L2/11-424

ISO/IEC JTC 1/SC 35 N 1743

DATE: 2011-10-28

ISO/IEC JTC 1/SC 35

User Interfaces

Secretariat: AFNOR

DOC TYPE:	**PDTR**
TITLE:	**PDTR 20007 "Information technology — Cultural and Linguistic Interoperability — Definition and relationship between symbols, icons, animated icons, pictograms, characters and glyphs”**
SOURCE:	**WG 5**
STATUS:	
ACTION ID:	**Vote**
DUE DATE:	**2012-01-29**
DISTRIBUTION:	**P, Def**
MEDIUM:	**E**
NO. OF PAGES:	

Secretariat of ISO/IEC JTC 1/SC 35 AFNOR — Philippe Magnabosco
11 rue Francis de Pressensé 93571 - La Plaine Cedex Saint-Denis - France
Telephone: +33 1 41 62 85 02; Facsimile: 33 1 49 17 90 00;
e-mail: philippe.magnabosco@afnor.org

ISO/IEC TC JTC 1/SC 35 N 1743

Date: 2011-10-28

ISO/IEC PDTR 20007

ISO/IEC TC JTC 1/SC 35/WG 5

Secretariat: AFNOR

Information technology — Cultural and Linguistic Interoperability — Definition and relationship between symbols, icons, animated icons, pictograms, characters and glyphs

Technologies de l'information — Interopérabilité linguistique et culturelle — Définition et relations entre symboles, icônes, icônes animées, pictogrammes, caractères et glyphes

Warning

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Document type: Technical Report
Document subtype:
Document stage: (20) Preparatory
Document language: E

C:\Users\emuller\AppData\Local\Temp\N1743_PDTR_20007_-_Definitions_and_relationship_between_symbols_icons_animated_icons_pictograms_characters_and_glyphs.doc STD Version 2.4a

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 2.

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 exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

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

ISO/IEC TR 20007 was prepared by Technical Committee ISO/TC JTC 1, *Information technology*, Subcommittee SC 35, *User interfaces*.

Introduction

It seems that many people misunderstand the limits of standardizing each of the concepts covered in this TR. As a case in point, ISO 7000 (Graphical symbols for use on equipment -- Index and synopsis) standardizes symbols with precise shapes, where, for example, the proportions are strictly established, while ISO/IEC 10646 (Universal Multiple-Octet Coded Character Set (UCS)) sometimes defines a coded character that maps an existing ISO 7000 symbols (which is practical for searching in technical documentation, for example), while any single coded character may be represented by a variety of different glyphs, thus open to a variety of shapes and proportions, as long as symbols remain recognizable (a glyph is not standardized for a given coded characters in this case, just the coding element is standardized unambiguously alongside its name). Some do not recognize that this is possible, but nevertheless both usages are internationally standardized and used with apparently contradicting requirements.

This TR tries to harmonize the apparent limitations of use of the different concepts involved in ISO and IEC context.

Information technology — Cultural and Linguistic Interoperability — Definition and relationship between symbols, icons, animated icons, pictograms, characters and glyphs

1 Scope

This Technical Report clearly defines each term related to ISO and IEC symbology in a single document and harmonizes difference of use and possible correspondence between different objects covering these concepts.

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.

IEC 60417 – Graphical symbols for use on equipment

IEC 80416-1:2008 – Basic principles for graphical symbols for use on equipment -- Part 1: Creation of graphical symbols for registration

ISO 7000 – Graphical symbols for use on equipment -- Index and synopsis

ISO/IEC 9541-1:1991 Information technology – Font information interchange – Part 1: Architecture

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

ISO/IEC 9995 – Information technology – Keyboard layouts for text and office systems (in particular part 7 on keyboard symbols)

ISO/IEC 10036 Information technology -- Font information interchange -- Procedures for registration of font-related identifiers

ISO/IEC 10646 – Information technology – Universal Multiple-Octet Coded Character Set

ISO/IEC FCD 11580 – Information technology – Framework for describing user interface objects, actions and attributes

ISO/IEC JTC 1/SC 35 FDIS 11581-10 – Information technology – Framework and general guidance

ISO/IEC 13251:2004 – Collection of graphical symbols for office equipment

ISO/IEC 14651 – Information technology -- International string ordering and comparison -- Method for comparing character strings and description of the common template tailorable ordering

