

ČESKÁ TECHNICKÁ NORMA

ICS 01.140.20; 35.040 **Říjen 2011**

Evropská pravidla řazení - Řazení znaků z latinského, řeckého, azbukového, gruzínského a arménského způsobu zápisu

ČSN
EN 13710
97 6031

European Ordering Rules – Ordering of characters from Latin, Greek, Cyrillic, Georgian and Armenian scripts

Regles de classement européen – Classement des caracteres latins, grecs, cyrilliques, géorgiens et arméniens

Europäische Sortierregeln – Sortierung von lateinischen, griechischen, kyrillischen, georgischen und armenischen Schriftzeichen

Tato norma je českou verzí evropské normy EN 13710:2011. 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 European Standard EN 13710:2011. It was translated by Czech Office for Standards, Metrology and Testing. It has the same status as the official version.

Národní předmluva

Informace o citovaných normativních dokumentech

ISO/IEC 10646:2003 nezavedena

ISO 12199:2000 dosud nezavedena

ISO/IEC 14651:2007 zavedena v ČSN ISO/IEC 14651:2011 (36 9169) Informační technologie – Mezinárodní řazení a porovnání řetězců – Metoda pro porovnávání znakových řetězců a popis obecné šablony pro přizpůsobení řazení

Související ČSN

ČSN ISO/IEC 15897:2002 (97 9707) Informační technologie – Procedury registrace kulturních prvků

ČSN 97 6030:1994 Abecední řazení

ČSN ISO/IEC 10367:1995 (36 9118) Informační technika – Normalizované kódované soubory grafických znaků pro použití v 8-bitových kódech

Vypracování normy

Zpracovatel: INFO 7, IČ 44266154, Ing. Jaroslav Ošlejšek

Technická normalizační komise: TNK 42 Výměna dat

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

EVROPSKÁ NORMA EN 13710
EUROPEAN STANDARD
NORME EUROPÉENNE
EUROPÄISCHE NORM Březen 2011

ICS 01.140.20; 35.040 Nahrazuje CR 14400:2001, ENV 13710:2000

Evropská pravidla řazení - Řazení znaků z latinského, řeckého, azbukového, gruzínského a arménského způsobu zápisu

European Ordering Rules - Ordering of characters from Latin, Greek, Cyrillic, Georgian and Armenian scripts

Regles de classement européen - Classement
des caracteres latins, grecs, cyrilliques, géorgiens
et arméniens

Europäische Sortierregeln - Sortierung
von lateinischen, griechischen, kyrillischen, georgischen und
armenischen Schriftzeichen

Tato evropská norma byla schválena CEN 2011-02-05.

Členové CEN jsou povinni splnit Vnitřní předpisy CEN/CENELEC, v nichž jsou stanoveny podmínky, za kterých se musí této evropské normě bez jakýchkoliv modifikací dát status národní normy. Aktualizované seznamy a bibliografické citace týkající se těchto národních norem lze obdržet na vyžádání v Řídicím centru nebo u kteréhokoliv člena CEN.

Tato evropská norma existuje ve třech oficiálních verzích (anglické, francouzské, německé). Verze v každém jiném jazyce přeložená členem CEN do jeho vlastního jazyka, za kterou zodpovídá a kterou notifikuje Řídicímu centru, má stejný status jako oficiální verze.

Členy CEN jsou národní normalizační orgány Belgie, Bulharska, České republiky, Dánska, Estonska, Finska, Francie, Chorvatska, Irska, Islandu, Itálie, Kypru, Litvy, Lotyšska, Lucemburska, Maďarska, Malty, Německa, Nizozemska, Norska, Polska, Portugalska, Rakouska, Rumunska, Řecka, Slovenska, Slovinska, Spojeného království, Španělska, Švédska a Švýcarska.

CEN

Evropský výbor pro normalizaci
European Committee for Standardization
Comité Européen de Normalisation
Europäisches Komitee für Normung

Řídicí centrum: Avenue Marnix 17, B-1000 Brusel

© 2011 CEN Veškerá práva pro využití v jakékoli formě a jakýmkoli prostředky Ref. č.
EN 13710:2011: E
jsou celosvětově vyhrazena národním členům CEN.

Obsah

1 Předmět normy 7

2 Citované normativní dokumenty 7

3 Definice 7

4 Shoda 8

5 Přizpůsobivost 8

6 Tabulka EOR odchylek 9

Příloha A (informativní) Principy evropských řadících pravidel 16

Příloha B (informativní) Řazení slova za slovem 32

Příloha C (informativní) Řazení podle pozice a stylu 33

Příloha D (informativní) Řazení smíšeného způsobu zápisu s jedním převládajícím způsobem zápisu 34

Příloha E (informativní) Definování národních odchylek založených na EOR 35

Příloha F (informativní) Moderní evropský způsob zápisu 39

Příloha G (informativní) EOR odchylka v syntaxi LDML 40

Bibliografie 47

Předmluva

Tento dokument (EN 13710:2011) byl připraven technickou komisí CEN/TC 304 „Informační a komunikační technologie – Požadavky na evropskou lokalizaci“ jejímž sekretariátem je DIN.

Této evropské normě je nutno nejpozději do září 2011 dát status národní normy, a to buď vydáním identického textu, nebo schválením k přímému používání, a národní normy, které jsou s ní v rozporu, je nutno zrušit nejpozději do září 2011.

Upozorňuje se na možnost, že některé prvky tohoto dokumentu mohou být předmětem patentových práv. CEN nelze činit odpovědným za identifikaci libovolného patentového práva nebo všech takových patentových práv.

