SLOVENSKI STANDARD
SIST EN 14603:2005

01-april-2005

±bzcfa UW]^g_UˈhY\ bc`c[ ]^UˈËˈ5ˈZUbi a Yf] b]ˈbUWcfˈg`]_cjb]\ ˈnbU_cjˈnUˈcdh] bc fUhdcnbUjcˈnbU_cj ˈC7F!6ˈËˈCVˈ]_Yˈ]bˈa YfYˈhjg_UbYˈg`]_Y

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

Informationstechnik - Alphanumerischer Bildzeichensatz für optische Zeichenerkennung
OCR-B - Formen und Abmessungen des gedruckten Bildes

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

Ta slovenski standard je istoveten z: EN 14603:2004

ICS:

35.040	Nabori znakov in kodiranje informacij	Character sets and information coding

SIST EN 14603:2005   **en**

iTeh STANDARD PREVIEW
(standards.iteh.ai)

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN 14603

December 2004

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

Ref. No. EN 14603:2004: E

EN 14603:2004 (E)

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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

2

Contents

EN 14603:2004 (E)

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.

4

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.

4.5
glyph image

An image of a glyph, as obtained from a glyph representation printed on paper or any other medium intended for OCR applications.

NOTE   The definition above of "coded character set" differs slightly from definitions in ISO/IEC standards, and the definition of "glyph image" is more limited. The definition of "composite glyph image" is specific to this standard (at the time of its publication).

5   Coding in OCR applications

This standard defines a set of glyph images, but does not specify corresponding characters, and relates no coding with the images. The images have been named as far as possible in the same way as the characters with corresponding glyphs in the ISO/IEC standard 10646-1 (see Bibliography), but this does not imply any normative association between the OCR-B glyph images according to this European Standard and the characters of either ISO/IEC 10646-1 or any other standard for coded character sets.

Printing and/or OCR applications based on this European Standard must therefore define, through reference to other standards or otherwise, the set of glyph images which is available for printing and/or shall be recognized, and for each image the corresponding character and its coding.

6   OCR-B styles

The OCR-B glyph images are defined by this standard in two different styles.

The "constant-strokewidth" style is intended primarily for printer equipment in which the width of the strokes of the images is less controllable. This is for instance the case for some types of mechanical printers.

The "letterpress" style is intended for printing equipment which can reproduce fine details with high accuracy. For aesthetic reasons, the strokewidths of the letterpress images are varied deliberately, and the stroke endings are specially designed.

The shapes of the glyph images for the two styles are specified (with the exception of the Euro sign glyph) by reference drawings. The constant-strokewidth style drawings show the centrelines of the strokes and the contours of the ends of the strokes; the letterpress style drawings show the complete outlines of the images (see clause 14). The general dimensions and the centrelines of the images are the same for both styles.

7   OCR-B sizes

Three sizes are specified for OCR-B glyph images in order to provide for use with a wide range of printing equipment possessing differing print quality characteristics.

NOTE   The metric and inch dimensions in this European Standard are rounded and therefore consistent but not exactly equal. Either system may be used but the two should not be intermixed.

The letterpress font is specified in size I (the smallest) only. It provides the option of a variable pitch in printing as is usual with letterpress.

The constant-strokewidth font is specified in three sizes: I, III and IV. Mechanisms using the constant-strokewidth font will usually maintain a fixed pitch.

NOTE   Size II which was defined in the original draft version of ISO 1073/II was not included in the final version.

The centrelines for the three sizes are simply related by appropriate horizontal and vertical scale factors. The factors for size III and size IV referred to size I are:

for size III: Vertical 1,333; horizontal 1,086
for size IV: Vertical 1,500; horizontal 1,500

This scale relationship does not apply to the outline shapes since nominal strokewidth is not strictly proportional to centreline dimensions. The strokewidths for each size are shown in the reference drawings.

The glyph image with the greatest height above the base line ("A" in figure 1) in each size is DIGIT EIGHT. The image with the greatest total height is SMALL LETTER J, because of its descender.

The centreline heights of the DIGIT EIGHT are:

    for size I: 2,40 mm (0,094 in)
    for size III: 3,20 mm (0,126 in)
    for size IV: 3,60 mm (0,142 in)

The widest glyph image in each size (except for the Alternative SMALL LETTER M) is DIGIT ZERO. Its centreline widths are:

    for size I: 1,40 mm (0,055 in)
    for size III: 1,52 mm (0,060 in)
    for size IV: 2,10 mm (0,083 in)

In fixed-pitch printing for OCR applications, the following minimum nominal pitches are recommended:

    size I: 2,54 mm (0,100 in)
    size III: 2,54 mm (0,100 in)
    size IV: 3,63 mm (0,143 in)

8 Typical dimensions of the nominal printed image

Typical dimensions for the nominal printed image of the letterpress font in size I are given in table 1. These dimensions are the heights above and below the horizontal base line of digits, capital and small letters, ascenders and descenders (see figure 1).

The shapes and dimensions of the constant-strokewidth glyph images are similar except that the stroke ends are rounded.

NOTE   It is recognized that some type-making and printing processes will not be able to produce sharp corners. Corners not given a specific radius should be as sharp as practicable. However, it is not necessary for OCR purposes that the radii of the corners of the nominal printed image be less than 0,08 mm (0,0035 in).

