

**Informační technologie - Mezinárodní řazení a porovnání
řetězců - Metoda pro porovnávání znakových řetězců a popis
obecné šablony
pro přizpůsobení řazení**

ČSN
ISO/IEC 14651
36 9169

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

Tato norma je českou verzí mezinárodní normy ISO/IEC 14651:2011. Překlad byl zajištěn Úřadem pro technickou normalizaci, metrologii a státní zkušebnictví. Má stejný status jako oficiální verze.

This standard is the Czech version of the International Standard ISO/IEC 14651:2011. It was translated by the Czech Office for Standards, Metrology and Testing. It has the same status as the official version.

Nahrazení předchozích norem

Touto normou se nahrazuje ČSN ISO/IEC 14651 (36 9169) z března 2003.

Národní předmluva

Změny proti předchozím normám

Třetí vydání zrušuje a nahrazuje druhé vydání (ISO/IEC 14651:2007), které prošlo technickou revizí.

Informace o citovaných dokumentech

ISO/IEC 10646:2011 nezavedena

Související ČSN

ČSN ISO/IEC 2022 Informační technologie - Struktura znakových kódů a metody rozšiřování

ČSN ISO/IEC 646 Informační technika. 7-bitový kódovaný soubor znaků ISO pro výměnu informací

ČSN ISO/IEC 6937 Informační technologie - Kódovaný soubor grafických znaků pro přenos textu -

Latinská abeceda

ČSN ISO/IEC 8859-1 Informační technologie – Jedním 8-bitovým bytem kódované soubory grafických znaků – Část 1: Latinská abeceda č. 1

ČSN ISO/IEC 8859-10 Informační technologie – Jedním 8-bitovým bytem kódované soubory grafických znaků – Část 10: Latinská abeceda č. 6

ČSN ISO/IEC 8859-13 Informační technologie – Jedním 8-bitovým bytem kódované soubory grafických znaků – Část 13: Latinská abeceda č. 7

ČSN ISO/IEC 8859-14 Informační technologie – Jedním 8-bitovým bytem kódované soubory grafických znaků – Část 14: Latinská abeceda č. 8 (keltština)

ČSN ISO/IEC 8859-15 Informační technologie – Jedním 8-bitovým bytem kódované soubory grafických znaků – Část 15: Latinská abeceda č. 9

ČSN ISO 7498-2:1993 Systémy na spracovanie informácií – Prepojenie otvorených systémov (OSI) – Základný referenčný model – Časť 2: Bezpečnostná architektúra

ČSN ISO/IEC 10118-1 (36 9930) Informační technologie – Bezpečnostní techniky – Hašovací funkce – Část 1: Všeobecně

ČSN ISO/IEC 10181-1 (36 9694) Informační technologie – Propojení otevřených systémů – Bezpečnostní struktury otevřených systémů – Část 1: Přehled

ČSN ISO/IEC 10181-4 (36 9694) Informační technologie – Propojení otevřených systémů – Bezpečnostní struktury otevřených systémů – Část 4: Struktura nepopíratelnosti

ČSN ISO/IEC 13888-2 (36 9787) Informační technologie – Bezpečnostní techniky – Nepopíratelnost – Část 2: Mechanismy používající symetrické techniky

ČSN ISO/IEC TR 14516 (36 9791) Informační technologie – Bezpečnostní techniky – Směrnice pro používání a řízení služeb důvěryhodných třetích stran

Vypracování normy

Zpracovatel: Ing. Vladimír Pračke, IČ 40654419

Technická normalizační komise: TNK 20 Informační technologie

Pracovník Úřadu pro technickou normalizaci, metrologii a státní zkušebnictví: Ing. Petr Wallenfels