ISO/IEC 14755 – Information Technology – Input methods to enter characters from the repertoire of ISO/IEC 10646 with a keyboard or other input device

ISO/IEC TR 15285:1998 – Information Technology – An operational model for characters and glyphs

ISO 17724:2003 – Graphical symbols – Vocabulary

ISO/IEC CD 24779-1 – Information Technology – Pictograms, Icons and Symbols for use with Biometric Systems -- Part 1: Overview

ISO/IEC WD 24779-2 – Information Technology – Pictograms, Icons and Symbols for use with Biometric Systems -- IEC 80416-1:2008, Basic principles for graphical symbols for use on equipment – Part 1: Creation of graphical symbols for registration; Part 2: Fingerprint applications

ITU-T Recommendation E.121 (2006), Pictograms, symbols and icons to assist users of the telephone and telefax services

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply. The definitions have been extracted from the different international standards that standardize them. References accompany each definition.

3.1
character
a member of a set of elements used for the organization, control, or representation of textual data; a character may be represented by a sequence of one or several coded characters [ISO/IEC 10646-1:2010]

3.2
code point
code position <deprecated>
any value in the Universal Character Set codespace [ISO/IEC 10646-1:2010]

NOTE 1 Values of the UCS codespace are integers (numbers) ranging from 0 to 10FFFF (hexadecimal [base 16] numeric representation)

3.3
coded character
an association between a character and a code point [ISO/IEC 10646-1:2010]

3.4
font
a collection of glyph images having the same basic design, e.g. Courier Bold Oblique. (ISO/IEC 9541-1: 1991)

3.5
glyph
a recognizable abstract graphic symbol which is independent of any specific design [ISO/IEC 9541-1:1991]

3.6
graphic character
a character, other than a control function, that has a visual representation normally handwritten, printed, or displayed [ISO/IEC 10646-1:2010]

3.7
graphic symbol (1)
the visual representation of a graphic character or of a composite sequence[ISO/IEC 10646-1:2010]

3.8
graphical symbol (2)
a visually perceptible figure with a particular meaning used to transmit information independently of language [IEC 80416-1, ISO 17724, etc.]

NOTE 1 Unique nature of graphical symbols is language independence. Therefore, the use of letters and punctuation marks as graphical symbol elements should be avoided.

NOTE 2 Graphical symbols are usually abstract representations that stand for something but that require learning on the part of users to take on their meaning.

3.9
icon (1)
a user interface [symbol / object] representing an object or a function of the computer system [FCD 11581-10 modified]

3.10
icon (2)
a symbol or combination of symbols in graphical user interfaces representing a function of the computer system [reference:?]

3.11
icon (3)
an object of manipulation of a function of the computer system through graphical user interfaces for computer applications [reference:?]

NOTE 1 Icons should be graphical representations that convey information with a minimum reliance on language.

NOTE 2 Icons have dynamic nature depending on the function of the computer system.

NOTE 3 Icons may be entirely abstract, like graphical symbols, or pictorial, like pictograms, or fall at some point between those extremes.

3.12
pictogram
a simplified pictorial representation, used to guide people and tell them how to achieve a certain goal [ITU-T Rec. E.121 Modified]

NOTE 1 Pictograms should be graphical representations that convey information with a minimum of reliance on language.

NOTE 2 Pictograms are, as far as possible, self-explanatory, and require little or no learning on the part of users.

NOTE 3 Pictorial representation can be two- or three-dimensional.

3.13
symbol
a visual (audible or tactile) sign, single letter, numeral, punctuation mark each of which has a fixed meaning [reference:?]

NOTE 1 Symbols are usually graphical representations that convey information with little reliance on language.

NOTE 2 Symbols are usually abstract representations that stand for something but that require learning on the part of users to take on their meaning.

NOTE 3 Examples of symbols are graphical symbols, graphic symbols, character symbols, chemical symbols, mathematical symbols, musical symbols, sex symbols, status symbols, tactile symbols, audible symbols.

