

---

---

**Information technology — Coded graphic  
character set for text communication —  
Latin alphabet**

*Technologies de l'information — Jeu de caractères graphiques codés pour  
la transmission de texte — Alphabet latin*

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 6937:2001](https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001)

[https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-  
1363fb9c7d1d/iso-iec-6937-2001](https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001)



**PDF disclaimer**

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO/IEC 6937:2001

<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001>

© ISO/IEC 2001

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 either ISO at the address below or ISO's member body in the country of the requester.

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](mailto:copyright@iso.ch)  
Web [www.iso.ch](http://www.iso.ch)

Printed in Switzerland

## Contents

	Page
Foreword	iv
Introduction	v
1 Scope	1
2 Conformance and implementation	1
2.1 Conformance	1
2.2 Implementation	2
3 Normative references	2
4 Terms and definitions	3
5 Notation, code table and names	5
5.1 Notation	5
5.2 Code table	5
5.3 Names	5
6 Specifications of SPACE, NO-BREAK SPACE and SOFT HYPHEN	6
7 Composition of the character repertoire	6
8 Specification of the coded character set	6
8.1 Character sets	6
8.2 Explanations concerning the code table	7
8.3 Coded representations of the graphic characters of the repertoire	7
9 Graphic character subrepertoires	8
10 Identification of options	9
10.1 Purpose and context of identification	9
10.2 Identification of coding method	9
10.3 Identification of primary and supplementary sets	9
10.4 Identification of subrepertoire	9
Annex A (normative) 7-bit code	20
Annex B (informative) Method of definition of short identifiers of this International Standard	23
Annex C (informative) Use of non-spacing diacritical marks	33
Annex D (informative) Use of Latin alphabetic characters in various languages	34
Annex E (informative) Alternative coded representation of the repertoire with no non-spacing diacritical marks	38
Annex F (informative) Main differences between the 1994 (second) edition of ISO/IEC 6937 and the present (third) edition of this International Standard	39
Bibliography	40

## 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.

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

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 national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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

International Standard ISO/IEC 6937 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

This third edition cancels and replaces the second edition (ISO/IEC 6937:1994), which has been technically revised.

Annex A forms a normative part of this International Standard. Annexes B, C, D, E and F are for information only.

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**  
<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001>

## Introduction

This International Standard specifies a repertoire of graphic characters and their coded representations, for use in text communication.

Although, in general, text (see 4.16) consists of characters and pictures, this International Standard applies only to text made up of characters.

The specifications are based on 8-bit coding; Annex A specifies the 7-bit code for the character set of this International Standard.

Other annexes include:

a) a description of the method used to define a short identifier for each character specified in this International Standard (Annex B);

b) a summary of the use of non-spacing diacritical marks in combination with letters of the basic Latin alphabetic characters (Annex C);

c) a summary of the use of Latin alphabetic characters in various languages (Annex D);

d) an alternative coded representation of the repertoire with no non-spacing diacritical marks (Annex E);

e) a summary of differences between the 1994 (second) edition of ISO/IEC 6937, and the present (third) edition of this International Standard (Annex F);

f) a bibliography.

ISO/IEC 6937:2001  
<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001>

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO/IEC 6937:2001

<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001>

# Information technology — Coded graphic character set for text communication — Latin alphabet

## 1 Scope

This International Standard

- a) specifies the coded representation of the characters;
- b) specifies a repertoire of the Latin alphabetic and non-alphabetic characters for the communication of text in many European languages using the Latin script;
- c) specifies rules for the definitions and use of graphic character subrepertoires, i.e. subsets of the specified character repertoire.

## 2 Conformance and implementation

### 2.1 Conformance

#### 2.1.1 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with this International Standard if all coded representations of characters within that CC-data-element conform to the mandatory requirements of this International Standard.

A claim of conformance shall identify:

- the subrepertoire in accordance with clause 9, if one has been adopted,
- the 7-bit coding in accordance with Annex A, if it has been adopted.

#### 2.1.2 Conformance of devices

A device is in conformance with this International Standard if it conforms to the requirements of 2.1.2.1 and either or both 2.1.2.2 and 2.1.2.3 below.

##### 2.1.2.1 Device description

A device that conforms to this International Standard shall be the subject of a description that identifies the means by which the user may supply characters to the device, or may recognize them when they are made available to the user, as specified respectively in 2.1.2.2 and 2.1.2.3 below.

