INTERNATIONAL STANDARD

ISO/IEC 14755

First edition
1997-08-15

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

Technologies de l'information — Méthodes de saisie de caractères du répertoire de l'ISO/CEI 10646 à l'aide d'un clavier ou d'autres unités d'entrée



Reference number
ISO/IEC 14755:1997(E)

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS

Not for Resale

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. Draft International Standards adopted by the joint technical committee are circulated to the national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 14755 has been prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 18, *Document processing and related communication.*

© ISO/IEC 1997
All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

ISO/IEC Copyright Office • Case postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

Introduction

Today, there is a well-known method in existence for inputting characters foreign to a given keyboard on certain personal computers. However this method is code-dependent and is limited to 8-bit coded character sets. There is a need to standardize such a method independently of coding even for these limited sets of characters.

There is also an International Standard, ISO/IEC 9995-3, for inputting on a standard 48-key keyboard the repertoire of characters belonging to those European languages using the Latin script. But that standard is limited to the Latin script, even if it opens the door to the defining supplementary groups for other scripts. In the meantime, until other groups are well defined and documented, there should be an easy standard way to enter non-Latin characters in a code-independent fashion. This would avoid the multiplicity of such methods, a situation that is never desirable for end-users.

Furthermore, ISO/IEC JTC 1 recently published ISO/IEC 10646, "Universal Multiple-Octet Coded Character Set (UCS)", which is a superset of the repertoires of all standard character sets published so far by ISO/IEC JTC 1. For this one large character set (UCS), there is no standard input method in existence today. But there will be an increasing need for one, which would also solve the problem of code independence.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

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

1 Scope

This International Standard defines methods that allow entry of characters belonging to the repertoire of single and multi-octet coding standards such as ISO/IEC 10646 in a code independent manner using a keyboard or other input/output devices. It is also expected that for implementations of different coding schemes of character sets, this method will be usable, provided that the target character sets have repertoires that are subsets of the universal multiple-octet coded character set (ISO/IEC 10646) or of any other standard character set.

More specifically, this International Standard defines four methods, which can be implemented in different scenarios according to the conformance clause:

- a basic method for entering a character which involves using its bit representation (canonical or abbreviated form) in ISO/IEC 10646 as a catalog number, whichever underlying code is used for that character;
- a method for entering standard keyboard symbols representing the functions used on the keyboard; this method is intended for entering characters corresponding to the visual representation of the keyboard function symbols (according to ISO/IEC 9995-7 symbols and ISO/IEC 10646 characters), with the help of the function keys themselves;
- a screen-selection entry method for selecting a character displayed on a screen for data entry;
- a feedback method that allows exact identification of characters shown on a screen, for subsequent data entry;

This International Standard is intended to complement existing national keyboard layouts or existing input methods optimized for national use. Hence it does not replace any national keyboard entry requirement but is rather a tool to ease entry of the complete character repertoire of ISO/IEC 10646 with the help of already existing national keyboards.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange.*

ISO/IEC 6937:1994, *Information technology — Coded graphic character set for text communication — Latin alphabet.*

ISO/IEC 8859-1:1987[1)], *Information processing — 8-bit single-byte coded character sets — Part 1: Latin alphabet No. 1.*

1) Currently under revision.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS Not for Resale

ISO/IEC 9995-1:1994, *Information technology — Keyboard layouts for text and office systems — Part 1: General principles governing keyboard layouts.*

ISO/IEC 9995-3:1994, *Information technology — Keyboard layouts for text and office systems — Part 3: Complementary layouts of the alphanumeric zone of the alphanumeric section.*

ISO/IEC 9995-7:1994, *Information technology — Keyboard layouts for text and office systems — Part 7: Symbols used to represent functions.*

ISO/IEC 10646-1:1993, *Information technology — Universal Multiple-Octet Coded Character Set — Part 1: Architecture and Basic Multilingual Plane.*

3 Definitions

For the purposes of this International Standard, the following definitions apply.

3.1 beginning sequence: a specific sequence or combination of typing of keys, or an implementation-defined mechanism, the effect of which is to put the system in a mode that will allow entering a specific input sequence according to a method described in this International Standard.

