ISO/IEC 14651:2020 (E)

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

Contents

Foreword

Introduction

1 Scope

2 Normative references

3 Terms and definitions

4 Symbols and conventions

5 Conformance

6 String comparison

- 6.1 Preparation of character strings prior to comparison
- 6.2 Key building and comparison
- 6.2.1 Preliminary considerations
- 6.2.1.1 Assumptions
- 6.2.1.2 Processing properties
- 6.2.2 Reference ordering key formation
- 6.2.2.1 General
- 6.2.2.2 Weighting elements to be ignored
- 6.2.2.3 Implicit weights computing
- 6.2.2.4 Subkeys formation
- 6.2.2.5 Formation of subkeys for the first three levels
- 6.2.2.6 Formation of the level 4 subkey
- 6.2.3 Reference comparison method for ordering character strings
- 6.2.4 Key ordering definition
- 6.3 Common Template Table: Formation and interpretation
- 6.3.1 General
- 6.3.2 BNF syntax rules for the Common Template Table in Annex A
- 6.3.2.1 General
- 6.3.2.2 Keyword usage
- 6.3.3 Well-formedness conditions
- 6.3.4 Interpretation of tailored tables
- 6.3.5 Evaluation of weight tables
- 6.3.6 Conditions for considering specific table equivalences
- 6.3.7 Conditions for results to be considered equivalent
- 6.4 Declaration of a delta
- 6.5 Name of the Common Template Table and name declaration

Annex A (normative) Common Template Table

Annex B (informative) Example tailoring deltas

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

Annex C (informative) Preparation

C.1 General considerations
C.2 Thai string ordering
C.2.1 General
C.2.2 Thai ordering principles
C.2.3 Vowel/consonant rearrangement
C.2.4 Example ordered strings
C.3 Handling of numeral substrings in collation
C.3.1 General
C.3.2 Handling of ”ordinary” numerals for natural numbers
C.3.3 Handling of positional numerals in other scripts
C.3.4 Handling of other non-pure positional system numerals or non-positional system numerals (e.g. Roman numerals)
C.3.5 Handling of numerals for whole numbers
C.3.6 Handling of positive positional numerals with fractional parts
C.3.7 Handling of positive positional numerals with fraction parts and exponent parts
C.3.8 Handling of date and time of day indications
C.3.9 Making numbers less significant than letters
C.3.10 Maintaining determinacy
C.4 Pre-processing Hangul
C.4.1 General
C.4.2 Step 1
C.4.3 Step 2
C.4.4 Step 3
C.4.5 Step 4

Annex D (informative) Tutorial on solutions brought by this document to problems of lexical ordering

D.1 General
D.2 Problems
D.3 Solution
D.4 Tailoring

Annex E (informative) Searching and fuzzy matches

Page count: 52