TECHNICAL SPECIFICATION

ISO/TS 12029

First edition
2002-03-15

Electronic imaging — Forms design optimization for electronic image management

Imagerie électronique — Optimisation de conception de formulaires pour la gestion d'images électronique

Reference number
ISO/TS 12029:2002(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years with a view to deciding whether it should be confirmed for a further three years, revised to become an International Standard, or withdrawn. In the case of a confirmed ISO/PAS or ISO/TS, it is reviewed again after six years at which time it has to be either transposed into an International Standard or withdrawn.

Attention is drawn to the possibility that some of the elements of this Technical Specification may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 12029 was prepared by Technical Committee ISO/TC 171, *Document imaging applications*, Subcommittee SC 2, *Application issues*.

Introduction

This Technical Specification discusses issues and provides guidance for the design of forms that are used for electronic capture of information. Features include colour dropouts, type fonts, printing screen tints, line width, data storage and other interrelated issues. It is necessary to balance conflicting requirements of user-friendliness and electronic capture. Making a form appealing by use of colour or graphics could assist users when they complete the form but could also decrease the form's scannability or other automated related functions. This conflict might require compromise in the design of a form.

Electronic imaging — Forms design optimization for electronic image management

1 Scope

This Technical Specification provides guidelines for the design of forms to be completed by users and scanned for processing by electronic image management (EIM) systems. These guidelines are limited to forms using roman characters.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Specification. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this Technical Specification are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 1073-1:1976, *Alphanumeric character sets for optical recognition — Part 1: Character set OCR-A — Shapes and dimensions of the printed image*

ISO 1073-2:1976, *Alphanumeric character sets for optical recognition — Part 2: Character set OCR-B — Shapes and dimensions of the printed image*

ISO 2471:1998, *Paper and board — Determination of opacity (paper backing) — Diffuse reflectance method*

ISO 11108, *Information and documentation — Archival paper — Requirements for permanence and durability*

ISO 12651:1999, *Electronic imaging — Vocabulary*

ISO 12653-1, *Electronic imaging — Test target for the black-and-white scanning of office documents — Part 1: Characteristics*

ISO 12653-2:2000, *Electronic imaging — Test target for the black-and-white scanning of office documents — Part 2: Method of use*

3 Terms and definitions

For the purposes of this Technical Specification, the terms and definitions given in ISO 12651 and the following apply.

3.1
alphanumeric
pertaining to a character set that contains letters, numbers and other characters, such as punctuation marks and symbols

3.2
dropout ink
ink of a colour that cannot be detected by a scanner

3.3
font
complete set of characters of a given size, weight and style of type, including capitals, small capitals and lower-case characters, together with figures, punctuation marks, ligatures, etc.

3.4
magnetic ink character recognition
MICR
machine recognition of digits printed with magnetizable ink

3.5
character pitch
number of characters per unit length of a line of print

3.6
recognition zone
area around a recognition data field that is free of other data

3.7
optical mark recognition
OMR
machine recognition of a mark such as a tick, cross or spot based on minimum area rather than shape of the mark

4 Layout and design

4.1 General

The design of a form which is the easiest for a person to complete can be in conflict with the most machine-readable form. For example, in a user-friendly layout, the following items, all interspersed with printed instructions next to specific areas, might be desirable:

— large print;

— colour-coded areas; and

— areas to be completed with both alphabetic and numeric information.

However, in a form designed for EIM, strict segregation of spaces for numeric and alphabetic information and instructional text within dropout colour areas can all be essential features. Optimum design can require a compromise between ideal user and scanner requirements. All logically connected information should be placed on the same page.

4.2 Data storage requirement

The designer should be aware of the impact on data storage requirements of line borders, screened tints and logos or other design elements with large areas of reversed print. Large reverses will make heavy demands on data storage. For all designs, particularly those having large areas of reversed print, the amount of data storage required should be determined and compared to the amount of data storage available within the system.

4.3 Page format

On each page of a form, the margin all around should not be less than 10 mm. If the document is bound, padded or has punched holes or die-cuts, the margin at that edge should be not less than 25 mm, and holes and die-cuts should be restricted to that margin.

Text and entry fields should not be within 6,5 mm of any crease or perforation.

4.4 Type

4.4.1 Typeface

There are two styles of typeface commonly used on forms, serif and sanserif, as illustrated in Figure 1.