##### 2.1.2.2 Originating devices

An originating device shall allow its user to supply any sequence of characters of the character repertoire, and shall be capable of transmitting their coded representations within a CC-data-element.

##### 2.1.2.3 Receiving devices

A receiving device shall be capable of receiving and interpreting any coded representation of characters that are within a CC-data-element, and that conform to 2.1.1 of this International Standard, and shall make the corresponding characters available to its user in such a way that the user can identify them among those of the repertoire, and can distinguish them from each other.

## 2.2 Implementation

The use of this character set requires definitions of its implementation in various media. For example, these could include magnetic and optical interchangeable media and transmission channels, thus permitting interchange of data to take place either indirectly by means of an intermediate recording on a physical medium, or by local connection of various units (such as input and output devices and computers) or by means of data transmission equipment.

The implementation of this coded character set in physical media and for transmission, taking into account the need for error checking, may be the subject of other International Standards.

## 3 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard 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/IEC 2022:1994, *Information technology - Character code structure and extension techniques*

ISO 2375:1985, *Data processing - Procedure for registration of escape sequences*

ISO/IEC 7350:1991, *Information technology - Registration of repertoires of graphic characters from ISO/IEC 10367*

ISO/IEC 10367:1991, *Information technology - Standardized coded graphic character sets for use in 8-bit codes*

ISO/IEC 10538:1991, *Information technology - Control functions for text communication*

ISO/IEC 10646-1:2000, *Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane*  
<http://www.iso.org/iso/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001>



## 4 Terms and definitions

For the purposes of this International Standard, the following terms and definitions apply:

### 4.1

#### **active position**

the character position which is to image the graphic symbol representing the next graphic character or relative to which the next control function is to be executed

### 4.2

#### **bit combination**

an ordered set of bits used for the representation of characters

### 4.3

#### **character**

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

### 4.4

#### **character position**

the portion of a display that is imaging or is capable of imaging a graphic symbol

### 4.5

#### **coded-character-data-element (CC-data-element)**

an element of interchanged information that is specified to consist of a sequence of coded representations of characters, in accordance with one or more identified standards for coded character sets

NOTE 1: In a communication environment in accordance with the Reference Model for Open Systems Interconnection of ISO 7498, a CC-data-element will form all or part of the information that corresponds to the Presentation-Protocol-Data-Unit (PPDU) defined in that International Standard.

<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-116849c7d1d5/iso-iec-6937-2001>

NOTE 2: When information interchange is accomplished by means of interchangeable media, a CC-data-element will form all or part of the information that corresponds to the user data, and not that recorded during formatting and initialization.

### 4.6

#### **coded character set; code**

a set of unambiguous rules that establishes a character set and the one-to-one relationship between the characters of the set and their bit combinations

### 4.7

#### **code extension**

the techniques for the encoding of characters that are not included in the character set of a given code

### 4.8

#### **code table**

a table showing the characters allocated to each bit combination in a code

### 4.9

#### **control character**

a control function the coded representation of which consists of a single bit combination

### 4.10

#### **control function**

an element of a character set that affects the recording, processing, transmission or interpretation of data, and that has a coded representation consisting of one or more bit combinations

**4.11 device:** A component of information processing equipment which can transmit, and/or receive, coded information within CC-data-elements

NOTE: It may be an input/output device in the conventional sense, or a process such as an application program or gateway function.

**4.12  
escape sequence**

a string of bit combinations that are used for control purposes in code extension procedures. The first of these bit combinations represents the control function ESCAPE

NOTE: Formats and rules regarding the use of escape sequences are specified in ISO/IEC 2022.

**4.13  
graphic character**

a character, other than a control function, that has a visual representation normally handwritten, printed or displayed, and that has a coded representation consisting of one or more bit combinations

**4.14  
graphic symbol**

a visual representation of a graphic character or of a control function

**4.15  
repertoire**

a specified set of characters that are represented by one or more bit combinations of a coded character set

**4.16  
text**

a representation of information for human comprehension that is intended for presentation in a two-dimensional form, for example printed on paper or displayed on a screen.

iTeh STANDARD PREVIEW

(standards.iteh.ai)

Text consists of symbols, phrases or sentences in natural or artificial languages, pictures, diagrams and tables

<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1565b7c7d1d/iso-iec-6937-2001>

NOTE: This International Standard applies only to text made up of characters.

