
**Information technology — Universal
Multiple-Octet Coded Character Set
(UCS) —**

**Part 1:
Architecture and Basic Multilingual Plane**

*Technologies de l'information — Jeu universel de caractères codés sur
plusieurs octets (JUC) —*

Partie 1: Architecture et plan multilingue de base

Reference number
ISO/IEC 10646-1:2000(E)



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.

© ISO/IEC 2000

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
Web www.iso.ch

Printed in Switzerland

Contents

	Page
1 Scope	1
2 Conformance	1
3 Normative references	2
4 Definitions	2
5 General structure of the UCS	4
6 Basic structure and nomenclature	4
7 General requirements for the UCS	8
8 The Basic Multilingual plane	8
9 Other planes	8
10 Private use groups, planes, and zones	8
11 Revision and updating of the UCS	9
12 Subsets	9
13 Coded representation forms of the UCS	9
14 Implementation levels	9
15 Use of control functions with the UCS	10
16 Declaration of identification of features	10
17 Structure of the code tables and lists	11
18 Block names	12
19 Characters in bi-directional context	12
20 Special characters	12
21 Presentation forms of characters	13
22 Compatibility characters	13
23 Order of characters	13
24 Combining characters	13
25 Special features of individual scripts	14
26 Code tables and lists of character names	15
27 CJK unified ideographs	304

Annexes

A Collections of graphic characters for subsets	879
B List of combining characters	885
C Transformation format for 16 planes of Group 00 (UTF-16)	890
D UCS Transformation Format 8 (UTF-8)	893
E Mirrored characters in Arabic bi-directional context	897
F Alternate format characters	899

G	Alphabetically sorted list of character names	904
H	The use of “signatures” to identify UCS	951
J	Recommendation for combined receiving/originating devices with internal storage	952
K	Notations of octet value representations	953
L	Character naming guidelines	954
M	Sources of characters	956
N	External references to character repertoires	959
P	Additional information on characters	961
Q	Code mapping table for Hangul syllables	964
R	Names of Hangul syllables	974
S	Procedure for the unification and arrangement of CJK ideographs	985

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 part of ISO/IEC 10646 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 10646-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 2, *Coded character sets*.

This second edition cancels and replaces the first edition (ISO/IEC 10646-1:1993), which has been technically revised. It also incorporates Amendments 1 to 13, 16 to 21 and 23, and Technical Corrigenda 1 and 2 to the first edition.

ISO/IEC 10646 consists of the following parts, under the general title *Information technology — Universal Multiple-Octet Coded Character Set (UCS)*:

- *Part 1: Architecture and Basic Multilingual Plane*
- *Part 2: Secondary Multilingual Plane for scripts and symbols, Supplementary Plane for CJK Ideographs, Special Purpose Plane*

Additional parts will specify other planes.

Annexes A to D form a normative part of this part of ISO/IEC 10646. Annexes E to S are for information only.

Introduction

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages (scripts) of the world as well as additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture and the Basic Multilingual Plane (BMP) of the UCS.

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

Part 1: Architecture and Basic Multilingual Plane

1 Scope

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input, and presentation of the written form of the languages of the world as well as of additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture, and

- defines terms used in ISO/IEC 10646;
- describes the general structure of the coded character set;
- specifies the Basic Multilingual Plane (BMP) of the UCS, and defines a set of graphic characters used in scripts and the written form of languages on a world-wide scale;
- specifies the names for the graphic characters of the BMP, and their coded representations;
- specifies the four-octet (32-bit) canonical form of the UCS: UCS-4;
- specifies a two-octet (16-bit) BMP form of the UCS: UCS-2;
- specifies the coded representations for control functions;
- specifies the management of future additions to this coded character set.

The UCS is a coding system different from that specified in ISO/IEC 2022. The method to designate UCS from ISO/IEC 2022 is specified in 16.2.

NOTE 1 - The Unicode Standard, Version 3.0, provides a set of characters, names, and coded representations that are identical with those in this Part 1 of this International Standard. It additionally provides details of character properties, processing algorithms, and definitions that are useful to implementors.

NOTE 2 - It is intended that character code positions for additional scripts and symbols will be allocated in this Part 1 of this International Standard when sufficient input and review

is provided by national standards organizations or other qualified experts.

2 Conformance

2.1 General

Whenever private use characters are used as specified in ISO/IEC 10646, the characters themselves shall not be covered by these conformance requirements.

2.2 Conformance of information interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with ISO/IEC 10646 if

- a) all the coded representations of graphic characters within that CC-data-element conform to clauses 6 and 7, to an identified form chosen from clause 13 or annex C or annex D, and to an identified implementation level chosen from clause 14;
- b) all the graphic characters represented within that CC-data-element are taken from those within an identified subset (clause 12);
- c) all the coded representations of control functions within that CC-data-element conform to clause 15.