Tento dokument nahrazuje ENV 13710:2000 a CR 14400:2001.

Ve vztahu k předchozímu vydání ENV 13710:2000 obsahuje toto vydání následující hlavní změny:

- a. ENV 13710:2000 a CR 14400:2001 byly sloučeny;
- b. dokument byl z části revidován a byl aktualizován.

Podle Vnitřních předpisů CEN/CENELEC jsou tuto evropskou normu povinny zavést národní normalizační organizace následujících zemí: Belgie, Bulharska, České republiky, Dánska, Estonska, Finska, Francie, Chorvatska, Irska, Islandu, Itálie, Kypru, Litvy, Lotyšska, Lucemburska, Maďarska, Malty, Německa, Nizozemska, Norska, Polska, Portugalska, Rakouska, Rumunska, Řecka, Slovenska, Slovinska, Spojeného království, Španělska, Švédsko a Švýcarska.

Úvod

Evropská norma poskytuje pravidla pro řazení vícejazyčných souborů dat do správně definované a intuitivní posloupnosti. Tato pravidla jsou určena pro data zapsaná v různých evropských jazycích, která musí být uvedena do předvídatelného pořadí, aby usnadnila uživateli z různých kulturních prostředí najít informaci. Ve stejném okamžiku je norma základem pro definici jazykově specifických profilů, které berou v úvahu pravidla dané jazykové komunity, stejně jako úplný pan-evropský soubor znaků v konzistentním pan-evropském způsobu.

Pravidla byla testována a široce přijata ve dvou předchozích specifikacích ENV 13710:2000 *Evropská pravidla řazení – Řazení znaků z latinského, řeckého a azbukového způsobu zápisu* a rozšířené specifikace CR 14400:2001 *Řazení znaků z latinského, řeckého, azbukového, gruzínského a arménského způsobu zápisu*. Tato evropská norma konsoliduje tyto normativní dokumenty do jedné kompatibilnější normy.

Tato evropská norma uspokojuje dvě rozdílné cílové skupiny implementátory softwaru na jedné straně a uživatele řadících aplikací na straně druhé.

Implementátoři softwaru potřebují mít jednoznačné strojově zpracovatelné pokyny, které mohou být snadno zavedeny do existujících a budoucích řadících aplikací. Tento cíl může být nejlépe dosažen definováním výchozí evropské tabulky v syntaxích dvou mezinárodně uznávaných specifikací:

- mezinárodní řadící normy ISO/IEC 14561:2007 Mezinárodní řazení řetězců a porovnání – Metoda pro porovnání znakových řetězců a popis obecné šablony pro přizpůsobení řazení, která se v této normě nazývá „delta“ (kapitola 6).
- „Locale Data Markup Language“ (LDML), který je použit na sběr (mezi jiným), řazených dat pro „Unicode Common Locale Data Repository (CLDR)“. CLDR staví na „Unicode Collation Algorithm (UCA)“, který je technicky profilem ISO/IEC 14651:2007 (informativní příloha G).

Uživatelé, kteří nemají konkrétní informační technologie, potřebují vysvětlit principy ve formě, která je ve větším souladu s národními normami pro řazení nebo s příslušnou praxí. Protože přizpůsobovací tabulky mohou být pro člověka obtížně čitelné, je vysvětlení principů přizpůsobovací tabulky uvedeno v informativních přílohách. Uživatelům, kteří nejsou obeznámeni s formální syntaxí přizpůsobovacích tabulek, se doporučuje seznámit se nejprve s těmito přílohami.

Hlavní normativní část této evropské normy specifikuje řazení znakových řetězců písmeno po písmenu. Informativní příloha A uvádí ekvivalentní informace pro člověka srozumitelnějším způsobem. Příloha B se zabývá řazením slova za slovem jako zvláštní formou řazení s vícenásobnými klíči. Informativní příloha C vysvětluje použití dalších řadících kritérií. Příloha D uvádí široce používané alternativy k hlavní části, zejména slučování několika způsobů zápisu do jednoho indexu přes implicitní transliteraci. Informativní příloha E uvádí návod na použití evropské normy jako základu pro vyjádření národních odchylek. Informativní příloha F uvádí základní repertoár znaků pro snadnější orientaci. Informativní příloha G vyjadřuje formální odchylku v syntaxi LDML.

V návaznosti na praxi ISO/IEC 14651, se na znaky odkazuje jako na UXXXX, kde X znamená jakoukoli hexadecimální číslici, které odkazuje na pozici kódu znaku v ISO/IEC 10646. Tato konvence je použita v celé této evropské normě.

1 Předmět normy

Tato evropská norma určuje řazení mezi dvěma řetězci znaků složených ze znaků *Moderních evropských způsobů zápisu (Modern European Skripts (MES))*, které jsou kolekcí ISO/IEC 10646:2003 nebo jejich podmnožin.

POZNÁMKA Kolekce 283 *Moderních evropských způsobů zápisu (MES) ISO/IEC 10646:2003* byla původně uvedena v CEN Workshop Agreement 13873:2000 *Multilingual European Subsets ISO/IEC 10646* jako *Multilingual European Subset Number 3* a byla následně začleněna jako kolekce v příloze A ISO/IEC 10646:2003 vedle sesterských kolekcí MES-1 a MES-2.

Pravidla řazení specifikovaná v této evropské normě jsou použitelná pouze pro soubory dat ve více než jednom evropském jazyce a v případě, že jsou určena pro multikulturní publikum. Doplní stávající národní normy či praxi této oblasti.

Konec náhledu - text dále pokračuje v placené verzi ČSN.