

	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 d'ordre de classement

Informationstechnik - Internationale Reihenfolge und Vergleich - Methode für Zeichenreihenvergleich und Beschreibung von gemeinsam festgelegten Reihenfolgen

Tato norma je českou verzí mezinárodní normy ISO/IEC 14651:2001. Mezinárodní norma ISO/IEC 14651:2001 má status české technické normy.

This standard is the Czech version of the International Standard ISO/IEC 14651:2001. The International Standard ISO/IEC 14651:2001 has the status of a Czech Standard.

Národní předmluva

Citované normy

ISO/IEC 10646-1:1993, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1:2002 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň

ISO/IEC 10646-1:1993/Amd. 1:1996, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Dodatek 1: Formát transformace pro 16 úrovně skupiny 00 (UTF-16).

ISO/IEC 10646-1:1993/Amd. 2:1996, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Amendment 2: UCS Formát transformace 8 (UTF-8).

ISO/IEC 10646-1:1993/Amd. 4:1996, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Dodatek 4.

ISO/IEC 10646-1:1993/Amd. 5:1998, nahrazena ISO/IEC 10646-1:2000 zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Amendment 5: Slabiky korejského písma (Hangul).

ISO/IEC 10646-1:1993/Amd. 6:1997, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Dodatek 6: Tibetština.

ISO/IEC 10646-1:1993/Amd. 7:1997, nahrazena ISO/IEC 10646-1:2000 zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Dodatek 7: 33 dodatečných znaků.

ISO/IEC 10646-1:1993/Amd. 9:1997, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Dodatek 9: Identifikátory pro znaky.

ISO/IEC 10646-1:1993/Amd.18:1999, nahrazena ISO/IEC 10646-1:2000, zavedena

v ČSN ISO/IEC 10646-1 (36 9143) Informační technologie - Universální víceoktetový kódovaný soubor znaků (UCS) - Část 1: Architektura a základní vícejazyčná úroveň Dodatek 18: Symboly a jiné znaky.

Vypracování normy

Zpracovatel: RNDr. Vratislav Datel, CSc., IČO 47572507

Technická normalizační komise: TNK 20 Informační technologie

Pracovník Českého normalizačního institutu: Ing. Petr Wallenfels

Strana 3

MEZINÁRODNÍ NORMA

Informační technologie -

ISO/IEC 14651

Mezinárodní řazení a porovnání řetězců -
vydání

První

**Metoda pro porovnávání znakových řetězců a popis
obecné šablony pro přizpůsobení řazení**

2001-02-15

ICS 35.040

Odmítnutí odpovědnosti za manipulaci s PDF

Tento PDF soubor obsahuje vložené typy písma (normální, kurzivní, tučné atd.). V souladu s licenční politikou Adobe tento soubor lze tisknout nebo prohlížet, avšak nesmí být editován, kromě těch vložených typů písma, které nepodléhají licenci a jsou instalovány v počítači, na kterém se editace provádí. Při používání tohoto souboru jsou jeho uživatelé odpovědní za to, že nebude porušena licenční politika Adobe. Ústřední sekretariát ISO nepřejímá za její porušení žádnou odpovědnost.

Adobe je obchodní značka „Adobe Systems Incorporated“.

Podrobnosti o softwarovém produktu, který vytváří tento PDF soubor, lze najít ve všeobecných informacích, které jsou k němu přiloženy; parametry na vytváření PDF jsou optimalizovány pro tisk.

Soubor je upraven tak, aby byl použitelný členskými organizacemi ISO. V případě nepravděpodobné události a problému, který se k ní vztahuje, informujte Ústřední sekretariát ISO. Jeho adresa je uvedena níže.

Všechna práva vyhrazena. Žádná část této normy nesmí být reprodukována nebo zpracována jakoukoli jinou formou, jako jsou například elektronické, mechanické prostředky, včetně fotokopí a mikrofilmu bez písemného povolení ISO, povolení lze vyžádat na níže uvedené adrese nebo u členské národní organizace v zemi žadatele.

