International Standard

ISO/IEC 14651

Seventh edition
2025-07

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 14651:2025(en)

© ISO/IEC 2025



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2025

All rights reserved. Unless otherwise specified, or required in the context of its implementation, 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
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **2**
3 Terms and definitions ... **2**
4 Symbols and conventions ... **3**
5 Conformance ... **3**
6 String comparison ... **4**
6.1 Preparation of character strings prior to comparison ... 4
6.2 Key building and comparison ... 5
6.2.1 Preliminary considerations ... 5
6.2.2 Reference ordering key formation ... 6
6.2.3 Reference comparison method for ordering character strings ... 8
6.2.4 Key ordering definition ... 8
6.3 Common Template Table: Formation and interpretation ... 9
6.3.1 General ... 9
6.3.2 BNF syntax rules for the Common Template Tables in Annex A ... 10
6.3.3 Well-formedness conditions ... 12
6.3.4 Interpretation of tailored tables ... 13
6.3.5 Evaluation of weight tables ... 14
6.3.6 Conditions for considering specific table equivalences ... 15
6.3.7 Conditions for results to be considered equivalent ... 15
6.4 Declaration of a delta ... 15
6.5 Names of the Common Template Tables and name declaration ... 17
Annex A (normative) **Common Template Tables** ... **18**
Annex B (informative) **Example tailoring deltas** ... **20**
Annex C (informative) **Preparation** ... **29**
Annex D (informative) **Tutorial on solutions brought by this document to problems of lexical ordering** ... **45**
Annex E (informative) **Searching and fuzzy matches** ... **49**
Bibliography ... **51**

© ISO/IEC 2025 – 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.

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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had notreceived notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

This seventh edition cancels and replaces the sixth edition (ISO/IEC 14651:2020), which has been technically revised.

The main changes are as follows:

— the Common Template Tables are now referenced externally, removing the need for frequent update to this document.

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 and www.iec.ch/national-committees.

© ISO/IEC 2025 – All rights reserved

Introduction

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

Common Template Tables require some tailoring in different local environments. Conformance to this document requires that all deviations from these templates, called "deltas", be declared to document resultant discrepancies.

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

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

© ISO/IEC 2025 – All rights reserved

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

1 Scope

This document defines a reference comparison method. This method is applicable to two or more 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 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 Tables (CTT) referenced by this document or from one of their tailoring. The format of the Common Template Table is described using the Backus-Naur Form (BNF). The format is used normatively within this document.

This document also defines syntax elements to tailor these Common Template Tables used by the reference comparison method. Furthermore, it defines requirements for a declaration of the differences (delta) between a collation table and a given Common Template Table including the tailoring elements.

These Common Template Tables describe an order for all characters encoded in the current and past ISO/IEC 10646 editions, including amendments. They allow for a specification of a fully deterministic ordering. These tables enable 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).

All these Common Template Tables have reference names which are related to a particular stage of development of the ISO/IEC 10646 Universal coded character set or a particular version of the Unicode Standard. These names and their relationship with ISO/IEC 10646 or the Unicode Standard repertoire are specified by an externally referenced document: Unicode Technical Standard, UTS #10, Unicode Collation Algorithm.

This document does not:

— mandate a specific comparison method; any equivalent method giving the same results is acceptable;

— mandate a specific format for describing or tailoring tables in a given implementation;

— mandate specific symbols to be used by implementations;

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

— mandate a context-dependent ordering;

— mandate any particular preparation of character strings prior to comparison.

NOTE 1 It is typical to do preparation of character strings prior to comparison even if it is not prescribed by this document (see Annex C).

NOTE 2 Annex D describes problems that gave way to this document with their anticipated solutions.

© ISO/IEC 2025 – All rights reserved

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/IEC 10646, *Information technology — Universal coded character set (UCS)*

Unicode Technical Standard. *Unicode Technical Standard #10, Unicode Collation Algorithm:*https://www.unicode.org/reports/tr10/

Common Template Tables. *Unicode Technical Standard UTS #10, Unicode Collation Algorithm, Appendix B: Synchronization with ISO/IEC 14651:*https://www.unicode.org/reports/tr10/#Synch_ISO14651

3 Terms and definitions

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

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

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

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

3.1
character string
sequence of characters considered as a single object

Note 1 to entry: A character string to be ordered does not normally include the characters that delimit it, as for example an “end of line” control character in a text file to be sorted.

3.2
collating symbol
symbol (3.12) used to specify weights assigned to a *collating element* (3.4)

3.3
collation table
weighting table
mapping from *collating elements* (3.4) to *weighting elements* (3.14)

3.4
collating element
sequence of one or more characters that are considered a single entity for *ordering* (3.7)

3.5
delta
list of the differences between a given *collation table* (3.3) and another one

Note 1 to entry: The given collation table, together with a given delta, forms a new collation table.

Note 2 to entry: Unless otherwise specified in this document, the term “delta” always refers to differences from a Common Template Table as defined in this document.

3.6
level
collation level
sequence number for a *subkey* (3.11) in the series of subkeys forming a key

© ISO/IEC 2025 – All rights reserved