Figure 1 — Comparison of serif and sanserif typefaces

Serif style is designed for ease of legibility, has variable line width within a character and a cross-line finishing a stroke of a letter. Sanserif has uniform line width within a character and no cross-lines. It is the style used in International Standards.

Serif type will inherently take more data storage capacity in a compressed image than sanserif type because more information has to be recorded for each character. With the most commonly used compression techniques, approximately 10 % more storage is required for a page printed in serif, as compared with one printed in sanserif type.

Sanserif generally requires less horizontal line space and more vertical height than the same point size serif type. Because of its uniform line width, it is preferred for photocopying, microfilming and scanning.

For forms which might be used in optical character recognition (OCR) applications, sanserif typefaces should be used.

For information on a form that is not required to be captured by scanning, the style of typeface used is not important.

4.4.2 Symbols

An OCR program can use a particular symbol to prompt an action. The forms designer should be aware of any such symbols and avoid their use other than as a prompt.

4.4.3 Spacing

In typesetting, character spacing can be either fixed or proportional. In fixed typesetting, each character takes up equal horizontal space. Proportional typesetting allows for characters of different width, such as that of “i” compared with “w” and automatically adjusts space between the individual characters to give a more natural appearance.

There should be a clear gap between characters. The recommended minimum gap is not less than the width of the vertical stroke of characters of the font.

The designer should also be concerned with the vertical spacing requirements of an OCR system. Although 4,2 mm vertical spacing is usually sufficient for typewritten entries, at least twice that amount of space is necessary for hand-printed entries and for separating entries for OCR.

4.4.4 Character pitch

Form design should allow no more than 0,4 characters per millimetre for character pitch.

4.4.5 Character size

In the printing industry, type size is usually specified in millimetres. In a computer or typewriter, type size is usually indicated in points. Fortunately, the printing industry is familiar with both systems and can easily translate requirements. The point (0,35 mm) is a unit derived from the height of metal slugs, once commonly but now rarely, used to set type. The size of character is not directly related to point size. For a given point size, the actual heights of the same upper-case character can be different for various typefaces. There is also variation in the ratio of heights of lower-case “e” to upper-case characters. This means that, for a given point size, even if upper-case characters of two different typefaces have the same height, there is a possibility that it is not so for lower-case characters. Because it is the size of the lower-case characters that will limit scannability, minimum acceptable point size should be determined by the height of the lower-case “e”. The recommended minimum height of the lower-case “e” is 1,4 mm.

If the EIM system is used as a transfer medium as part of overall processing of the information extracted from a form, the minimum type size used shall allow for any degradation of image quality resulting from subsequent parts of the process.

4.4.6 Weight

The weight of a type font is its relative line thickness, ranging from light to extra bold. Font weight directly affects the number of dots or pixels used to display a character of an electronic image. Different weights can also be used to emphasize or reduce significance of text blocks or captions for the user.

4.4.7 Type family

Design variations on a basic typeface can include italic, condensed, expanded and others. Form designers should try to keep the number of type families used within a form to a minimum to project an uncomplicated appearance that is pleasing to the eye. EIM systems, and particularly OCR software, can also benefit from limited use of type families.

4.5 Machine printed stylized information

4.5.1 General

Information may be presented on a form as a bar code or in OCR or MICR characters These bar codes and stylized character sets are especially designed for automated processing and are machine readable with high accuracy.

Machine reading is not always wholly accurate. The degree of accuracy achievable can be improved if forms have error-checking features built into their design. Whenever possible, forms should be designed to use a second source of information for cross-checking. When calculation is involved, both subtotal and entry figures should appear on forms so that the processing system can recalculate the subtotal to compare it with the amount read. Other examples of information for cross-checking are account number/customer name and version number/issue date.

4.5.2 OCR fonts

OCR fonts such as Farrington 7B, OCR-A and OCR-B are available with numeric only and alphanumeric character sets for automated recognition. Data encoded using OCR fonts shall be printed according to the relevant International Standards, as given in Table 1.

Table 1 — OCR font standards

Code types supported	International Standard
OCR-A Numeric	ISO 1073-1
OCR-A Alphanumeric	ISO 1073-1
OCR-B Numeric	ISO 1073-2
OCR-B Alphanumeric	ISO 1073-2