

---

---

**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*

Without  
Copyright

Withdrawn



**COPYRIGHT PROTECTED DOCUMENT**

© ISO/IEC 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing 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](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

Foreword .....	v
Introduction.....	vi
<b>1</b> <b>Scope</b> .....	<b>1</b>
<b>2</b> <b>Conformance</b> .....	<b>2</b>
<b>3</b> <b>Normative references</b> .....	<b>2</b>
<b>4</b> <b>Terms and definitions</b> .....	<b>2</b>
<b>5</b> <b>Symbols and abbreviations</b> .....	<b>4</b>
<b>6</b> <b>String comparison</b> .....	<b>4</b>
<b>6.1</b> <b>Preparation of character strings prior to comparison</b> .....	<b>4</b>
<b>6.2</b> <b>Key building and comparison</b> .....	<b>4</b>
<b>6.2.1</b> <b>Preliminary considerations</b> .....	<b>4</b>
<b>6.2.2</b> <b>Reference ordering key formation</b> .....	<b>5</b>
<b>6.2.3</b> <b>Reference comparison method for ordering character strings</b> .....	<b>7</b>
<b>6.3</b> <b>Common Template Table: formation and interpretation</b> .....	<b>8</b>
<b>6.3.1</b> <b>BNF syntax rules for the Common Template Table in Annex A</b> .....	<b>8</b>
<b>6.3.2</b> <b>Well-formedness conditions</b> .....	<b>11</b>
<b>6.3.3</b> <b>Interpretation of tailored tables</b> .....	<b>12</b>
<b>6.3.4</b> <b>Evaluation of weight tables</b> .....	<b>13</b>
<b>6.3.5</b> <b>Conditions for considering specific table equivalences</b> .....	<b>14</b>
<b>6.3.6</b> <b>Conditions for results to be considered equivalent</b> .....	<b>14</b>
<b>6.4</b> <b>Declaration of a delta</b> .....	<b>14</b>
<b>6.5</b> <b>Name of the Common Template Table and name declaration</b> .....	<b>16</b>
<b>Annex A</b> (normative) <b>Common Template Table</b> .....	<b>17</b>
<b>Annex B</b> (informative) <b>Example tailoring deltas</b> .....	<b>18</b>
<b>B.1</b> <b>Example 1 — Minimal tailoring</b> .....	<b>18</b>
<b>B.2</b> <b>Example 2 — Reversing the order of lowercase and uppercase letters</b> .....	<b>18</b>
<b>B.3</b> <b>Example 3 — Canadian delta and benchmark</b> .....	<b>18</b>
<b>B.4</b> <b>Example 4 — Danish delta and benchmark</b> .....	<b>21</b>
<b>B.5</b> <b>Example 5 — A tailoring for Khmer</b> .....	<b>24</b>
<b>Annex C</b> (informative) <b>Preparation</b> .....	<b>26</b>
<b>C.1</b> <b>General considerations</b> .....	<b>26</b>
<b>C.2</b> <b>Thai string ordering</b> .....	<b>26</b>
<b>C.2.1</b> <b>Thai ordering principles</b> .....	<b>27</b>
<b>C.2.2</b> <b>Vowel/consonant rearrangement</b> .....	<b>27</b>
<b>C.2.3</b> <b>Example ordered strings</b> .....	<b>29</b>
<b>C.3</b> <b>Handling of numeral substrings in collation</b> .....	<b>29</b>
<b>C.3.1</b> <b>Handling of ‘ordinary’ numerals for natural numbers</b> .....	<b>30</b>
<b>C.3.2</b> <b>Handling of positional numerals in other scripts</b> .....	<b>32</b>
<b>C.3.3</b> <b>Handling of other non-pure positional system numerals or non-positional system numerals (e.g. Roman numerals)</b> .....	<b>32</b>
<b>C.3.4</b> <b>Handling of numerals for whole numbers</b> .....	<b>33</b>
<b>C.3.5</b> <b>Handling of positive positional numerals with fractional parts</b> .....	<b>34</b>
<b>C.3.6</b> <b>Handling of positive positional numerals with fraction parts and exponent parts</b> .....	<b>35</b>
<b>C.3.7</b> <b>Handling of date and time of day indications</b> .....	<b>36</b>
<b>C.3.8</b> <b>Making numbers less significant than letters</b> .....	<b>37</b>
<b>C.3.9</b> <b>Maintaining determinacy</b> .....	<b>37</b>

- C.4 Preprocessing Hangul.....38
- C.4.1 Step 1 .....38
- C.4.2 Step 2 .....38
- C.4.3 Step 3 .....39
- C.4.4 Step 4 .....39
- Annex D (informative) Tutorial on solutions brought by this International Standard to problems of lexical ordering .....40
- D.1 Problems.....40
- D.2 Solution.....41
- D.3 Tailoring.....42
- Bibliography.....44

Withdrawn

## 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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

ISO/IEC 14651 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

This third edition cancels and replaces the second edition (ISO/IEC 14651:2007), which has been technically revised. It also incorporates the Amendment ISO/IEC 14651:2007/Amd.1:2009.

## 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 14652 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.

Withdrawn

# 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 ISO/IEC 10646:2011. 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 multiple-octet 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 1 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).

NOTE 2 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* 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 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.

ISO/IEC 10646:2011, *Information technology — Universal Coded Character Set (UCS)*