MEZINÁRODNÍ NORMA

Informační technologie – ISO/IEC 14651

Mezinárodní řazení a porovnávání řetězců – Třetí vydání

Metoda pro porovnávání znakových řetězců a popis 2011-08-15

obecné šablony pro přizpůsobení řazení

ICS 35.040

Obsah

Předmluva 6

Úvod 7

1 Předmět normy 8

2 Shoda 8

3 Citované normativní dokumenty 9

4 Termíny a definice 9

5 Symboly a zkrácené termíny 10

6 Porovnání řetězců 10

6.1 Příprava řetězců znaků před porovnáním 10

6.2 Tvorba klíče a porovnávání 11

6.2.1 Předběžné úvahy 11

6.2.2 Tvorba klíče referenčního řazení 12

6.2.3 Referenční porovnávací metoda pro řazení řetězců znaků 13

6.3 Obecná šablona tabulky: tvorba a interpretace 14

6.3.1 Pravidla BNF syntaxe pro Obecnou šablonu tabulky v příloze A 14

6.3.2 Podmínky dobré tvarovatelnosti 16

6.3.3 Interpretace přizpůsobených tabulek 17

6.3.4 Vyhodnocování váhových tabulek 18

6.3.5 Podmínky, za kterých lze konkrétní tabulky považovat za ekvivalentní 19

6.3.6 Podmínky, za kterých lze výsledky považovat za ekvivalentní 19

6.4 Deklarace odchylky 19

6.5 Název Obecné šablony tabulky a deklarace názvu 20

Příloha A (normativní) Obecná šablona tabulky 21

Příloha B (informativní) Ukázkové odchylky přizpůsobení 22

B.1 Příklad 1 – Minimální přizpůsobení 22

B.2 Příklad 2 – Obrácení pořadí malých a velkých písmen 22

B.3 Příklad 3 – Kanadská odchylka a typová úloha 22

B.4 Příklad 4 – Dánská odchylka a typová úloha 24

B.5 Příklad 5 – Přizpůsobení pro khmerštinu 27

Příloha C (informativní) Příprava 29

C.1 Všeobecné úvahy 29

Strana

C.2 Řazení řetězců v thajštině 29

C.2.1 Zásady řazení thajštiny 29

C.2.2 Změna uspořádání samohlásek/souhlásek 30

C.2.3 Ukázka seřazených řetězců 31

C.3 Ošetření číselných podřetězců při uspořádávání 31

C.3.1 Ošetření 'běžných, číslovek pro přirozená čísla 32

C.3.2 Ošetření pozičních číslovek v jiných písmech 34

C.3.3 Ošetření dalších číslovek ne zcela pozičních systémů nebo číslovek nepozičních systémů (například římské číslovky) 34

C.3.4 Ošetření číslovek pro celá čísla 34

C.3.5 Ošetření kladných pozičních číslovek se zlomkovou částí 36

C.3.6 Ošetření kladných pozičních číslovek se zlomkovou částí a exponentovou částí 36

C.3.7 Ošetření indikací data a denního času 37

C.3.8 Snižování významu čísel oproti písmenům 38

C.3.9 Zachování určitosti 38

C.4 Zpracování písma Hangul 39

C.4.1 Krok 1 39

C.4.2 Krok 2 39

C.4.3 Krok 3 39

C.4.4 Krok 4 40

Příloha D (informativní) Ukázka aplikace této normy na řešení problémů lexikálního řazení 41

D.1 Problémy 41

D.2 Řešení 42

D.3 Přizpůsobování 43

Bibliografie 44



DOKUMENT CHRÁNĚNÝ COPYRIGHTEM

© ISO/IEC 2011

Veškerá práva vyhrazena. Pokud není specifikováno jinak, nesmí být žádná část této publikace reprodukována nebo používána v jakékoliv formě nebo jakýmkoliv způsobem, elektronickým nebo mechanickým, včetně fotokopíí a mikrofilmů, bez písemného svolení buď od organizace ISO na níže uvedené adrese, nebo od členské organizace ISO v zemi žadatele.

ISO copyright office

Předmluva

ISO (Mezinárodní organizace pro normalizaci) a IEC (Mezinárodní elektrotechnická komise) tvoří specializovaný systém celosvětové normalizace. Národní orgány, které jsou členy ISO nebo IEC, se podílejí na vypracování mezinárodních norem prostřednictvím technických komisí ustavených příslušnými organizacemi pro jednotlivé obory technické činnosti. Technické komise ISO a IEC spolupracují v oborech společného zájmu. Práce se zúčastňují také další vládní a nevládní mezinárodní organizace, s nimiž ISO a IEC navázaly pracovní styk. V oblasti informační technologie zřídily ISO a IEC společnou technickou komisi ISO/IEC JTC 1.

Návrhy mezinárodních norem jsou vypracovávány v souladu s pravidly danými směrnicemi ISO/IEC, část 2.

Hlavním úkolem společné technické komise je vypracování mezinárodních norem. Návrhy mezinárodních norem přijaté společnou technickou komisí jsou rozesílány národním členům k hlasování. Vydání mezinárodní normy vyžaduje souhlas alespoň 75 % hlasujících národních orgánů.

Upozorňuje se na možnost, že některé prvky tohoto dokumentu mohou být předmětem patentových práv. ISO a IEC nelze činit odpovědné za identifikaci jakéhokoliv nebo všech patentových práv.

ISO/IEC 14651 vypracovala společná technická komise ISO/IEC JTC 1, *Informační technologie*, subkomise SC 2 *Kódované soubory znaků*.

Třetí vydání zrušuje a nahrazuje druhé vydání (ISO/IEC 14651:2007), jehož je technickou revizí. Zpracovává také změnu ISO/IEC 14651:2007/Amd.1:2009.

