INTERNATIONAL STANDARD

ISO/IEC 14651

Fourth edition
2016-02-15

Corrected version
2016-05-01

Information technology — International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering

Technologies de l'information — Classement international et comparaison de chaînes de caractères — Méthode de comparaison de chaînes de caractères et description du modèle commun et adaptable d'ordre de classement

Reference number
ISO/IEC 14651:2016(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

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 2, *Coded character sets*.

This fourth edition cancels and replaces the third edition (ISO/IEC 14651:2011), which has been technically revised.

This corrected version of ISO/IEC 14651:2016 incorporates the following corrections: in Annex C, Thai characters have been corrected.

Introduction

This International Standard provides a method, applicable around the world, for ordering text data, and provides a Common Template Table which, when tailored, can meet a given language's ordering requirements while retaining reasonable ordering for other scripts.

The Common Template Table requires some tailoring in different local environments. Conformance to this International Standard requires that all deviations from the template, called "deltas", be declared to document resultant discrepancies.

This International Standard describes a method to order text data independently of context.

ISO/IEC TR 30112 has specifications for ordering that informatively complement the specifications in this International Standard and indicates where additional information can be sought on ordering keywords defined in this International Standard.

Information technology — International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering

1 Scope

This International Standard defines the following.

— A reference comparison method. This method is applicable to two character strings to determine their collating order in a sorted list. The method can be applied to strings containing characters from the full repertoire of ISO/IEC 10646. This method is also applicable to subsets of that repertoire, such as those of the different ISO/IEC 8-bit standard character sets, or any other character set, standardized or not, to produce ordering results valid (after tailoring) for a given set of languages for each script. This method uses collation tables derived either from the Common Template Table defined in this International Standard or from one of its tailorings. This method provides a reference format. The format is described using the Backus-Naur Form (BNF). This format is used to describe the Common Template Table. The format is used normatively *within* this International Standard.

— A Common Template Table. A given tailoring of the Common Template Table is used by the reference comparison method. The Common Template Table describes an order for all characters encoded in the Unicode 8.0 standard, included in ISO/IEC 10646:2014 and its Amendment 1. It allows for a specification of a fully deterministic ordering. This table enables the specification of a string ordering adapted to local ordering rules, without requiring an implementer to have knowledge of all the different scripts already encoded in the Universal Coded Character Set (UCS).

NOTE 1 This Common Template Table is to be modified to suit the needs of a local environment. The main worldwide benefit is that, for other scripts, often no modification is required and the order will remain as consistent as possible and predictable from an international point of view.

NOTE 2 The character repertoire used in this International Standard is equivalent to that of the Unicode Standard version 6.0.

— A reference name. The reference name refers to this particular version of the Common Template Table, for use as a reference when tailoring. In particular, this name implies that the table is linked to a particular stage of development of the ISO/IEC 10646 Universal coded character set.

— Requirements for a declaration of the differences (delta) between the collation table and the Common Template Table.

This International Standard does *not* mandate the following.

— A specific comparison method; any equivalent method giving the same results is acceptable.

— A specific format for describing or tailoring tables in a given implementation.

— Specific symbols to be used by implementations, except for the name of the Common Template Table.

— Any specific user interface for choosing options.

— Any specific internal format for intermediate keys used when comparing, nor for the table used. The use of numeric keys is not mandated either.

— A context-dependent ordering.

— Any particular preparation of character strings prior to comparison.

NOTE It is normally necessary to do preparation of character strings prior to comparison even if it is not prescribed by this International Standard (see Annex C).

Although no user interface is required to choose options or to specify tailoring of the Common Template Table, conformance requires always declaring the applicable delta, a declaration of differences with this table. It is recommended that processes present available tailoring options to users.

2 Conformance

A process is conformant to this International Standard if it produces results identical to those that result from the application of the specifications given in 6.2 to 6.5.

A declaration of conformity to this International Standard shall be accompanied by a statement, either directly or by reference, of the following:

— the number of levels that the process supports; this number shall be at least three;

— whether the process supports the forward position processing parameter;

— whether the process supports the backward processing parameter and at which level;

— the tailoring *delta* described in 6.4 and how many levels are defined in the delta;

— if a preparation process is used, the method used shall be declared.

It is the responsibility of implementers to show how their delta declaration is related to the table syntax described in 6.3, and how the comparison method they use, if different from the one mentioned in Clause 6, can be considered as giving the same results as those prescribed by the method specified in Clause 6. The use of a preparation process is optional and its details are not specified in this International Standard.

3 Normative references

The following referenced documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10646:2014, *Information technology — Universal Coded Character Set (UCS)*

ISO/IEC 10646:2014/Amd.1:2015, *Information technology — Universal Coded Character Set (UCS) / Amendment 1: Cherokee supplement and other characters*

