Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 35, *User interfaces*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

© ISO/IEC 2020 – All rights reserved

Introduction

This document defines general mechanisms to specify cultural conventions. It also defines formats for a number of specific cultural conventions in the areas of character classification and conversion, sorting, number formatting, monetary formatting, date formatting, message display, addressing of persons, postal address formatting, and telephone number handling.

The benefits from this document are:

Rigid specification Using this document, a user can rigidly specify a number of the cultural conventions that apply to their information technology environment.

Cultural adaptability If an application has been designed and built in a culturally neutral manner, the application can use the specifications as data to its application programming interfaces (APIs), and thus the same application can accommodate different users in a culturally acceptable way to each of the users, without change of the binary application.

Productivity This document specifies cultural conventions and how to specify data for them. With that data, an application developer is released from getting the different information to support all the cultural environments for the expected customers of the product. The application developer is assured of culturally correct behaviour as specified by the customer, and more markets can potentially be reached as customers can provide the data themselves for markets that were not targeted.

Uniform behaviour When a number of applications share one cultural specification, which may be supplied from the user or provided by the application or operating system, their behaviour for cultural adaptation becomes uniform.

The specification formats are independent of platforms and specific encoding and they are designed to be usable from a wide range of programming languages.

A number of cultural conventions, such as spelling, hyphenation rules and terminology, are not specifiable with this document, but the document provides mechanisms to define new categories and also new keywords within existing categories. An internationalized application can take advantage of information provided with the FDCC-set (such as the language) to provide further internationalized services to the user.

This document defines a format compatible with the one used in ISO/IEC 14651.

This document is upward compatible with elements of ISO/IEC/IEEE 9945, especially those on POSIX locales and charmaps – a locale or charmap conformant to POSIX specifications will also be conformant to specifications in this document, while the reverse condition will not hold. Some of the descriptions are intended to be coded in text files to be used via APIs developed for a number of systems which comply with ISO/IEC/IEEE 9945.

This document has enhanced functionality in a number of areas such as ISO/IEC 10646 support, more classification of characters, transliteration, dual (multi) currency support, enhanced date and time formatting, personal name writing, postal address formatting, telephone number handling, keyboard handling, and management of categories. There is enhanced support for character sets including ISO/IEC 2022 handling and an enhanced method to separate the specification of cultural conventions from an actual encoding via a description of the character repertoire employed. A standard set of values for all the categories has been defined covering the repertoire of ISO/IEC 10646.

 © ISO/IEC 2020 – All rights reserved

This document has been developed to align with ISO/IEC/IEEE 9945. The major extensions from ISO/IEC/IEEE 9945 are listed in Annex A.

A rationale for elements of this document is found in Annex B.

A BNF specification of the syntax for formats in this document is given in Annex C.

The relation to the taxonomy of ISO/IEC TR 24785 is listed in Annex D.

A listing of the implementation of the specifications of this document in the GNU libc compiler product is given in Annex E.

The relation between formats and APIs of this document is listed in Annex F.

A guideline for a method to bind APIs of other programming languages to APIs defined in this document is specified in Annex G.

© ISO/IEC 2020 – All rights reserved

Information technology — Specification methods for cultural conventions

1 Scope

This document specifies description formats and functionality for the specification of cultural conventions, description formats for character sets, and description formats for binding character names to ISO/IEC 10646, as well as a set of default values for some of these items.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 639 (all parts), *Codes for the representation of names of languages*

ISO/IEC 2022, *Information technology — Character code structure and extension techniques*

ISO 3166 (all parts), *Codes for the representation of names of countries and their subdivisions*

ISO 4217, *Codes for the representation of currencies*

ISO 8601, *Date and time — Representations for information interchange*

ISO/IEC 9899, *Information technology — Programming languages — C*

ISO/IEC/IEEE 9945, *Information technology — Portable Operating System Interface (POSIX) Base Specifications, Issue 7*

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

ISO/IEC 14651, *Information technology — International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering*

ISO/IEC 15897:2011, *Information technology — User interfaces — Procedures for the registration of cultural elements*

ISO 15924, *Information and documentation — Codes for the representation of names of scripts*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

© ISO/IEC 2020 – All rights reserved

3.1 Bytes and characters

3.1.1
byte
individually addressable unit of data storage that is equal to or larger than an octet, used to store a character or a portion of a character

Note 1 to entry: A byte is composed of a contiguous sequence of bits, the number of which is implementation defined. The least significant bit is called the low-order bit; the most significant bit is called the high-order bit.

3.1.2
character
member of a set of elements used for the organization, control or representation of data

3.1.3
coded character
sequence of one or more bytes representing a single character

3.1.4
text file
file that contains characters organized into one or more lines

3.2 Cultural and other major concepts

3.2.1
cultural convention
data item for information technology that may vary dependent on language, territory, or other cultural habits

3.2.2
FDCC
formal definition of a cultural convention
cultural convention put into a formal definition scheme

3.2.3
FDCC-set
set of FDCCs
subset of a user's information technology environment that depends on language and cultural conventions

Note 1 to entry: The FDCC-set is a superset of the "locale" term in C and POSIX.

3.2.4
charmap
definition of a mapping between symbolic character names and character codes, plus related information

3.2.5
repertoiremap
definition of a mapping between symbolic character names and characters for the repertoire of characters used in a FDCC-set

Note 1 to entry: This is further described in Clause 7.

 © ISO/IEC 2020 – All rights reserved

3.3 FDCC-related categories

3.3.1
character class
named set of characters sharing an attribute associated with the name of the class

3.3.2
collation
logical ordering of strings according to defined precedence rules

3.3.3
collating element
smallest entity used to determine logical ordering

Note 1 to entry: See collating sequence. A collating element consists of either a single character, or two or more characters collating as a single entity. The LC_COLLATE category in the associated FDCC-set determines the set of collating elements.

3.3.4
multicharacter collating element
sequence of two or more characters that collate as an entity

Note 1 to entry: For example, in some languages two characters are sorted as one letter, as in the case for Danish and Norwegian "aa".

3.3.5
collating sequence
relative order of collating elements as determined by the setting of the LC_COLLATE category in the applied FDCC-set

3.3.6
equivalence class
set of collating elements with the same primary collation weight

Note 1 to entry: Elements in an equivalence class are typically elements that naturally group together, such as all accented letters based on the same letter. The collation order of elements within an equivalence class is determined by the weights assigned on any subsequent levels after the primary weight.

4 Notations

4.1 Notation for defining syntax

In this document, the description of an individual record in a FDCC-set is done using the syntax notation given in the following.

The syntax notation:

"<format>",[<arg1>,<arg2>,...,<argn>]

The <format> is given in a format string enclosed in double quotes, followed by a number of parameters, separated by commas. It is similar to the format specification defined in ISO/IEC/IEEE 9945 and the format specification used in C language printf() function. The format of each parameter is given by an escape sequence:

%s specifies a string

%d specifies a decimal integer

© ISO/IEC 2020 – All rights reserved