

Information technology - Alphanumeric glyph image set for optical character recognition OCR-B - Shapes and dimensions of the printed image

Information technology - Alphanumeric glyph image set for optical character recognition OCR-B - Shapes and dimensions of the printed image

EESTI STANDARDI EESSÕNA

NATIONAL FOREWORD

<p>Käesolev Eesti standard EVS-EN 14603:2005 sisaldab Euroopa standardi EN 14603:2004 ingliskeelset teksti.</p> <p>Käesolev dokument on jõustatud 22.02.2005 ja selle kohta on avaldatud teade Eesti standardiorganisatsiooni ametlikus väljaandes.</p> <p>Standard on kättesaadav Eesti standardiorganisatsioonist.</p>	<p>This Estonian standard EVS-EN 14603:2005 consists of the English text of the European standard EN 14603:2004.</p> <p>This document is endorsed on 22.02.2005 with the notification being published in the official publication of the Estonian national standardisation organisation.</p> <p>The standard is available from Estonian standardisation organisation.</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p>Käsitlusala: This European Standard defines a set of glyph images designated OCR-B, intended primarily for use in Optical Character Recognition (OCR) applications, but suitable also for visual, i.e. human, reading. It does not relate any coding scheme with these images (see clause 5)</p>	<p>Scope: This European Standard defines a set of glyph images designated OCR-B, intended primarily for use in Optical Character Recognition (OCR) applications, but suitable also for visual, i.e. human, reading. It does not relate any coding scheme with these images (see clause 5)</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

ICS 35.020

Võtmesõnad:

ICS

English version

**Information technology - Alphanumeric glyph image set for
optical character recognition OCR-B - Shapes and dimensions
of the printed image**

Technologies de l'information - Jeu d'images de glyphe
alphanumérique pour la reconnaissance optique de
caractères OCR-B - Formes et dimensions de l'image
imprimée

This European Standard was approved by CEN on 17 June 2004.

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 Central Secretariat 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 Central Secretariat has the same status as the official versions.

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



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

Management Centre: rue de Stassart, 36 B-1050 Brussels

Foreword

This document (EN 14603:2004) has been prepared by Technical Committee CEN/TC 304, "Information and communication technologies – European localization requirements", the secretariat of which is held by SIS.

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 June 2005, and conflicting national standards shall be withdrawn at the latest by June 2005.

The document is based on the International Standard ISO 1073/II, "*Alphanumeric character set for optical recognition – Part II: Character set OCR-B – Shapes and dimensions of the printed image*".

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

Contents

1	Scope	5
2	Conformance.....	5
3	Normative references	5
4	Terms and definitions	5
5	Coding in OCR applications	6
6	OCR-B styles.....	6
7	OCR-B sizes	6
8	Typical dimensions of the nominal printed image.....	7
9	OCR-B glyph image set.....	8
9.1	Subset 1: Minimal alphanumeric subset.....	8
9.2	Subset 2: Basic alphanumeric subset.....	8
9.3	Subset 3: Extended alphanumeric subset	9
9.4	Subset 4: Options subset	9
10	Index table	10
10.1	Availability of glyph images	10
10.2	Identification of drawings	10
10.3	Application considerations	10
11	Use of diacritical marks	18
11.1	Diacritical mark repertoire	18
11.2	Composite glyph images	18
11.3	Rules for glyph image combinations	18
12	Use of the LOW LINE glyph	19
13	SPACE	19
14	Glyph image shape definitions	19
14.1	Reference drawings	19
14.2	Availability of duplicates	19
14.3	Type dimensions	19
14.4	Constant-strokewidth font, size I.....	19
14.5	Constant-strokewidth font, size III.....	20
14.6	Constant-strokewidth font, size IV.....	20
14.7	Letterpress font, size I	20
15	Printing the letterpress and constant-strokewidth fonts	20
16	Illustration of OCR-B	20
Annex A (normative) Definition of Euro sign glyph image (ISO/IEC 9541-3 syntax)		22
Annex B (informative) Main differences between ISO 1073/II-1976 and this European Standard		23
Annex C (informative) Notes on the implementation of OCR-B		24
Annex D (informative) Glyph-repertoire extension needs identified in JTC 1/SC 2 revision process		25
Annex E (informative) Illustrations of reference drawings		29
Annex F (informative) Availability of reference drawings.....		32
Bibliography		33

Introduction

Optical Character Recognition technology, OCR, came into use in the 1960s, and some specialized OCR fonts were designed at the time. In 1976 two such fonts were formally standardized by ISO, designated OCR-A and OCR-B, in the standard ISO 1073 parts I and II, respectively.

