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

EESTI STANDARDIKESKUS EVS
ESTONIAN CENTRE FOR STANDARDISATION

EESTI STANDARDI EESSÕNA

NATIONAL FOREWORD

Käesolev Eesti standard EVS-EN 13710:2011 sisaldab Euroopa standardi EN 13710:2011 ingliskeelset teksti.	This Estonian standard EVS-EN 13710:2011 consists of the English text of the European standard EN 13710:2011.
Standard on kinnitatud Eesti Standardikeskuse 31.03.2011 käskkirjaga ja jõustub sellekohase teate avaldamisel EVS Teatajas.	This standard is ratified with the order of Estonian Centre for Standardisation dated 31.03.2011 and is endorsed with the notification published in the official bulletin of the Estonian national standardisation organisation.
Euroopa standardimisorganisatsioonide poolt rahvuslikele liikmetele Euroopa standardi teksti kättesaadavaks tegemise kuupäev on 16.03.2011.	Date of Availability of the European standard text 16.03.2011.
Standard on kättesaadav Eesti standardiorganisatsioonist.	The standard is available from Estonian standardisation organisation.

ICS 01.140.20, 35.040

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN 13710

March 2011

ICS 01.140.20; 35.040

Supersedes CR 14400:2001, ENV 13710:2000

English Version

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

Règles de classement européen - Classement des
caractères latins, grecs, cyrilliques, géorgiens et arméniens

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

This European Standard was approved by CEN on 5 February 2011.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Management Centre: Avenue Marnix 17, B-1000 Brussels

Ref. No. EN 13710:2011: E

Contents

Foreword

This document (EN 13710:2011) has been prepared by Technical Committee CEN/TC 304 "Information and communications technologies - European localization requirements", the secretariat of which is held by DIN.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by September 2011, and conflicting national standards shall be withdrawn at the latest by September 2011.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN [and/or CENELEC] shall not be held responsible for identifying any or all such patent rights.

This document supersedes ENV 13710:2000 and CR 14400:2001.

This edition incorporates the following main changes made with respect to the previous edition of ENV 13710:2000 mainly as follows:

a)   ENV 13170:2000 and CR 14400:2001 have been consolidated;

b)   the document has been partly revised and has been brought up to date.

According to the CEN/CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and United Kingdom.

Introduction

This European Standard provides rules for ordering multilingual lists into a well-defined and intuitive sequence. These rules are intended for data from different European languages that must be brought into a predictable order that makes it easy for users from multiple cultural backgrounds to find information. At the same time the standard is a basis for the definition of language-specific profiles taking the rules of a given language community into account at the same time as the total pan-European character set in a consistent, pan-European manner.

The rules have been tested and widely adopted in two predecessor specifications, ENV 13710:2000 *European Ordering Rules — Ordering of characters from the Latin, Greek and Cyrillic scripts* and its companion and extension specification CR 14400:2001 *European Ordering Rules - Ordering for Latin, Greek, Cyrillic, Georgian and Armenian scripts*. This European Standard consolidates these two documents into one technically largely upwards compatible standard.

This European Standard caters for two different target groups, software implementers on the one hand and users of ordering applications on the other.

Software implementers need unambiguous, machine-processable guidelines, which can readily be loaded into existing and future ordering applications. This goal can best be achieved by defining a european default ordering table in the syntaxes of two internationally relevant specifications in the field:

— the international ordering standard ISO/IEC 14651:2007 International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering, of which the present standard is a "delta" (Clause 6);

— the Locale Data Markup Language (LDML) which is used to capture (amongst others) the collation data for the Unicode Common Locale Data Repository (CLDR). CLDR builds on the Unicode Collation Algorithm (UCA) which is technically a profile of ISO/IEC 14651:2007 (informative Annex G).

Users with no specific ICT background, however, need an explanation of the principles in a form more in line with existing national ordering standards or relevant practice. Tailoring tables can be difficult to read for human readers, so an explanation of the principles behind that table is given in the informative annexes. Users not familiar with the formal syntax of the tailoring table are advised to consult those annexes first.

The normative main part of this European Standard specifies letter-by-letter ordering of character strings. Informative Annex A presents equivalent information in a more human-oriented way. Informative Annex B deals with word-by-word ordering as a special form of ordering with multiple keys. Informative Annex C explains the use of further ordering criteria. Informative Annex D presents a widely used alternative to the main part, namely the amalgamation of several scripts in one index via implicit transliteration. Informative Annex E gives guidance on the use of this European Standard as the basis for expressing national deltas. Informative Annex F lists the underlying character repertoire for ease of reference. Informative Annex G expresses the formal delta in the LDML syntax.

Following the practice of ISO/IEC 14651 characters are referenced as UXXXX where *X* stands for any hexadecimal digit and refers to the code position of that character in ISO/IEC 10646. This convention is used throughout this European Standard.

1 Scope

This European Standard specifies the order between two character strings composed of characters from the *Modern European Scripts* (MES) collection of ISO/IEC 10646:2003 or subsets of it.

NOTE Collection 283 *Modern European Scripts* (MES) of ISO/IEC 10646:2003 was originally specified in CEN Workshop Agreement 13873:2000 *Multilingual European Subsets of ISO/IEC 10646* as *Multilingual European Subset Number 3* and was subsequently incorporated as a collection in Annex A of ISO/IEC 10646:2003 alongside its sister collections MES-1 and MES-2.

The ordering rules specified in this European Standard are only applicable for lists of data in more than one European language and when this data is intended for a multicultural audience. They complement existing national standards or practices in the field.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10646:2003, *Information technology — Universal Multi-Octet Coded Character Set (UCS)*

ISO 12199:2000, *Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet*

ISO/IEC 14651:2007, *Information technology –– International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering*

3 Terms and definitions

For the purposes of this document, the following terms and definitions of ISO/IEC 10646:2003 and of ISO/IEC 14651:2007 apply.

3.1
character
member of a set of elements used for the organisation, control, or representation of data
[ISO/IEC 10646:2003]

3.2
character string
sequence of characters considered as a single object
[ISO/IEC 14651:2007]

3.3
collating symbol
symbol used to specify weights assigned to a collating element
[ISO/IEC 14651:2007]

3.4
collating element
sequence of one or more characters that are considered a single entity for ordering
[ISO/IEC 14651:2007]