**4.17  
text communication; communication of text**

the transfer of text by means of telecommunications

NOTE: In the context of this International Standard, text communication is by means of binary-coded representations of characters.

**4.18  
user**

a person or other entity that invokes the services provided by a device

NOTE 1: This entity may be a process such as an application program if the "device" is a code convertor or a gateway function, for example.

NOTE 2: The characters, as supplied by the user or made available to the user, may be in the form of codes local to the device, or of non-conventional visible representations, provided that 2.1.2 above is satisfied.

## 5 Notation, code table and names

### 5.1 Notation

The bits of the bit combinations of the 8-bit code are identified by  $b_8$ ,  $b_7$ ,  $b_6$ ,  $b_5$ ,  $b_4$ ,  $b_3$ ,  $b_2$  and  $b_1$ , where  $b_8$  is the highest-order, or most significant bit and  $b_1$  is the lowest-order, or least significant bit.

The bit combinations may be interpreted to represent numbers in the range 0 to 255 in binary notation by attributing the following weights to the individual bits:

Bit	$b_8$	$b_7$	$b_6$	$b_5$	$b_4$	$b_3$	$b_2$	$b_1$
Weight	128	64	32	16	8	4	2	1

In this International Standard, the bit combinations are identified by notations of the form  $xx/yy$ , where  $xx$  and  $yy$  are numbers in the range 00 to 15. The correspondence between the notations of the form  $xx/yy$  and the bit combinations consisting of the bits  $b_8$  to  $b_1$ , is as follows:

- $xx$  is the number represented by  $b_8$ ,  $b_7$ ,  $b_6$  and  $b_5$  where these bits are given the weights 8, 4, 2 and 1, respectively;
- $yy$  is the number represented by  $b_4$ ,  $b_3$ ,  $b_2$  and  $b_1$  where these bits are given the weights 8, 4, 2 and 1, respectively.

The notations of the form  $xx/yy$  are the same as the ones used to identify code table positions, where  $xx$  is the column number and  $yy$  is the row number (see 5.2).

### 5.2 Code table

An 8-bit code table consists of 256 positions arranged in 16 columns and 16 rows. The columns and rows are numbered 00 to 15.

The code table positions are identified by notations of the form  $xx/yy$ , where  $xx$  is the column number and  $yy$  is the row number.

The positions of the code table are in one-to-one correspondence with the bit combinations of the code. The notation of a code table position, of the form  $xx/yy$ , is the same as that of the corresponding bit combination.

### 5.3 Names

This International Standard assigns one name to each character. In addition, it specifies an acronym for the three characters SPACE, NO-BREAK SPACE and SOFT HYPHEN and a graphic symbol for the other graphic characters. By convention, only capital letters, space and hyphen are used for writing the names of characters. It is intended that the acronym and this convention be retained in all translations of the text of this International Standard.

The names chosen to denote graphic characters are intended to reflect their customary meaning. However, this International Standard does not define and does not restrict the meanings of graphic characters. Neither does it specify a particular style or font design for imaging the graphic characters.

The character names are aligned with those of ISO/IEC 10646-1.

## 6 Specifications of SPACE, NO-BREAK SPACE and SOFT HYPHEN

**6.1 SPACE (SP):** A graphic character that has a visual representation consisting of the absence of a graphic symbol. Its coded representation is 02/00.

**6.2 NO-BREAK SPACE (NBSP):** A graphic character, the visual representation of which consists of the absence of a graphic symbol, for use when a line break is to be prevented in the text as presented.

**6.3 SOFT HYPHEN (SHY):** A graphic character that is imaged by a graphic symbol identical with, or similar to, that representing HYPHEN-MINUS, for use when a line break has been established within a word.

## 7 Composition of the character repertoire

The repertoire of the graphic characters defined in this International Standard consists of

a) SPACE (SP)

and of 332 characters as follows

b) Latin alphabetic characters comprising

1) the 52 capital and small letters of the basic Latin alphabet,

2) accented letters, the graphic representations of which consist of combinations of basic Latin letters with diacritical marks,

3) special alphabetic characters which are neither basic Latin letters nor combinations of basic Latin letters with diacritical marks;

c) non-alphabetic characters, such as digits, fractions, punctuation and diacritical marks, monetary symbols etc.

The repertoire, excluding SPACE, is specified in Table 4. In each table entry, the first column specifies the name of the character. The second column specifies its coded representation (see 8.3).

