TECHNICAL REPORT

ISO/IEC TR 20007

First edition
2014-05-01

Information technology — Cultural and linguistic interoperability — Definitions and relationship between symbols, icons, animated icons, pictograms, characters and glyphs

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

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC 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 Technical Report. As a case in point, ISO 7000 standardizes graphical symbols with precise shapes, where, for example, the proportions are strictly established, while ISO/IEC 10646 sometimes defines a coded character that maps an existing ISO 7000 symbol (which is practical for searching in technical documentation, for example). However, any single coded character can 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 character in this case, the coding element is standardized unambiguously alongside its name only). Some do not recognize that this is possible; nevertheless, both usages are internationally standardized and used with apparently contradicting requirements.

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

Information technology — Cultural and linguistic interoperability — Definitions 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 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

2.1
character
member of a set of elements used for the organization, control, or representation of textual data

Note 1 to entry: A character may be represented by a sequence of one or several coded characters

[SOURCE: ISO/IEC 10646:—, 4.5]

2.2
code point
DEPRECATED: code position
value in the Universal Character Set codespace

[SOURCE: ISO/IEC 10646:—, 4.10]

Note 1 to entry: Values of the Universal Character Set (UCS) codespace are integers (numbers) ranging from 0 to 10FFFF (hexadecimal [base 16] numeric representation)

2.3
coded character
association between a character and a code point

[SOURCE: ISO/IEC 10646:—, 4.8]

2.4
font
collection of glyph images having the same basic design, e.g. Courier Bold Oblique

[SOURCE: ISO/IEC 9541-1:1991]

2.5
glyph
recognizable abstract graphic symbol which is independent of any specific design

[SOURCE: ISO/IEC 9541-1:1991]

2.6
graphic character
character, other than a control function, that has a visual representation normally handwritten, printed, or displayed

[SOURCE: ISO/IEC 10646:—, 4.29]

2.7
graphic symbol
visual representation of a graphic character or of a composite sequence

[SOURCE: ISO/IEC 10646:—, 4.30]

2.8
graphical symbol
visually perceptible figure with a particular meaning used to transmit information independently of language

[SOURCE: IEC 80416-1:2008, 3.4, ISO 17724:2003, 31]

Note 1 to entry: The 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 to entry: Graphical symbols are usually abstract representations that stand for something, but that require learning on the part of users to take on their meaning.

2.9.1
icon
user interface (symbol or object) representing an object or a function of the computer system

[SOURCE: ISO/IEC 11581-10:2010, 3.4, modified]

2.9.2
icon
symbol or combination of symbols in graphical user interfaces representing a function of the computer system

2.9.3
icon
object of manipulation of a function of the computer system through graphical user interfaces for computer applications

Note 1 to entry: Icons should be graphical representations that convey information with a minimum reliance on language.

Note 2 to entry: Icons have dynamic nature depending on the function of the computer system.

Note 3 to entry: Icons may be entirely abstract, such as graphical symbols, or pictorial, such as pictograms, or fall at some point between those extremes.

2.10
pictogram
simplified pictorial representation, used to guide people and tell them how to achieve a certain goal

[SOURCE: ITU-T Recommendation E.121, modified]

Note 1 to entry: Pictograms should be graphical representations that convey information with a minimum of reliance on language.

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

Note 3 to entry: Pictorial representation can be two- or three-dimensional.

2.11
symbol
visual (audible or tactile) sign, single letter, numeral, punctuation mark each of which has a fixed meaning

Note 1 to entry: Symbols are usually graphical representations that convey information with little reliance on language.

Note 2 to entry: 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 to entry: Examples of symbols are graphical symbols, graphic symbols, character symbols, chemical symbols, mathematical symbols, musical symbols, sex symbols, status symbols, tactile symbols and audible symbols.

3 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 representing letters of an alphabet (which, at the time of publication of this Technical Report, 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 carry all 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 hand, to codify the computerized visual representation of a symbol, and on the other hand, to represent an entity associated with an object or and action in computer applications.

4 Limits and strengths of each different concept

Symbols standardized under ISO 7000 or IEC 80416-1 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 cannot 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 might become a weakness: icons on computers are rendered using pixels, for example, and depending on screen resolution, the exact proportions might not be physically respected. Furthermore, the state of actions and objects (for example: a "trash bin" [metaphor for deleted objects] might be empty, full, available, in the process of being emptied or restored, etc.) are represented by icons that might change shade, colour, even shape, and icons themselves might 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 UCS), characters, which might occasionally correspond to ISO symbols (standardized under ISO 7000 or IEC 80416-1), 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 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 80416-1 might 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 might be strongly impacted by cultural and linguistic differences (a padlock can be considered 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 can mean nothing to somebody who does not understand this language or does not understand a strongly cultural-related abbreviated term. 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.

5   Properties of each (different) concept

5.1   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 make reference to ISO/IEC 14651, which standardizes a customizable method for sorting character strings based on the UCS.

5.2   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.

5.3   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 can be conceived as the same by a user. For example an uppercase A and a lowercase a might be considered the same, or the letter a with any accent can be considered the same for example when searching. Also different spelling (for example the 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 14651-tailored 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.

6   Relationship between the different concepts

The flow of relationship between the different concepts may be simplified as follows: a simple sketch (which can be considered an original glyph and as an original pictogram) can become a standardized symbol and later be potentially (although not necessarily) encoded as a coded character and become semantically searchable directly in documentation (although searching can also be achieved by more

complex mechanisms than character coding). 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 3, and the terms and definitions in Clause 2 prescribes additional 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   Additional concepts are encodings as a set of coded characters and coded character set switching rules; a transformation format (such as 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 can 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 14882 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.

7   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. Furthermore, 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, such as Chinese, Japanese and Korean, or for generalized input or restricted input media such as 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, which are adapted to specific linguistic or specialized environments, exist. ISO/IEC 9995-9, under development, aims at providing elaborate methods to input multiple writing systems using the universal character set (see ISO/IEC 10646) or other character sets, by using a keyboard.

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

8   Databases, sets and repositories

There are a number of resources in ISO and IEC and other places where the concepts of this Technical Report and instantiations thereof can be found. The following are examples.

ISO/IEC 10646 can be seen as of the most comprehensive compendium of characters in existence, as its scope aims at standardizing all characters in the world (past, present and future) in a single code. 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 database 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.

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