Úvod

Tato mezinárodní norma poskytuje celosvětově použitelnou metodu pro řazení textových dat a poskytuje Obecnou šablonu tabulky (Common Template Table), která, je-li přizpůsobena, může splňovat dané požadavky na jazykové řazení, přičemž zachovává přiměřené řazení pro jiné druhy písma.

Obecná šablona tabulky vyžaduje jistá přizpůsobení v různých lokálních prostředích. Shoda s touto mezinárodní normou vyžaduje, aby všechny odlišnosti od šablony, nazývané „odchytky“, které jsou deklarovány, dokumentovaly výsledné nepravidelnosti.

Tato mezinárodní norma popisuje metodu řazení textových dat nezávisle na kontextu.

ISO/IEC TR 14652 obsahuje specifikace pro řazení, které informativně doplňuje specifikace v této mezinárodní normě a ukazuje, kde mohou být nalezeny doplňující informace o klíčových slovech pro řazení definovaných v této mezinárodní normě.

1 Předmět normy

Tato mezinárodní norma řeší následující problematiku.

- Referenční srovnávací metodu. Tato metoda je aplikovatelná na dva řetězce znaků k určení jejich relativního pořadí. Metoda může být aplikována na řetězce, které obsahují úplný soubor znaků z ISO/IEC 10646. Tato metoda je rovněž aplikovatelná na podmnožiny takového souboru znaků, jako jsou různé 8bitové soubory normalizovaných znaků v ISO/IEC nebo jakýkoliv jiný normalizovaný nebo nenormalizovaný soubor znaků, který slouží k vytvoření výsledného řazení, které je platné (po přizpůsobení) pro danou množinu jazyků a pro každé písmo. Tato metoda používá tabulky uspořádání odvozené buď z Obecné šablony tabulky popsané v této mezinárodní normě, nebo z jednoho z jejích přizpůsobení. Tato metoda poskytuje referenční formát. Tento formát je popsán užitím Backus-Naurovy formy (BNF). Tento formát je použit k popisu Obecné šablony tabulky. Formát je použit normativně v rámci této mezinárodní normy.
- Obecnou šablonu tabulky. Dané přizpůsobení Obecné šablony tabulky je používáno referenční porovnávací metodou. Obecná šablona tabulky popisuje pořadí všech znaků kódovaných v ISO/IEC 10646:2011. To dovoluje stanovení plně deterministického řazení. Tato tabulka umožňuje stanovit řazení řetězců přizpůsobené místním pravidlům řazení, aniž by vyžadovala, aby implementátor měl znalosti všech různých písem již kódovaných v Univerzální kódovaný soubor znaků (UCS).

POZNÁMKA 1 Tuto Obecnou šablonu tabulky je třeba upravit tak, aby vyhovovala potřebám místního prostředí. Hlavní užitek v celosvětovém měřítku je, že pro další písma často není vyžadována další úprava a pořadí zůstane z mezinárodního hlediska co nejshodnější a nejpředvídatelnější.

POZNÁMKA 2 Soubor znaků použitý v této mezinárodní normě je ekvivalentní k standardu Unicode verze 6.0.

- Referenční název. Referenční název odkazuje na určitou konkrétní verzi Obecné šablony tabulky pro použití jako reference při přizpůsobení. Tento název zejména naznačuje, že tabulka je spojena s určitým stádiem vývoje ISO/IEC 10646 Univerzální víceoktetový kódovaný soubor znaků.
- Požadavky na vyhlášení rozdílů (odchylky) mezi porovnávací tabulkou a Obecnou šablonou tabulky.

Tato mezinárodní norma *nenařizuje* následující.

- Konkrétní porovnávací metodu; jakákoliv ekvivalentní metoda, která dává stejné výsledky, je přijatelná.
- Konkrétní formát pro popis a přizpůsobení tabulek v dané implementaci.
- Konkrétní symboly, které mají být použity v implementacích, kromě Obecné šablony tabulky.
- Jakékoliv konkrétní uživatelské rozhraní pro výběr možností.
- Jakýkoliv konkrétní interní formát pro přechodné klíče použité při porovnávání, jakož ani pro použitou tabulku. Rovněž nenařizuje použití číselných klíčů.
- Kontextově závislé řazení.
- Jakoukoliv konkrétní přípravu řetězců znaků před porovnáním.

POZNÁMKA 1 Přípravu řetězců znaků před porovnáním je obvykle nutno provést, i když to není touto mezinárodní normou předepsáno (viz příloha C).

POZNÁMKA 2 Ačkoliv k volbě možností nebo stanovení přizpůsobení Obecné šablony tabulky není vyžadováno uživatelské prostředí, shoda vždy vyžaduje deklarování vhodné odchylky, tedy deklaraci rozdílů od této tabulky. Doporučuje se, aby procesy předkládaly uživatelům dostupné možnosti přizpůsobení.

Konec náhledu - text dále pokračuje v placené verzi ČSN.