3.2 canonical form: the form with which characters of the UCS are specified using 32 bits (or 8 hexadecimal digits) to represent each character.

3.3 catalog number: identification of a character by its canonical form, with the difference that leading zeroes may be omitted.

3.4 compose character: a function which selects a graphic character which has not been allocated on the keyboard by associating other allocated characters.

3.5 control key: a key representing the function *Control* as described in ISO/IEC 9995-7 (symbol: ⎈).

3.6 ending sequence: a specific sequence or combination of typing of keys, or an implementation-defined mechanism, the effect of which is to terminate the generating of a character whose selection was begun by the *beginning sequence* in conformance with a method described in this International Standard .

3.7 hexadecimal: hexadecimal numbering is a counting system analogous to the decimal system, but uses base 16 instead of base 10. In the hexadecimal counting system, when a count reaches 16 in any number position a carry of 1 occurs into the next more significant number position, instead of when that count reaches 10 as in the decimal system. As there are no digits available beyond 9, the first 6 letters of the Latin alphabet (or of any alphabet if the Latin script is not used) are used to represent the extra hexadecimal "digits" 10 (A), 11 (B), 12 (C), 13 (D), 14 (E), 15 (F). In this International Standard, hexadecimal numbers are used to refer to the UCS, the hexadecimal coding of which is considered as equivalent to a catalog numbering system to select characters. Hexadecimal notation exists as a shortcut to represent groups of 4 bits (there are 16 combinations possible with permutations of 4 bits, whose values can only be a zero or a one); it also takes less characters to express a number in hexadecimal than in decimal.

3.8 level: a logical state of a keyboard providing access to a collection of graphic characters or elements of graphic characters. Usually these graphic characters or elements of graphic characters logically belong together, such as the capital forms of letters. In certain cases the level selected may also affect function keys.

3.9 Level 2 Select: a function which selects the set of characters or functions allocated to the level 2 of the keyboard in an active group (symbol: ⇧).

3.10 UCS: the Universal Multiple-Octet Coded Character Set standard known as ISO/IEC 10646-1 and its extensions to come.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

4 Symbols and abbreviations

In this International Standard, identification of specific characters of the UCS repertoire will be by means of symbols of the form Uxxxxxxxx . In this form, the occurrences of xx which follow the letter "U" represent the catalog number of a character. Such a catalog number refers to the UCS. This is in practice a way to be independent of any specific code. The UCS contains all the coded graphic characters ever defined by ISO/IEC. Whenever possible other short mnemonic identifiers will be used in comments in addition to the printing of the characters themselves.

When the name of a character is used (standard or conventional) in this text, it is surrounded in the text by the LESS-THAN SIGN and the GREATER-THAN SIGN, as for example, in the string <*SPACE*>, which represents the character **SPACE**.

5 Requirements

Prerequisites

The keyboard in use shall have an alphanumeric section. This alphanumeric section shall provide a space bar (which generates the character <*SPACE*>), the ten decimal digits and the first 6 letters of the Latin alphabet if the Latin script is used, or the first six letters of any other alphabet if a different script is used. These are used to represent hexadecimal numbers (see definition for hexadecimal numbering system), the catalog numbers of all UCS characters. The system on which the keyboard operates shall be programmed in such a way that the sequence of events required by some methods (in particular method 5.2) be possible. This International Standard is hence typically not applicable to a mechanical typewriter.

5.1 Basic method

The basic method operates as follows:

While the *beginning sequence* has been activated, typing the catalog number of a character and ending the "catalog number" by the *ending sequence* shall generate a coded graphic character equivalent to the one corresponding to this "catalog number" in the UCS or indicate an error in compliance with 5.5 if the selected character cannot be generated by the system.

Whenever possible no distinction should be made between upper case and lower case letters (or shape variants) for entering the first 6 letters of the alphabet as hexadecimal digits.

If, with this method, an option for successive multiple character entry is provided by a conforming system or application, then it is recommended that the SPACE BAR should be used as the equivalent of a temporary ending sequence ***that generates the character previously entered and then at the same time as the equivalent of a new*** beginning sequence ***for another character. This acts as an optional locking mechanism for the mode in which the basic method operates.***

NOTE —In the following examples, it is assumed that the *beginning sequence* consists in the combined use of keys *Level 2 Select* (⇧) and *Control* (❄). The ending sequence corresponds to the releasing of these keys. This should only be considered as an example of implementation for these sequences.

