ISO/IEC 14651:2011-08 (E)

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

Contents

Page

Foreword v

Introduction vi

1 Scope 1

2 Conformance 2

3 Normative references 2

4 Terms and definitions 2

5 Symbols and abbreviations 4

6 String comparison 4
6.1 Preparation of character strings prior to comparison 4
6.2 Key building and comparison 4
6.2.1 Preliminary considerations 4
6.2.2 Reference ordering key formation 5
6.2.3 Reference comparison method for ordering character strings 7
6.3 Common Template Table: formation and interpretation 8
6.3.1 BNF syntax rules for the Common Template Table in Annex A 8
6.3.2 Well-formedness conditions 11
6.3.3 Interpretation of tailored tables 12
6.3.4 Evaluation of weight tables 13
6.3.5 Conditions for considering specific table equivalences 14
6.3.6 Conditions for results to be considered equivalent 14
6.4 Declaration of a delta 14
6.5 Name of the Common Template Table and name declaration 16

Annex A (normative) Common Template Table 17

Annex B (informative) Example tailoring deltas 18

B.1 Example 1 -- Minimal tailoring 18
B.2 Example 2 -- Reversing the order of lowercase and uppercase letters 18
B.3 Example 3 -- Canadian delta and benchmark 18
B.4 Example 4 -- Danish delta and benchmark 21
B.5 Example 5 -- A tailoring for Khmer 24

Annex C (informative) Preparation 26

C.1 General considerations 26
C.2 Thai string ordering 26
C.2.1 Thai ordering principles 27
C.2.2 Vowel/consonant rearrangement 27
C.2.3 Example ordered strings 29
C.3 Handling of numeral substrings in collation 29
C.3.1 Handling of `ordinary' numerals for natural numbers 30

C.3.2 Handling of positional numerals in other scripts 32
C.3.3 Handling of other non-pure positional system numerals or non-positional system numerals (e.g. Roman numerals) 32
C.3.4 Handling of numerals for whole numbers 33
C.3.5 Handling of positive positional numerals with fractional parts 34
C.3.6 Handling of positive positional numerals with fraction parts and exponent parts 35
C.3.7 Handling of date and time of day indications 36
C.3.8 Making numbers less significant than letters 37
C.3.9 Maintaining determinacy 37
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