ISO copyright office

Obsah

Strana

Předmluva

.....
..... 6

Úvod

.....
..... 7

1 Předmět normy

.....
.. 8

2 Shoda

.....
..... 8

3 Normativní odkazy

..... 9

4 Termíny a definice

..... 10

5 Symboly a zkratky

.....
10

6 Porovnávání řetězců

..... 11

6.1 Příprava znakových řetězců před porovnáním.....

11

6.2 Tvorba klíče a porovnání.....

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: utváření a interpretace..... 14
6.3.1	Pravidla syntaxe BNF pro Obecnou tabulku šablony v Příloze A..... 14
6.3.2	Dobrá tvarovatelnost podmínek..... 17
6.3.3	Interpretace přizpůsobených tabulek..... 18
6.3.4	Vyhodnocení váhových tabulek..... 20
6.3.5	Podmínky, za kterých lze dvě tabulky považovat za ekvivalentní..... 20
6.3.6	Podmínky, za kterých lze výsledky porovnání považovat za ekvivalentní..... 20
6.4	Deklarace odchylky..... 20
6.5	Název Obecné šablony tabulky a deklarace názvu..... 22
Příloha A	(normativní) Obecná šablona tabulky..... 23
Příloha B	(informativní) Příklad..... 24
B.1	Příklad 1 - Minimální přizpůsobení..... 24
B.2	Příklad 2 - Obrácené pořadí malých a velkých písmen..... 24

B.3	Příklad 3 - Kanadská odchylka a typová úloha	24
B.4	Příklad 4 - Dánská odchylka a typová úloha	27
Příloha C (informativní)		
	Příprava	31
C.1	Všeobecné úvahy	31
C.2	Řazení řetězce v thajštině - příprava zahrnující písmo vyžadovaná pro získání správného řazení	31
C.2.1	Zásady řazení thajštiny	31
C.2.2	Algoritmické hledisko	33
C.3	Ošetření číselných podřetězců při uspořádávání	34
C.3.1	Ošetření „řadových“ číslovek pro přirozená čísla	34
C.3.2	Ošetření pozičních číslovek v jiných písmech	37
C.3.3	Ošetření dalších číslovek z nečistých pozičních systémů nebo číslovek z nepozičních systémů	37
C.3.4	Ošetření číslovek pro celá čísla	37
C.3.5	Ošetření kladných pozičních číslovek se zlomkovou částí	38
C.3.6	Ošetření kladných pozičních číslovek se zlomkovou a exponentovou částí	39
C.3.7	Ošetření indikací data a denního času	40

C.3.8 Vytváření čísel s menším významem než písmena.....	41
C.3.9 Udržování určitosti.....	41
Příloha D (informativní) Ukázka aplikace této normy na řešení problémů lexikálního řazení.....	42
D.1 Problémy.....	43
D.2 Řešení.....	43
D.3 Přizpůsobení.....	44
Bibliografie.....	46

Předmluva

ISO (Mezinárodní organizace pro normalizaci) a IEC (Mezinárodní elektrotechnická komise) spolu tvoří celosvětový normalizační systém. Národní instituce, které jsou členy ISO a IEC, se účastní prací na mezinárodních normách prostřednictvím svých technických komisí ustavených těmito institucemi tak, aby reprezentovaly jednotlivé technické obory. Technické komise ISO a IEC spolupracují v oblastech společného zájmu. Další mezinárodní organizace, vládní i nevládní, se ve spolupráci s ISO a IEC také účastní těchto prací.

Mezinárodní normy jsou připravovány v souladu s pravidly uvedenými ve směrnících v části 3 ISO/IEC.

V oblasti informační technologie ISO a IEC zřídily společný technický výbor ISO/IEC JTC 1. Návrhy mezinárodních norem přijaté technickými komisemi se rozesílají členům ISO k hlasování. Vydání mezinárodní normy vyžaduje souhlas nejméně 75 % hlasujících členů.

Upozorňujeme na možnost, že některé prvky obsažené v této části ISO/IEC 13249 mohou být předmětem patentových práv. ISO a IEC nesmí být činěno zodpovědným za identifikaci kteréhokoliv nebo všech takových patentových práv.