A claim of conformance shall identify the adopted form, the adopted implementation level and the adopted subset by means of a list of collections and/or characters.

2.3 Conformance of devices

A device is in conformance with ISO/IEC 10646 if it conforms to the requirements of item a) below, and either or both of items b) and c).

NOTE - The term device is defined (in 4.18) as a component of information processing equipment which can transmit and/or receive coded information within CC-data-elements. A device may be a conventional input/output device, or a process such as an application program or gateway function.

A claim of conformance shall identify the document that contains the description specified in a) below, and shall identify the adopted form(s), the adopted

implementation level, the adopted subset (by means of a list of collections and/or characters), and the selection of control functions adopted in accordance with clause 15.

a) Device description: A device that conforms to ISO/IEC 10646 shall be the subject of a description that identifies the means by which the user may supply characters to the device and/or may recognize them when they are made available to the user, as specified respectively, in subclauses b), and c) below.

b) Originating device: An originating device shall allow its user to supply any characters from an adopted subset, and be capable of transmitting their coded representations within a CC-data-element in accordance with the adopted form and implementation level.

c) Receiving device: A receiving device shall be capable of receiving and interpreting any coded representation of characters that are within a CC-data-element in accordance with the adopted form and implementation level, and shall make any corresponding characters from the adopted subset available to the user in such a way that the user can identify them.

Any corresponding characters that are not within the adopted subset shall be indicated to the user. The way used for indicating them need not distinguish them from each other.

NOTE 1 - An indication to the user may consist of making available the same character to represent all characters not in the adopted subset, or providing a distinctive audible or visible signal when appropriate to the type of user.

NOTE 2 - See also annex J for receiving devices with re-transmission capability.

3 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 10646. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this part of ISO/IEC 10646 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/IEC 6429:1992, *Information technology — Control functions for coded character sets*.

4 Terms and definitions

For the purposes of this part of ISO/IEC 10646, the following terms and definitions apply:

4.1 Basic Multilingual Plane (BMP): Plane 00 of Group 00.

4.2 block: A contiguous range of code positions to which a set of characters that share common characteristics, such as script, are allocated. A block does not overlap another block. One or more of the code positions within a block may have no character allocated to it.

4.3 canonical form: The form with which characters of this coded character set are specified using four octets to represent each character.

4.4 CC-data-element (coded-character-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.

4.5 cell: The place within a row at which an individual character may be allocated.

4.6 character: A member of a set of elements used for the organization, control, or representation of data.

4.7 character boundary: Within a stream of octets the demarcation between the last octet of the coded representation of a character and the first octet of that of the next coded character.

4.8 coded character: A character together with its coded representation.

4.9 coded character set: A set of unambiguous rules that establishes a character set and the relationship between the characters of the set and their coded representation.

4.10 code table: A table showing the characters allocated to the octets in a code.

4.11 collection: A set of coded characters which is numbered and named and which consists of those coded characters whose code positions lie within one or more identified ranges.

NOTE - If any of the identified ranges include code positions to which no character is allocated, the repertoire of the collection will change if an additional character is assigned to any of those positions at a future amendment of this International Standard. However it is intended that the collection number and name will remain unchanged in future editions of this International Standard.

4.12 combining character: A member of an identified subset of the coded character set of ISO/IEC 10646 intended for combination with the preceding non-combining graphic character, or with a sequence of combining characters preceded by a non-combining character (see also 4.14).

NOTE - This part of ISO/IEC 10646 specifies several subset collections which include combining characters.

4.13 compatibility character: A graphic character included as a coded character of ISO/IEC 10646 primarily for compatibility with existing coded character sets.

4.14 composite sequence: A sequence of graphic characters consisting of a non-combining character followed by one or more combining characters (see also 4.12).

NOTE 1 - A graphic symbol for a composite sequence generally consists of the combination of the graphic symbols of each character in the sequence.

NOTE 2 - A composite sequence is not a character and therefore is not a member of the repertoire of ISO/IEC 10646.

4.15 control function: An action that affects the recording, processing, transmission, or interpretation of data, and that has a coded representation consisting of one or more octets.

4.16 default state: The state that is assumed when no state has been explicitly specified.

4.17 detailed code table: A code table showing the individual characters, and normally showing a partial row.

4.18 device: A component of information processing equipment which can transmit and/or receive coded information within CC-data-elements. (It may be an input/output device in the conventional sense, or a process such as an application program or gateway function.)

4.19 fixed collection: A collection in which every code position within the identified range(s) has a character allocated to it, and which is intended to remain unchanged in future editions of this International Standard.

4.20 graphic character: A character, other than a control function, that has a visual representation normally handwritten, printed, or displayed.

4.21 graphic symbol: The visual representation of a graphic character or of a composite sequence.