Example a. Entering character **À**

1. Depress the *Level 2 Select* (⇧) key and hold it.
2. Depress the *Control* key (❄) and hold it.
3. Type the sequence 00C0.
4. Release the *Control* (❄) key and the *Level 2 Select* (⇧) key.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS Not for Resale

Result: Character **À** will be generated, as the entered sequence represents the UCS catalog number of the letter **À**, regardless of the code used in the system, provided that this character is available in the actual character set.

Please note that the canonical form of this character is 000000C0, but that the user has here chosen to not type the first four zeros. It would also have been allowed not to type any leading zero at all (typing C0 would have been valid).

If the system uses the ISO/IEC 8859-1 coding, the bit combination whose notation is 12/00 will correspond to the actual coding.

If the system uses the ISO/IEC 6937 code, the character will be coded on two octets and will correpond to two consecutive values whose notation in that standard is 12/01, 04/01, which respectively represent a non-spacing grave accent and the upper-case latin letter A, unaccented.

If the system does not have character **À** available or has no integral support for it in any manner, the computer should issue a warning, such as an emitted beep sound indicating that the character has not been accepted.

Example b. Entering the Chinese character **Shàn** (善Japanese **Zen**, Korean **Seon**)

1. Depress the *Level 2 Select* (⇧) key and hold it.

2. Depress the *Control* (❆) key and hold it.

3. Type the sequence 5584.

4. Release the *Control* (❆) key and the *Level 2 Select* (⇧) key.

Result: Character 善 whose canonical form is 00005584 will be generated, as it represents the UCS catalog number of the Chinese character whose Mandarin phonetics corresponds to **Shàn**, in the Latin script transliteration system known as "hanyu pinyin", officially used in China. This character is also known in Japan and can be transliterated, in Japanese Romaji, as **Zen**. It is also used in Korea and pronounced **Seon**. The ideograph means **kindness** in English.

This character will be generated regardless of the code used in the system, provided that this character is available in the actual character set.

Please note that the canonical form of this character is 00005584, but that the user has here chosen to not type the first four zeros, which is a valid practice according to this International Standard.

According to ISO/IEC 10646 the following are actual examples of implementations in national standards:

If the system uses GB 2312-1980 coding, the coding generated (in hexadecimal notation) will be 4946.

If the system uses JIS X 0208-1990 coding, the coding generated (in hexadecimal notation) will be 4131.

If the system uses KS C 5601-1987 coding, the coding generated (in hexadecimal notation) will be 603C.

If the system does not have this character available or has no integral support for it in any manner, the computer should issue a warning, such as an emitted beep sound indicating that the character has not been accepted.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS Not for Resale

Example c. Entering character \

1. Depress the *Level 2 Select* (⇧) key and hold it.
2. Depress the *Control* (❃) key and hold it.
3. Type the sequence 5C.
4. Release the *Control* (❃) key and the *Level 2 Select* (⇧) key.

Result: Character \ will be generated, as it represents the UCS catalog number of <REVERSE SOLIDUS>, regardless of the code used in the system, provided that this character is available in the actual character set.

Please note that the canonical form of this character is 0000005C, but that the user has here chosen to not type the first six zeros, which is a valid practice according to this International Standard.

If the system uses the ISO/IEC 8859 coding, the bit combination whose notation is 05/12 will correspond to the actual coding (this is also notation 5/12 according to ISO/IEC 646 7-bit-coded character set).

If the system does not support character \ or has no integral support for it in any manner, the computer should issue a warning, such as an emitted beep sound indicating that the character has not been accepted.

5.2 Keyboard symbols entry method (formerly known as *Meta-entry method*)

Certain function keys of keyboards have been associated a symbol in ISO/IEC 9995-7. For those symbols which are also associated an appropriate character in the UCS, the entry of the character can be done using the function key itself. This can be used as follows, to select that character (say, for searching that character in online documentation, or simply editing such documentation for users of internationalized applications):

After the *beginning sequence* has been activated, depressing a function key **and releasing it** shall enter the character assigned in the UCS to the symbol representing this function in ISO/IEC 9995-7 (provided that the character exists in the UCS). As some functions are activated with the help of a qualifier key such as *Level 2 Select* (⇧) key, itself a function key, the resulting character shall be generated only once all keys have been released by the user.

