

CEN

CWA 16108

WORKSHOP

March 2010

AGREEMENT

ICS 35.180

English version

Functional Multilingual Extensions to European Keyboard Layouts

This CEN Workshop Agreement has been drafted and approved by a Workshop of representatives of interested parties, the constitution of which is indicated in the foreword of this Workshop Agreement.

The formal process followed by the Workshop in the development of this Workshop Agreement has been endorsed by the National Members of CEN but neither the National Members of CEN nor the CEN Management Centre can be held accountable for the technical content of this CEN Workshop Agreement or possible conflicts with standards or legislation.

This CEN Workshop Agreement can in no way be held as being an official standard developed by CEN and its Members.

This CEN Workshop Agreement is publicly available as a reference document from the CEN Members National Standard Bodies.

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

Contents

Foreword	3
Introduction	4
1 Scope	5
2 Informative references	5
2.1 Overview	5
2.2 Formal reasons for expanding the character repertoire	5
2.2.1 General	5
2.2.2 The Lisbon treaty	5
2.2.3 UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions	6
2.2.4 European Charter for Regional or Minority Languages	6
2.2.5 On the recording of surnames and forenames in civil status registers, Convention relative à l'indication des noms et prénoms dans les registres de l'État Civil, CIEC 14	6
2.2.6 German-Polish Treaty on good neighbourly relations and friendly cooperation	7
3 Background	8
4 Usage scenarios	8
5 Proposed solutions	9
5.1 A pragmatic approach to extend current keyboards	9
5.2 A global input method	10
5.3 Ongoing work on international standardization	10
6 Summary	11
7 Assessment	12
8 Further considerations	13
Annex 1: A quick guide to the basic setting of the keyboard layout standard SFS 5966	14
Annex 2: Description of a proposal for a Global Input Method	18
Annex 3: A tentative "Five Scripts Keyboard" layout design	22

Foreword

The production of this document which addresses the considerations and guidance for functional multilingual extensions to European keyboard layouts, primarily for users of the Latin script, was agreed by the CEN Workshop Functional Multilingual Extensions to European Keyboard Layouts (WS/MEEK) in the Workshop's Kick-Off meeting on 2008-01-25.

This document has been developed through the collaboration of a number of contributing partners in WS/MEEK. The WS/MEEK representation gathers a wide mix of interests, coming from academia, public administrations, IT suppliers, and other interested experts. The present CWA (CEN Workshop Agreement) has received the support of representatives from each of these sectors. There have been several WS/MEEK internal review/endorsement rounds of this CWA. In addition, parallel to the last internal round, the CWA has been the subject of a public comment period from 2009-10-13 to 2009-11-14 (following an earlier period from 2009-06-08 to 2009-08-28). The final physical meeting of the Workshop took place on 2009-09-03 and the final on-line meeting on 2009-11-17. The final text of this CWA was submitted to CEN for publication on 2009-12-30.

This document has been developed in close liaison with ISO/IEC JTC 1/SC 35, "User Interfaces", and the Unicode Consortium. SC 35 is responsible within ISO for keyboard layouts; the participation in ISO is through the National Standards Organizations. Members of the Unicode Consortium include major computer corporations, software producers, database vendors, government ministries, research institutions, international agencies, various user groups, and interested individuals.

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 CEN Workshop Agreement has only been made in English.

Experts from the following organizations have approved this document:

- Nokia
- Lionbridge
- Red Hat
- University of Helsinki, Finland
- University of Applied Sciences Worms, Germany
- New Bulgarian University, Bulgaria
- Finnish Standards Association SFS, Finland
- Standard Norge, Norway
- Bulgarian Institute for Standardisation, Bulgaria
- ISO/IEC JTC 1/SC 35
- The Unicode Consortium
- Ministry for Education, Finland
- Ministry for Education, France
- Ministry of State Administration and Administrative Reform, Bulgaria
- European Commission
- CSC - IT Center for Science, Finland

This CEN Workshop Agreement is publicly available as a reference document from the National Members of CEN: AENOR, AFNOR, BSI, CSNI, CYS, DIN, DS, ELOT, EVS, IBN, IPQ, IST, HZN, LVS, LST, MSA, MSZT, NEN, NSAI, ON, PKN, SEE, SIS, SIST, SFS, SN, SNV, SUTN and UNI.

Comments or suggestions from the users of the CEN Workshop Agreement are welcome and should be addressed to the CEN-CENELEC Management Centre.

Introduction

Year 2008 was proclaimed the International Year of Languages by the United Nations General Assembly. It was also declared the European Year of Intercultural Dialogue by the European Parliament and the Council of the European Union. The MEEK Workshop can be seen as a small step in support of these themes.

The current CEN membership consists of 31 countries: 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.

These countries have several official languages and even more regional and minority languages. Some of the regional and minority languages have certain official status, either by national law or by the ratification of the applicable Council of Europe Treaty. Of the official languages, Bulgarian is written with Cyrillic letters and Greek with Greek letters, whereas all the others are written with Latin letters. The Latin repertoire includes several letters beyond the “basic a-to-z” and a very large number of letters with diacritics, some of which are perceived as basic letters in one or more languages, and also collated separately (e.g., å, ä and ö after z in Swedish and Finnish).

In addition, many more native and immigrant languages written in a number of different scripts are spoken in Europe.

1 Scope

This CWA is aimed to assist in the preparation of functional multilingual extensions to European keyboard layouts. They are aimed to allow “ordinary users” and Public Authorities to input primarily Latin-script characters, especially in the light of current and potential future legal and common educational requirements. The CWA builds upon existing implementations of common official and de facto standards for national and regional computer keyboard layouts and related input methods. Based on this, practical guidance is given on handling multilingual data entry requirements, taking account of existing international standards in the field. Outside the scope of this CWA are the needs of specialist usage, such as the need to regularly and rapidly generate large volumes of text in a wide range of languages.

Since the European single market allows for free movement of people and goods, one should be able to correctly enter the names of people, places, products, and companies and other legal entities in a consistent, easily comprehensible manner. This implies that the capability should exist in all kinds of applications, including those based on open source software that are traditionally based on freely available information.

The CWA does not define any specific, let alone Pan-European, keyboard layout. Liaison has been established with ISO/IEC JTC1/SC35 to minimize the risk of further divergence between actual implementations and formal standards (ISO/IEC 9995 series and other relevant standards). Liaison has also been established with the Unicode Consortium.

2 Informative references

2.1 Overview

In the past, the number of simultaneously representable characters in a computing environment was severely limited because of the 8 bit character encoding schemes in use. These encodings only allowed 256 unique values. In fact, even fewer characters could be represented, as some of the values were reserved for control characters. As a consequence, each country has historically concentrated on having its own official languages properly supported in ICT, more recently also its own regional and minority languages.

The expanding European single market, however, with free movement of goods and people, is drastically changing the support requirements. For both accuracy and politeness, all kinds of proper names should be communicated in their correct form. This change is driven not only by increasing business and legal reasons but also as a personal choice by many individuals in appreciation of cultural diversity.

2.2 Formal reasons for expanding the character repertoire

2.2.1 General

As witnessed by key legal documents cited here, respect for cultural diversity is recognized as a human right in International and European law.

2.2.2 The Lisbon treaty

“The Union shall respect cultural, religious and linguistic diversity.” (Charter of Fundamental Rights of the European Union, Article 22)