4.22 group: A subdivision of the coding space of this coded character set; of 256 x 256 x 256 cells.

4.23 high-half zone: a set of cells reserved for use in UTF-16 (see annex C); an RC-element corresponding to any of these cells may be used in UTF-

16 as the first of a pair of RC-elements which represents a character from a plane other than the BMP.

4.24 interchange: The transfer of character coded data from one user to another, using telecommunication means or interchangeable media.

4.25 interworking: The process of permitting two or more systems, each employing different coded character sets, meaningfully to interchange character coded data; conversion between the two codes may be involved.

4.26 low-half zone: a set of cells reserved for use in UTF-16 (see annex C); an RC-element corresponding to any of these cells may be used in UTF-16 as the second of a pair of RC-elements which represents a character from a plane other than the BMP.

4.27 octet: An ordered sequence of eight bits considered as a unit.

4.28 plane: A subdivision of a group; of 256 x 256 cells

4.29 presentation; to present: The process of writing, printing, or displaying a graphic symbol.

4.30 presentation form: In the presentation of some scripts, a form of a graphic symbol representing a character that depends on the position of the character relative to other characters.

4.31 private use plane: A plane within this coded character set the contents of which is not specified in ISO/IEC 10646 (see clause 10)

4.32 RC-element: a two-octet sequence comprising the R-octet and the C-octet (see 6.2) from the four octet sequence (in the canonical form) that corresponds to a cell in the coding space of this coded character set.

4.33 repertoire: A specified set of characters that are represented in a coded character set.

4.34 row: A subdivision of a plane; of 256 cells.

4.35 script: A set of graphic characters used for the written form of one or more languages.

4.36 supplementary plane: A plane that accommodates characters which have not been allocated to the Basic Multilingual Plane.

4.37 unpaired RC-element: An RC-element in a CC-data element that is either:

- an RC-element from the high-half zone that is not immediately followed by an RC-element from the low-half zone, or
- an RC-element from the low-half zone that is not immediately preceded by a high-half RC-element from the high-half zone.

4.38 user: A person or other entity that invokes the service provided by a device. (This entity may be a process such as an application program if the “device” is a code converter or a gateway function, for example.)

4.39 zone: A sequence of cells of a code table, comprising one or more rows, either in whole or in part, containing characters of a particular class (for example see clause 8).

5 General structure of the UCS

The general structure of the Universal Multiple-Octet Coded Character Set (referred to hereafter as “this coded character set”) is described in this explanatory clause, and is illustrated in figures 1 and 2. The normative specification of the structure is given in the following clauses.

The value of any octet is expressed in hexadecimal notation from 00 to FF in ISO/IEC 10646 (see annex K).

The canonical form of this coded character set – the way in which it is to be conceived – uses a four-dimensional coding space, regarded as a single entity, consisting of 128 three-dimensional groups.

NOTE - Thus, bit 8 of the most significant octet in the canonical form of a coded character can be used for internal processing purposes within a device as long as it is set to zero within a conforming CC-data-element.

Each group consists of 256 two-dimensional planes. Each plane consists of 256 one-dimensional rows, each row containing 256 cells. A character is located and coded at a cell within this coding space or the cell is declared unused.

In the canonical form, four octets are used to represent each character, and they specify the group, plane, row and cell, respectively. The canonical form consists of four octets since two octets are not sufficient to cover all the characters in the world, and a 32-bit representation follows modern processor architectures.

The four-octet canonical form can be used as a four-octet coded character set, in which case it is called UCS-4.

The first plane (Plane 00 of Group 00) is called the Basic Multilingual Plane. The Basic Multilingual Plane includes characters in general use in alphabetic, syllabic, and ideographic scripts together with various symbols and digits.

The subsequent planes are regarded as supplementary or private use planes, which will accommodate additional graphic characters (see clause 9).

The planes that are reserved for private use are specified in clause 10. The contents of the cells in private use zones are not specified in ISO/IEC 10646.

Each character is located within the coded character set in terms of its Group-octet, Plane-octet, Row-octet, and Cell-octet.

In addition to the canonical form, a two-octet BMP form is specified. Thus, the Basic Multilingual Plane can be used as a two-octet coded character set identified as UCS-2.

Subsets of the coding space may be used in order to give a sub-repertoire of graphic characters.

A UCS Transformation Format (UTF-16) is specified in annex C which can be used to represent characters from 16 planes of group 00, additional to the BMP, in a form that is compatible with the two-octet BMP form.

Another UCS Transformation Format (UTF-8) is specified in annex D which can be used to transmit text data through communication systems which are sensitive to octet values for control characters coded according to the 8-bit structure of ISO/IEC 2022, and to ISO/IEC 4873. UTF-8 also avoids the use of octet values according to ISO/IEC 4873 which have special significance during the parsing of file-name character strings in widely-used file-handling systems.