NOTE — In the following examples, it is assumed that the *beginning sequence* consists in the combined use of keys *Level 2 Select* (⇧) and *Control* (❃). This should only be considered as an example of implementation for this sequence.

Example a. Entering the character **<RIGHTWARDS ARROW TO BAR>** (⇥)

1. Depress the *Level 2 Select* (⇧) key and hold it.
2. Depress the *Control* (❃) key.
3. Release both the *Control* (❃) key and the *Level 2 Select* (⇧) key.

   At this stage, the keyboard is in a logical state where depressing any function key combination will potentially generate a character instead of activating the associated function.

4. Depress the *Tabulation Right* (⇥) function key.
5. Release the *Tabulation Right* (⇥) function key.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS Not for Resale

Result: Character **<RIGHTWARDS ARROW TO BAR>** (⇥) will be generated, as it represents the UCS character whose normal presentation form corresponds to ISO/IEC 9995-7 symbol for the *Tabulation Right* (⇥) function. The canonical value of this character is 000021E5 in the UCS.

If the system does not have this character available or has no integral support for it in any manner, the computer should issue a warning, such as an emitted beep sound indicating that the character has not been accepted.

Example b. Entering the character **<UPWARDS WHITE ARROW>** (⇧)

1. Depress the *Level 2 Select* (⇧) key and hold it.
2. Depress the *Control* (✲) key.
3. Release both the *Control* (✲) key and the *Level 2 Select* (⇧) key.

   At this stage, the keyboard is in a logical state where depressing any function key combination will potentially generate a character instead of activating the associated function.
4. Depress the *Level 2 Select* (⇧) function key.
5. Release the *Level 2 Select* (⇧) function key.

Result: Character **<UPWARDS WHITE ARROW>** (⇧) will be generated, as it represents the UCS character whose normal presentation form corresponds to ISO/IEC 9995-7 symbol for the *Level 2 Select* (⇧) function. The canonical value of this character is 000021E7 in the UCS.

Example c. Entering the character **<LEFTWARDS ARROW TO BAR>** (⇤)

1. Depress the *Level 2 Select* (⇧) key and hold it.
2. Depress the *Control* (✲) key.
3. Release both the *Control* (✲) key and the *Level 2 Select* (⇧) key.

   At this stage, the keyboard is in a logical state where depressing any function key combination will potentially generate a character instead of activating the associated function.
4. Depress the *Level 2 Select* (⇧) function key.
5. Depress the *Tabulation Right* (⇥) function key.
6. Release all keys.

NOTE — This scenario is an example only and does not imply any obligation to have a keyboard on which the *Tabulation Left* (⇤) function is allocated to the same key as the *Tabulation Right* (⇥) key nor that the *Tabulation Right* (⇥) function is allocated a dedicated key. It only reflects a common practice.

Result: Character **<LEFTWARDS ARROW TO BAR>** (⇤) will be generated, as it represents the UCS character whose normal presentation form corresponds to ISO/IEC 9995-7 symbol for the *Tabulation Left* (⇤) function, assumed here to be placed on the same key as the *Tabulation Right* (⇥) function (this one being accessible with level 1 of the keyboard, the first one being accessible after selection of level 2). The canonical value of this character is 000021E4 in the UCS.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS Not for Resale

If the system does not have this character available or has no integral support for it in any manner, the computer should issue a warning, such as an emitted beep sound indicating that the character has not been accepted.

5.3 Screen-selection entry method

In the screen-selection entry method the user may select for entry a character that is displayed on a screen associated with the input keyboard or device. The selection is achieved by the use of a pointing device and a selection mechanism, such as a button that may be pressed (or clicked). When the user selects a character as described above, the system shall generate the coded character that corresponds to the character selected on the screen. The feedback method described in 5.4 below shall be used to indicate to the user the actual coded character in use.