4 Purpose of each different concept

The purpose of a **symbol** is to carry a meaning. A **pictogram** is a symbol as simple as possible whose purpose is to carry a symbolic meaning easy to understand for humans, ideally in an intuitive way, independently of language and culture. In ISO and IEC, standardized symbols are codified with strict forms. The purpose of a **glyph** is similar to that of a symbol, but goes beyond, in that it may also apply to a symbol that has become codified more abstractly over history, as for example glyphs that represent letters of an alphabet (which nowadays have no meaning by themselves, while a mere symbol is intended to have a meaning). Sets of glyphs usually grouped in a given style are called **fonts**. The purpose of a **character** is to group similar glyphs (even of different fonts) so that they all be recognizable as similar by humans, to all carry the same meaning, and to encompass all glyphs with the same meaning. Finally the purpose of a **coded character** is to codify a character for its transmission and processing (sorting, searching, matching, text structuring, etc.) by computers, independently of their presentation. The purpose of an **icon** is, on one side, to codify the computerized visual representation of a symbol, and on the other side to represent an entity associated with an object or and action in computer applications.

5 Limits and strengths of each different concept

Symbols standardized under ISO 7000 or IEC 80416 are destined to be reproduced directly on equipment (they are typically silk-printed or engraved). Their main limitation is also their strength: they shall be reproduced in their strict proportions and hence can not be confused with other symbols because no tolerance is allowed. The intent is that once learnt by humans, they are recognized without any doubt.

For computer applications, though, this strength may become a weakness: icons on computers are rendered using pixels, for example, and depending on screen resolution, the exact proportions may not be physically respected. Furthermore, the state of actions and objects (example: a “trash can” [metaphor for deleted objects] may be empty, full, available, in process of being emptied or restored, etc.) are represented by icons that may change shade, colour, even shape, and icons themselves may become animated objects, something that does not happen when a symbol is silk-printed on equipment with exact proportions.

At the other end of the spectrum, in the world of coded characters (standardized under ISO/IEC 10646 – the Universal Character Set [UCS]) , characters – which may occasionally correspond to ISO symbols (standardized under ISO 7000 or ISO 80416) – may be represented by any even vaguely corresponding glyph, depending on font style, or on rendition engines, so that humans can recognize them depending on environment, on accessibility requirements, or simply on personal preferences. That said, coded characters have a major strength: they can be searched, sorted, processed, and transformed by machines, without confusion. They can also be interchanged within different coding schemes, provided that their character names - the ultimate human identifiers that make two coded characters be considered the same - are shared in these two coding schemes. Because character names may vary between different human languages (and also have non-standardized synonyms within the same language) even for the Universal character set, the ultimate character identifier, nowadays, is its coded value in the UCS.

NOTE The name of a symbol standardized under ISO 7000 or IEC 80416 may not be the same as its name under ISO/IEC 10646 for different reasons: historical reasons, parallel development, unification purposes between similar-looking glyphs, and so on.

Symbols and pictograms also have their weakness per se: even if the intent of a symbol’s developer is that they be recognized intuitively, this may be strongly impacted by cultural and linguistic differences (a padlock may be considered as something which represent unavailability without a key in a given language while it only represents a fixed state [“Numlock”, for example, is ambiguously “decoded” in languages other than English and French] in another language); also, a symbol represented by letters or a word in one given language may

mean nothing to somebody who does not understand this language or does not understand a strongly-cultural-related abbreviation. Of course, once a symbol has been learned and become universal in usage, it becomes a powerful communication tool between speakers of different languages living in different cultural environments.

6 Properties of each different concept

6.1 Searchability

The entity that is easiest to search among all these concepts is the coded character. Pictures are difficult to retrieve otherwise, even with highly sophisticated pattern matching processes. For humans, alphabets, syllabaries and ideographic sets can be visually searched within a sorted list if their collating sequence is well established and the object of systematically learned searching methods. For character collation, one may refer to ISO/IEC 14651 which standardizes a customizable method for sorting character strings based on the UCS.

6.2 Presentation

Symbols, pictograms, icons and even characters can be presented under different forms: visual, audio, tactile, with different levels of precision. For accessibility purposes, one needs to find ways to make sure that the different representations are recognized without ambiguity by humans. Fuzziness is possible if the goal is more or less informal.

6.3 Shape, precise representation, fuzzy representation, encoding, animation, temporal representation, etc.

Some concepts are for a number of purposes not well defined from the user's point of view. Items that are technically different may be conceived as the same by a user. For example an uppercase A and a lowercase a may be considered the same, or the letter a with any accent may be considered the same for example when searching. Also different spelling (for example of transcribed Cyrillic or Arabic names like Chernobyl/Tjernobyl or Gadaffi/Kadafi) should be considered the same (fuzzy match). A character or graphic symbol looking similar but in different encodings could be considered the same in some cases. ISO/IEC-14651tailored data may be used to choose different levels of matching precision (for example irrespective of letter case, or irrespective of letter accents, or precise match). More complex matching is often done by industry Internet search engines.

