a variety of sources across national and regional boundaries, and in any language.
The direct users of the standard are:

- implementers of information technology and communications systems,
- organisations such as global companies, cultural institutions, and government agencies, including developers of global shopping malls and similar multilingual content providers,
- institutions charged with cataloging and preserving national and regional cultural history,
- developers of other standards dependent on coded character technology, in ISO, CEN, ECMA, IETF, W3C and JTC1.

Requirements for ISO/IEC 10646

The totality of the requirements of these various users can be summarised as follows.

Universal coverage

- All languages in which users conduct day-to-day business, or for which they must collect records,
- Complete coverage of historic scripts for preservation and publication of cultural heritage material.

Single Coherent Coding Architecture

- A single coherent coding architecture holds down costs of product implementations and provides leverage for easy incremental upgrading in step with future extensions of the character repertoire.
- This lowers the barriers to entry into, or to creation of, specialized and niche markets.

Stable, expandable, flexible

- End-users' investment in their encoded data must be preserved over a very long time period, whether commercial data or material relating to world cultural history. Thus the intended life span for 10646 is much longer than for other types of IT standards,
- The standard must allow for easy future expansion to cover all known scripts and notation systems,
- End-users' need easy upgrade of their systems whenever the character repertoire is extended.

Support Interoperability

- The global internet removes physical and national barriers and raises the requirements and opportunities for interoperability of electronic data.
- End-users find severe interoperability problems with today's fragmented solutions.

Measurement of market relevance

ISO/IEC 10646 satisfies all of the requirements described above, and is being developed in close cooperation with the Unicode Consortium, which is a group of market-driven organisations. This is a clear demonstration of its market relevance.

It is already used in key emerging technologies such as the world-wide-web, essential for the world wide integration of IT, with significant benefits to all aspects of world trade and lives of people.

A further demonstration is the high level of participation in the development, including:

- the largest number of NB's for any SC of JTC1, itself the biggest TC of ISO,
- many non-ISO countries and outside experts and scholars,
- many other liaison organizations besides Unicode Consortium.

Achievements

Users and implementations of ISO/IEC 10646 today

ISO/IEC 10646 has been widely adopted since its first published edition in 1993. Implementations include operating systems, applications, internet browsers, programming languages and development tools. In this context, RFC 2277 has established that all future internet protocols must be able to support ISO/IEC 10646.

ISO/IEC 10646 2nd Edition

ISO/IEC 10646-1 and all its amendments and corrigenda will be consolidated and published as the 2nd edition early in the year 2000. This will cover almost all the scripts that are in current use worldwide.

Synchronization with the Unicode Standard

The Unicode Standard, Version 3.0 and the second edition of ISO/IEC 10646 will contain the identical repertoire, and will be published simultaneously. This equivalence and synchronization gives a high level of confidence to all types of users to invest in the adoption of these standards for the long-term future.

Anticipated effort and completion of remaining work

The possible content of ISO/IEC 10646 is large, but it has been fully surveyed by SC2/WG2 and is not unbounded. It can be grouped into characters for the following categories of use:

1. major living languages with established markets
2. minor living languages with emerging markets
3. mathematical and other notations used in scientific and technical endeavors
4. historical languages of interest to religious and scholarly communities
5. other notational systems and scripts needed by specific communities

The number of historic scripts contained in item 4 is well known, as is the approximate number of characters to be included. The overall scale of the encoding to be done is approximately the same as the numbers that have already been encoded. The main effort for historic scripts lies in gathering sufficient, detailed and authoritative information about them, rather than in the number of characters to be encoded.

The current status and estimates for completion are as shown below, where "(-1)" identifies Part 1 of ISO/IEC 10646, and "(-2)" identifies Part 2.

	at 2nd Edition (-1)	extensions (1st priority)	extensions (2nd priority)
1	almost complete	(-1) 2002	(-2) 2004
2	almost complete	(-1) 2002	(-2) 2004
3	partially complete	2004	2008
4	mostly incomplete	2004	2008
5	a few items in progress	per user demand	per user demand

After the 2nd edition is finalised, towards the end of 1999, the frequency and length of WG meetings will be reduced, but work will continue at a steady rate.

Development and maintenance procedure

The development process of ISO/IEC 10646 is by nature additive. Characters will neither be changed, nor renamed or removed once they are coded. Thus an incremental development process, using Amendments, is appropriate. Where user communities such as IETF have time sensitive requirements then an immediate amendment of modest size may be required. The bulk of the work will be less time critical, and will be bundled into collections for inclusion in an occasional large amendment

WG2 adopts a process of distributed and cooperative development, where the original submitters, experts from interested national bodies, the relevant user communities and liaison organizations (particularly the Unicode Consortium) take an active role in resolving many of the technical issues in a given proposal, ensuring that it is mature before it is finally approved by WG2. This cooperative process has proved very

effective, benefitting from the ease and speed with which proposals may be distributed and reviewed via electronic mail and the world-wide web. Meetings of the full membership of WG2 will be needed only infrequently.

Although devolved development and a registration process have recently been proposed, to speed up development and reduce its cost, the experience of implementors in the past shows that registration processes for coded character sets have led to technically incoherent results and fragmented solutions that fail to satisfy the requirements described earlier in this paper. Preservation of the architectural coherence of the standard requires continuing centralized ownership and maintenance by SC2/WG2.

Conclusion

JTC1 considers linguistic and cultural adaptability a strategic cornerstone of IT. The lack of such adaptability is recognized by JTC1 as the ultimate trade barrier. The universal character set, ISO/IEC 10646, is the essential fundamental building block of this cultural and linguistic adaptability. Therefore, WG2 has a special responsibility in meeting the stated requirements of the users of this universal character set to enable them to meet that objective.

-----end of document---