This basic method is passive by nature, but it is recommended that its implementation be accompanied by user-friendly services to invoke small subsets of characters selected according to precise searching criteria, as for example the character shape, the han character radicals, the identification of an alphabet, the spacialized nature of certain series (for example mathematical operators). Once a subset selected, it is recommend to give access to a menu allowing to selcet the desired character for example in typing a number, in addition to the possibility to use a pointing device to select it.

Examples of subset invocation

Example a. Direct subset identification

One system may extend the use of the *beginning sequence* to enter a mode in which, if non hexadecimal characters are used as the first character entered, the sequence entered could identify the subset in full, as for example, the name of subsets documented in the UCS, such as IPA EXTENSION (International Phonetic Alphabet), GREEK, MATHEMATICAL OPERATORS, and so on.

Example b. Sub-method identification

A sub-method identifier could also be used in the same mode, such as the keyword RADICAL, for example, which could give access to the traditional Han radicals, after which answering the menu, one could enter the number of remaining strokes in the character, which would give access to a second menu with a final list for selecting the right character.

5.4 Feedback method for identifying displayed characters for later input

In the feedback identification method the user may obtain the identification of a character that is displayed on the screen associated with the input keyboard or device. The selection of the character to be identified is achieved by the use of a pointing device and a selection mechanism, such as a button that may be pressed (or clicked). When the user selects a character as described above, the system shall inform the user of the identification of the selected character. The identification may be done by means of a display on the screen, or by voice-output, or by any other appropriate means. The identification shall include the catalog number of the character as a sequence of hexadecimal digits. If the system uses an internal character coding method different from that of the UCS, then it is recommended that the indications to the user should also include:

— the internal code position of the character in a suitable notation, and
— the name of the character in the natural language of the user ***(when this is applicable)***.

Example scenario

Let's suppose that shape **A** is displayed on a screen to represent LATIN CAPITAL LETTER A and that the user is French-speaking. Let's also suppose that feedback is generated on a screen under the form of a virtual slide which is displayed as long as a character is pointed by a mouse, used as pointing device.

The user points this character with his mouse.

The computer then displays the canonical form 00000041 whatever the coding actually used by the system.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS Not for Resale

According to the recommendation made in this International Standard, in this scenario, the actual coding used by the system is also displayed (it could be, as in a well known actual example, hexadecimal C1) as well as the string **LETTRE MAJUSCULE LATINE A**, corresponding to the character name in the user's natural language.

This example is used to show the necessity of such a function. The UCS is a huge character set, in which many characters can share the same shape, or external representation, while they actually represent different conceptual characters, which is reflected in the internal coding as also being different.

In this example, shape **A** is displayed on a display screen, and the conceptual character to which this shape corresponds can be three-fold: capital Latin letter A, capital Greek letter Alpha or capital Cyrillic letter A. In order to have provisions for this the UCS uses three canonical values for these different characters. Hence it becomes important for a user to know what the actual character is if it is to be used for input later, and for retrieval at a much later stage. The feedback method will positively identify the right character without any possible ambiguity for further input.

5.5 Unexpected conditions

It is recommended that whenever a character entered by any of one of these aforementioned methods cannot be supported by the applicative environment, a warning should be issued to the user by either sound or visual indication, or by any sensible means of attracting the user's attention. Whenever possible, data integrity of the character entered should be preserved even if it is not possible to display it or if it is not possible to process it properly with full applicative support. The preservation of data integrity should then allow interchanging of the character to another application which can better deal with this character. This recommendation can be achieved by a variety of different technical means which are implementation-defined.

6 Conformance

An application, or a complete system performing functions for all applications running under it, used in conjunction with a keyboard, conforms to this International Standard if it conforms to either of two scenarios:

Scenario A shall be compliance with 5.1 and 5.4 and 5.5;

Scenario B shall be compliance with 5.3 and 5. 4 and 5.5;

Both scenarios may also optionally include compliance with 5.2.

Any claim of conformance to this International Standard shall declare to which scenario(s) it is compliant and whether it offers support for 5.2.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS
Not for Resale

© ISO/IEC

ICS 35.040

Descriptors: data processing, information interchange, text processing, graphic characters, coded character sets, input-output units, input.

Price based on 8 pages

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS

Not for Resale