Similar fuzzy matching requirements may be needed for the glyph concepts, for instance finding a related glyph in another font for missing glyphs, or finding a replacement font.

7 Relationship between the different concepts

The flow of relationship between the different concepts can be simplified as follows: a simple sketch (which can be considered as an original glyph and as an original pictogram) can become a standardized symbol and later be encoded as a coded character and become semantically searchable directly in documentation. In parallel, such a sketch can be represented by a computerized icon used as a metaphor to represent an object or an action processed by a computer.

The description in clause 4, and the definitions in clause 3 prescribes further relations between the different concepts. ISO/IEC TR 24785 has a taxonomy of some of the concepts.

The “character” family of concepts has a symbolic and a coded counterpart. The symbolic concept does not need coding, it is essential for human recognition, and they are the fundamental concept for computing. The atomic concept here is the character concept. The coded counterpart has the coded character as the atomic concept. Then a coded character set is a collection of coded characters.

NOTE Further concepts are encodings as a set of coded characters and coded character set switching rules; a transformation format (like UTF-8) can then be applied on the encoding, and finally at transfer-encoding can be applied on the transformation format, this whole hierarchy is what is normally processed, for instance on the Internet. .Characters may

be represented by character strings, for example ISO/IEC 8879 SGML and ISO/IEC 15445 HTML has ü for ü, and the ISO/IEC 9899 C and ISO/IEC 14884 C++ has \Uxxxx for a UCS character in the source input.

The “glyph” family of concepts has the glyph as its atomic concept, and a “font” generally groups different glyphs sharing some style attributes. Each glyph is normally connected to a character or graphic symbol. Correspondence can often be found from a glyph ID in the ISO/IEC 10036 registry to a ISO/IEC 10646 character.

Icons (static or animated) and pictograms are not collected into different bigger collections, and are not normally related to characters nor glyphs.

8 Input, Process and Output Considerations (from drawing to search, via representation on different media and encoding)

The “character” family of concepts are the concepts most oriented towards computer processing. ISO/IEC TR 11017 describes ways of processing and APIs for this family of concepts. Further, there are needs for accessibility, for instance blind or deaf persons, both for inputting and outputting. Some of these methods may also be useful in a general environment, for example spoken input or output. For complicated scripts like Chinese, Japanese and Korean, or for generalized input or restricted input media like mobile telephones, specific input methods may be applied. APIs for processing characters for transliteration and transcribing (for example with limited display capabilities or limited reading recognition skills of the reader) may also be available. A very basic universal input method is specified in ISO/IEC 14755. More user-friendly keyboard handling and input methods exist that are adapted to specific linguistic or specialized environments. ISO/IEC 9995-9, under development, aims at providing methods to have elaborate keyboard handling and input methods accessible for multi-writing-system input of the universal character set (ISO/IEC 10646) and other character sets.

APIs processing glyphs and fonts are required in some areas, for example in office processing packages, or general desktop software.

9 Databases, Sets and Repositories

There are a number of ressources in ISO and IEC and other places where the concepts of this Technical Report and instantiations thereof can be found. Examples are:

ISO/IEC 10646 can be seen as a universal registry of characters, as its scope comprises all characters in the world. Different coded character sets (not all International Standards) are registered with ISO/IEC 2375 and ISO/IEC 15897. Character repertoires and encodings are registered in ISO/IEC 15897. Some industry registration databases also exist (notably the IANA registry).

ISO/IEC maintains a data base of graphical symbols and their descriptions based on International Standards ISO 7000 and IEC 60417.

ISO 7000 and ISO 7001 define a registry of pictograms, symbols and icons and the subsequent registries' content? [to be validated by national bodies during ballot or to be removed]

ISO/IEC 10036 defines a registry of fonts and related glyphs, with also a relation to graphic symbols (composed of one or more characters).

10 Further considerations

[Open to national contributions during ballot – if no contribution this section will be removed]

[Examples of topics: More about icons, pictograms, what to do with animated icons when relating them to symbols or characters, relationship with descriptive speech, etc.]