ISO 1073 was developed by the ISO Technical Committee ISO/TC97, *Computers and information processing*. At the creation of ISO/IEC JTC 1, responsibility for ISO 1073 was transferred to JTC 1/SC 2, *Coded character sets*.

In order to enlarge the set of characters covered by the standard, especially with special letters used in European-origin languages, a revision of the standard was initiated in 1994 by JTC 1/SC 2, and progressed through three consecutive Committee Drafts. Since however testing of the proposed character set extensions could not be accomplished, the JTC 1/SC 2 revision was discontinued in 1999.

With the introduction of the Euro sign a need – primarily European – to add that character to the OCR-B set was recognized. CEN/TC304 therefore decided to develop an OCR-B glyph image shape for the character, verify its recognition properties, and include it in a European version of the OCR-B standard; see CEN/TC304 reports referenced in the Bibliography. The decided-on glyph image shape is specified in Annex A.

For reasons of continuity, and also to facilitate possible future CEN – ISO/IEC cooperation on OCR-B, it was decided to use the current ISO text with only the necessary minimum of changes as a basis for the CEN standard, even though the ISO text was developed in an OCR-technology situation rather different from the one existing when this CEN standard is published. In particular, the ISO standard text's division into clauses was kept as far as possible, although some restructuring might have been desirable.

A description of the main differences between this European Standard and ISO 1073/II is given in Annex B. General information on the implementation of the OCR-B shapes, taken from ISO 1073/II, has been included in Annex C.

In connection with the verification of the recognition properties of the Euro sign, some limited verification was also done on special letters identified during the JTC 1/SC 2 revision work as needed in OCR-B. The extent of this verification is not sufficient for the inclusion of the letters in the OCR-B repertoire at present, but the issue is described in Annex D, as a basis for possible future inclusion work.

1 Scope

This European Standard defines a set of glyph images designated OCR-B, intended primarily for use in Optical Character Recognition (OCR) applications, but suitable also for visual, i.e. human, reading. It does not relate any coding scheme with these images (see clause 5).

This European Standard is based on the ISO standard 1073 part II. It differs from that standard in extending normatively the set of glyph images with the Euro currency sign; but also in deleting some glyphs not relevant in present-day OCR processing. It further adds information on a number of glyph images corresponding to characters specific to some European-origin languages.

NOTE In ISO 1073 Part II the term "character" is used not only in its strict sense, but also to mean the printed images used for their visual, i.e. printed, representations. In this European Standard the term "glyph image" is used in the latter sense.

This European Standard contains information on nominal dimensions for the glyph images. Tolerances, printing quality and other characteristics of the formats needed to satisfy interchange requirements are covered in other standards (see clause 3).

The glyph image set contains 117 glyph images comprising digits, capital and small letters, diacritical marks, and symbols. It also contains a definition for SPACE.

The diacritical marks are designed for combination with small letters to produce composite glyph images complementing the basic image repertoire.

2 Conformance

A printing or OCR reading device is in conformance with this standard if it can generate/recognize, for either or both of the defined styles (see clause 6) and in one or more of the specified sizes (see clause 7), all or part of the specified glyph image subsets (see clause 9).

A claim of conformance shall specify all the images in (each of) the style(s) and size(s) generated/recognized. Such a specification shall take the form of a reference to one of the subsets, a list of the images generated/recognized, or a combination of those.

Additionally, a printing or OCR reading device must claim conformance to International Standard ISO 1831 (see clause 3).

Printed images produced by an OCR-B printing device are in conformance with this standard if their nominal shapes and dimensions are in accordance with their respective reference drawing(s) and, in the case of the Euro sign glyph image, with Annex A (see clause 14); with the claimed conformance to tolerances and printing quality factors specified in standard ISO 1831 considered.

3 Normative references

This European Standard incorporates by dated or undated reference, provisions from other publications. These normative references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this European Standard only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

ISO 1831-1980, *Printing specifications for optical character recognition*.

ISO/IEC 9541-3:1994, *Information technology — Font information interchange — Part 3: Glyph shape representation*

OCR-B character reference drawings and glyph definition (see clause 14).

4 Terms and definitions

For the purposes of this European Standard, the following terms and definitions apply:

4.1 character

a member of a set of elements used for the organisation, control or representation of data.

4.2 coded character set

a set of characters, defined by unambiguous rules that establish the character set and the relationship between the characters of the set and their coded representations.

4.3 composite glyph image

An image printed on paper or any other medium intended for OCR applications, obtained by superimposing two or more glyph images on the same area.

4.4 glyph

A recognizable abstract graphic symbol which is independent of any specific design.