NOTE 1: A survey of the use of Latin characters in various languages is included in Annex D.

NOTE 2: Use of the following characters: LATIN CAPITAL LETTER L WITH MIDDLE DOT, LATIN SMALL LETTER L WITH MIDDLE DOT and LATIN SMALL LETTER N PRECEDED BY APOSTROPHE, is deprecated, and they should better be encoded as 'l' / 'L' followed by MIDDLE DOT, and APOSTROPHE followed by 'n', respectively.

## 8 Specification of the coded character set

### 8.1 Character sets

The coded representations of the graphic characters of the repertoire defined in this International Standard make use of the character SPACE and of two character sets, that is "a primary set" and a "supplementary set".

The primary set shall consist of the graphic characters of the basic G0 set identified by international registration number 6, represented by bit combinations 02/01 to 07/14. The characters of the primary set shall not be used in combination with each other to generate graphic characters of the repertoire defined in this International Standard. The primary set contains the letters of the basic Latin alphabet, some spacing diacritical marks and a number of non-alphabetic characters.

The supplementary set contains the graphic characters of the G1 set identified by international register number 156, represented by bit combinations 10/00 to 11/15 and 13/00 to 15/15, and non-spacing diacritical marks, represented by bit combinations 12/00 to 12/15. The graphic characters consist of a number of characters used in addition to those in the primary set.

A non-spacing diacritical mark shall be used only in combination with certain basic Latin letters, or with SPACE.

The allowed combinations of non-spacing diacritical marks and letters are the ones needed to represent the accented letters included in Table 4. This set of combinations is summarized in Annex C.

The code table for the primary and the supplementary sets of graphic characters is given in Table 1. Shaded positions denote bit combinations which are reserved as specified in 8.2.

The names of the characters in the primary set are specified in Table 2.

The names of the characters and non-spacing diacritical marks of the supplementary set are specified in Table 3. In order to stress that non-spacing diacritical marks are not characters, the names given to them are printed in lower case italics.

NOTE: The shaded positions 00/00 to 01/15 and 07/15 to 09/15 are outside the scope of this International Standard.

## 8.2 Explanations concerning the code table

**8.2.1** Bit combinations 10/04 and 10/06 are reserved for future standardization, and shall not be used.

**8.2.2** The non-spacing diacritical marks of column 12 are used only in combination with certain basic Latin letters, or with SPACE (see Annex C). The graphic symbols shown in column 12 represent diacritical marks as separate graphic characters.

**8.2.3** Bit combinations 12/00, 12/09 and 12/12 are reserved for possible allocation of additional diacritical marks, and shall not be used.

**8.2.4** Bit combinations 13/08 to 13/11 and 14/05 are reserved for future standardization, and shall not be used.

## 8.3 Coded representations of the graphic characters of the repertoire

The coded representations of the graphic characters of the repertoire defined in this International Standard are specified in Table 4. The formats of the coded representations are as follows:

a) Accented letters

<https://standards.iteh.ai/catalog/standards/sist/41a7f2aa-97b6-4a4e-b441-1363fb9c7d1d/iso-iec-6937-2001>

Each accented letter is represented by a sequence of bit combinations consisting of the coded representation of the relevant non-spacing diacritical mark (an element of the supplementary set), followed by the coded representation of the relevant basic Latin letter (an element of the primary set).

b) Diacritical marks as separate graphic characters

The diacritical marks that are elements of the primary set (GRAVE ACCENT, CIRCUMFLEX ACCENT and TILDE) are represented as separate graphic characters by the corresponding single bit combination in the range 02/01 to 07/14.

The other ten of the diacritical marks of column 12 are represented as separate graphic characters by a sequence of bit combinations consisting of the coded representation of the relevant non-spacing diacritical mark (an element of the supplementary set), followed by the coded representation of the character SPACE, i.e. the bit combination 02/00.

c) All other graphic characters of the repertoire

Any graphic character of the repertoire, other than an accented letter or a diacritical mark as a separate graphic character that is not an element of the primary set, is an element of either the primary set or the supplementary set and is represented by the corresponding single bit combination in the range 02/01 to 07/14 or 10/00 to 15/15.

Depending of the code extension techniques used, a bit combination, representing an element of either the primary or the supplementary set may have to be preceded by a code extension function invoking the character set concerned.