TECHNICAL REPORT

ISO/IEC TR 30112

First edition
2014-06-15

Corrected version
2014-08-01

Information technology — Specification methods for cultural conventions

Technologies de l'information — Méthodes de spécification des conventions culturelles

Reference number
ISO/IEC TR 30112:2014(E)

© ISO/IEC 2014



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2014

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**

Introduction ... **v**

1 Scope ... **1**

2 Normative references ... **1**

3 Terms and definitions, and notations ... **1**
3.1 Terms and definitions ... 1
3.2 Notations ... 3

4 FDCC-set ... **6**
4.1 FDCC-set description ... 7
4.2 LC_IDENTIFICATION ... 10
4.3 LC_CTYPE ... 12
4.4 LC_COLLATE ... 40
4.5 LC_MONETARY ... 49
4.6 LC_NUMERIC ... 52
4.7 LC_TIME ... 53
4.8 LC_MESSAGES ... 59
4.9 LC_XLITERATE ... 59
4.10 LC_NAME ... 61
4.11 LC_ADDRESS ... 62
4.12 LC_TELEPHONE ... 64
4.13 LC_PAPER ... 65
4.14 LC_MEASUREMENT ... 66
4.15 LC_KEYBOARD ... 66

5 CHARMAP ... **66**
5.1 Character Set Description Text ... 66
5.2 WIDTH section ... 70

6 REPERTOIREMAP ... **70**

Annex A (informative) **Differences from the ISO/IEC 9945 standard** ... **104**

Annex B (informative) **Rationale** ... **106**

Annex C (informative) **BNF Grammar** ... **117**

Annex D (informative) **Relation to taxonomy** ... **122**

Annex E (informative) **Implementation in glibc** ... **125**

Annex F (informative) **Index** ... **126**

Bibliography ... **135**

© ISO/IEC 2014 – All rights reserved

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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).

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

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, SC 35, *User interfaces*.

This corrected version of ISO/IEC/TR 30112:2014 incorporates the following corrections: the cover page, the header on page 1 and the reference number have been corrected to reflect that this document is a Technical Report, and not an International Standard.

 © ISO/IEC 2014 – All rights reserved

Introduction

This Technical Report defines general mechanisms to specify cultural conventions, and it 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.

There are a number of benefits coming from this Technical Report:

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

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

Productivity: This Technical Report specifies cultural conventions and how to specify data for them. With that data an application developer is relieved from getting the different information to support all the cultural environments for the expected customers of the product. The application developer is thus ensured of culturally correct behaviour as specified by the customer, and possibly more markets may be reached as customers may have the possibility to 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 targeted 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 Technical Report, but the Technical Report provides mechanisms to define new categories and also new keywords within existing categories. An internationalized application may take advantage of information provided with the FDCC-set (such as the language) to provide further internationalized services to the user.

This Technical Report defines a format compatible with the one used in the International string ordering standard, ISO/IEC 14651. This Technical Report is upward compatible with parts of the ISO/IEC 9945 POSIX standard, especially those on POSIX locales and charmaps. The major extensions from that text are listed in annex A. This Technical Report 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 2014 – All rights reserved

Information technology — Specification methods for cultural conventions

1 Scope

This Technical Report 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, plus a set of default values for some of these items.

The specification is upward compatible with POSIX locale specifications - a locale conformant to POSIX specifications will also be conformant to specifications in this Technical Report, while the reverse condition will not hold. Some of the descriptions are intended to be coded in text files to be used via Application Programming Interfaces, that are expected to be developed for a number of systems which comply with ISO/IEC 9945. An alignment effort has been undertaken for this specification to be aligned with ISO/IEC 9945.

2 Normative references

The following referenced documents are indispensable for the application 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.

This document contains no normative references.

3 Terms and definitions, and notations

3.1 Terms and definitions

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

3.1.1 Bytes and characters

3.1.1.1
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

3.1.1.2
character
member of a set of elements used for the organization, control or representation of data

3.1.1.3
coded character
sequence of one or more bytes representing a single character

3.1.1.4
text file
file that contains characters organized into one or more lines

© ISO/IEC 2014 – All rights reserved

3.1.2 Cultural and other major concepts

3.1.2.1
cultural convention
data item for information technology that may vary dependent on language, territory, or other cultural habits

3.1.2.2
FDCC
Formal Definition of a Cultural Convention
cultural convention put into a formal definition scheme

3.1.2.3
FDCC-set
set of Formal Definitions of Cultural Conventions (FDCC's)
definition of the 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.

3.1.2.4
charmap
definition of a mapping between symbolic character names and character codes, plus related information

3.1.2.5
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 6.

3.1.3 FDCC categories related

3.1.3.1
character class:
named set of characters sharing an attribute associated with the name of the class

3.1.3.2
collation
logical ordering of strings according to defined precedence rules

3.1.3.3
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.

3.1.3.4
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".

3.1.3.5
collating sequence
relative order of collating elements as determined by the setting of the LC_COLLATE category in the applied FDCC-set

© ISO/IEC 2014 – All rights reserved