
**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*

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

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.

This document is a preview generated by EVS

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 8859-1:1987¹⁾, *Information processing — 8-bit single-byte coded character sets — Part 1: Latin alphabet No. 1*.

1) Currently under revision.

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.