Mezinárodní norma ISO/IEC 14651 byla připravena technickou komisí ISO/IEC JTC 1 *Informační technologie*, subkomisí SC 22, *Programovací jazyky, jejich prostředí a rozhraní systémového software*.

Příloha A tvoří normativní část této mezinárodní normy. Přílohy B, C a D jsou pouze informativní.

Strana 7

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 “odchyly”, které jsou deklarovány, dokumentovaly výsledné nepravidelnosti.

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

Návrh technické zprávy ISO/IEC DTR 14652 (ve vývoji) obsahuje specifikace pro řazení, které informativně doplňuje specifikace v této mezinárodní normě a kde doplňující informace mohou být nalezeny na klíčových slovech pro řazení definovaných v této mezinárodní normě.

Strana 8

1 Předmět normy

Tato mezinárodní norma obsahuje následující definice.

- Referenční srovnávací metody. 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 10646-1. Tato metoda je rovněž aplikovatelná na podmnožiny takového souboru znaků, jako jsou různé 8-bitové soubory normalizovaných znaků v ISO/IEC nebo na jakýkoliv 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 buď odvozené z Obecné šablony tabulky popsané v této mezinárodní normě, nebo jedno z jejích přizpůsobení.
- Referenčního formátu. 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.

- Obecné šablony tabulky. Referenční porovnávací metoda používá dané přizpůsobení Obecné šablony tabulky. Obecná šablona tabulky popisuje pořadí pro všechny znaky kódované v prvním vydání ISO/IEC 10646-1 až do Dodatku č. 7. To umožňuje specifikaci plně deterministického řazení. Tato tabulka umožňuje přizpůsobit specifikaci řazení řetězců místním pravidlům řazení bez požadavku, aby implementátor měl znalosti všech různých písem již kódovaných v UCS.

POZNÁMKA 1 Obecnou šablonu tabulky lze modifikovat, aby vyhovovala lokálnímu prostředí. Hlavní užitek v celosvětovém měřítku je, že pro další písma často nemusí být vyžadována modifikace a že pořadí zůstane tak konzistentní, jak je možné a předpokládané z mezinárodního hlediska.

POZNÁMKA 2 Soubor znaků použitý v této mezinárodní normě je ekvivalentní s Unicode Standard verze 2.1.

- Referenčního názvu. Referenční název odkazuje na určitou verzi Obecné šablony tabulky pro použití jako reference při přizpůsobení. Speciálně, z tohoto názvu zvláště vyplývá, že tabulka je spojena s určitou etapou vývoje ISO/IEC 10646 Universal multiple-octet coded character set.
- Požadavků pro deklaraci rozdílů (odchylka) mezi porovnávací tabulkou a Obecnou šablonou tabulky.

Tato mezinárodní norma neopravňuje přikazovat:

- určitou porovnávací metodu; jakákoliv ekvivalentní metoda, která dává stejné výsledky je přijatelná;
- specifický formát pro popisování nebo přizpůsobování tabulek v dané implementaci;
- specifické symboly mají být použity implementacemi kromě názvu CommonTemplate Table;
- jakékoli specifické uživatelské rozhraní pro volitelné možnosti;
- jakýkoliv specifický vnitřní formát pro přechodné klíče použité při porovnávání, ani pro použitou tabulku a nenařizuje použití číselných klíčů
 - kontextově závislé řazení;
 - jakoukoliv přípravu řetězců znaků před porovnáním.

POZNÁMKA 1 V praxi je nutné provést přípravu řetězců znaků před porovnáním dokonce, i když to není předepsáno touto mezinárodní normou (viz informativní příloha C).

POZNÁMKA 2 Ačkoliv k volbě možností nebo specifickému přizpůsobení Obecné šablony tabulky není požadováno žádné uživatelské rozhraní, shoda vždy vyžaduje deklarování aplikovatelné odchylky, deklaraci rozdílů od této tabulky. Uživatelům se doporučuje, aby zpracovali použitelné možnosti přizpůsobení.

-- Vynechaný text --