The dimensions in table 1 are for general information only. The values for individual glyph images are obtainable from the reference drawings.
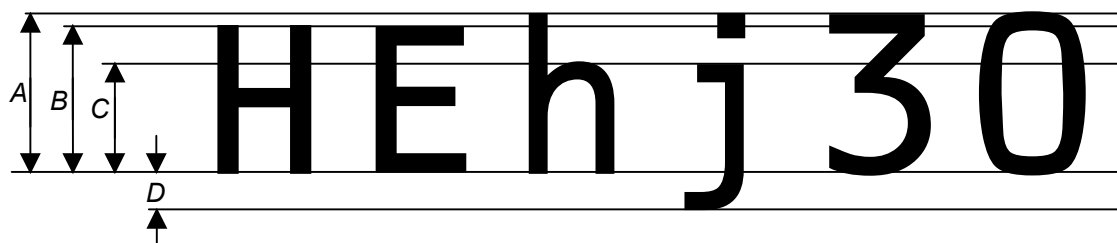

Figure 1 — Heights above and below base line

iTeh STANDARD PREVIEW

(standards.iteh.ai)

Table 1 — Typical dimensions A, B, C and D

Size	millimeters				inches			
	A	B	C	D	A	B	C	D
I	2,66	2,46	1,83	0,60	0,105	0,097	0,072	0,024

EN 14603:2004 (E)

9   OCR-B glyph image set

The full set contains 117 glyph images and a definition for SPACE (see clause 13). Four subsets are defined in 9.1 – 9.4 below.

9.1   Subset 1: Minimal alphanumeric subset

This subset applies to sizes I, III and IV in constant-strokewidth font and to size I in letterpress font. It contains 21 glyph images and SPACE:

0 1 2 3 4 5 6 7 8 9

C E N S T X Z

< + > | (SPACE)

9.2   Subset 2: Basic alphanumeric subset

This subset applies to sizes I and IV in constant-strokewidth font and to size I in letterpress font. It contains 25 glyph images in addition to subset 1, i.e. a total of 46 glyph images and SPACE:

0 1 2 3 4 5 6 7 8 9

A B C D E F G H I J K L M

N O P Q R S T U V W X Y Z

< + > * – = / . , | (SPACE)

8

9.3 Subset 3: Extended alphanumeric subset

This subset applies to sizes I and IV in constant-strokewidth font and to size I in letterpress font.

It contains 51 glyph images in addition to subset 2, i.e. 97 glyph images in all (and SPACE); in particular the images corresponding to those characters listed in ISO/IEC 646 as unique, alternative, and International Reference Version:

! " # € $ £ ¤ % & ' ( ) * + , - . /

0 1 2 3 4 5 6 7 8 9 : ; < = > ?

@ A B C D E F G H I J K L M N O

P Q R S T U V W X Y Z [\ ] ^ _

' a b c d e f g h i j k l m n o

p q r s t u v w x y z { | } ~   (SPACE)

9.4 Subset 4: Options subset

This subset applies to sizes I and IV in constant-strokewidth font and to size I in letterpress font. It contains 8 capital and 5 small national letters, 4 diacritical marks and 3 further glyph images.

Images from this subset shall be used only in conjunction with subset 3. A printing or OCR reading device may generate/recognize any of the images of this subset. The images generated/recognized by the device shall be listed in the claim of conformance (see clause 2).

Ä Å Æ IJ Ñ Ö Ø Ü   å æ ij ø ß

´ ¨ ^

,

§ ¥ m

EN 14603:2004 (E)

10 Index table

10.1 Availability of glyph images

All glyph images except the Euro sign are available in size I as constant-strokewidth font. All glyph images including the Euro sign are available in size I as letterpress font.

Only the images of the minimal alphanumeric subset (subset 1) are available in size III as constant-strokewidth font.

All images are available in size IV as constant-strokewidth font, with the exception of the Euro sign and VERTICAL LINE.

10.2 Identification of drawings

In the following table each image is given with the number of its reference drawing(s) and the subset(s) in which it is comprised.

The drawings are identified as follows :

    L: for letterpress font, size I
    C: for the constant-strokewidth font, size I
    III: for the constant-strokewidth font, size III

As stated in 14.6, the shapes for size IV are derived from those of size I for the constant-strokewidth font (designated by C).

10.3 Application considerations

Application advice is given in the column "Remarks".

It is recommended that prospective users of this European Standard consult manufacturers before deciding on a particular set of glyph images.

The ordering of the glyph images in the table does not imply any kind of coding scheme.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Table 2 — OCR-B glyph image set

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
1	1	1 L, C, III	DIGIT ONE	1 2 3	
2	2	2 L, C, III	DIGIT TWO	1 2 3	
3	3	3 L, C, III	DIGIT THREE	1 2 3	
4	4	4 L, C, III	DIGIT FOUR	1 2 3	
5	5	5 L, C, III	DIGIT FIVE	1 2 3	
6	6	6 L, C, III	DIGIT SIX	1 2 3	
7	7	7 L, C, III	DIGIT SEVEN	1 2 3	
8	8	8 L, C, III	DIGIT EIGHT	1 2 3	
9	9	9 L, C, III	DIGIT NINE	1 2 3	