6 Basic structure and nomenclature

6.1 Structure

The Universal Multiple-Octet Coded Character Set as specified in ISO/IEC 10646 shall be regarded as a single entity.

This entire coded character set shall be conceived of as comprising 128 groups of 256 planes. Each plane shall be regarded as containing 256 rows of characters, each row containing 256 cells. In a code table representing the contents of a plane (such as in figure 2), the horizontal axis shall represent the least significant octet, with its smaller value to the left; and the vertical axis shall represent the more significant octet, with its smaller value at the top.

Each axis of the coding space shall be coded by one octet. Within each octet the most significant bit shall be bit 8 and the least significant bit shall be bit 1. Accordingly, the weight allocated to each bit shall be

bit 8	bit 7	bit 6	bit 5	bit 4	bit 3	bit 2	bit 1
128	64	32	16	8	4	2	1

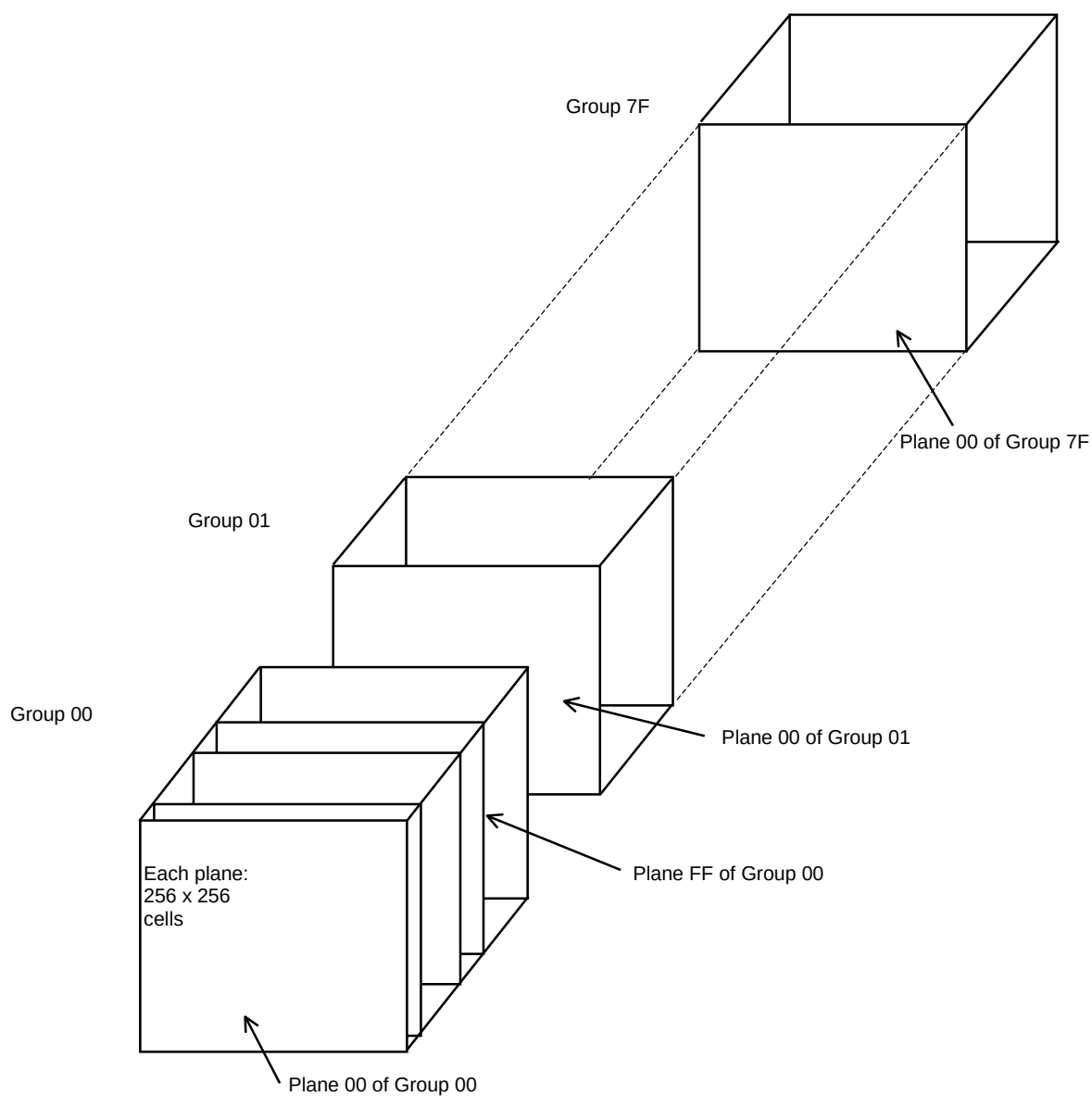
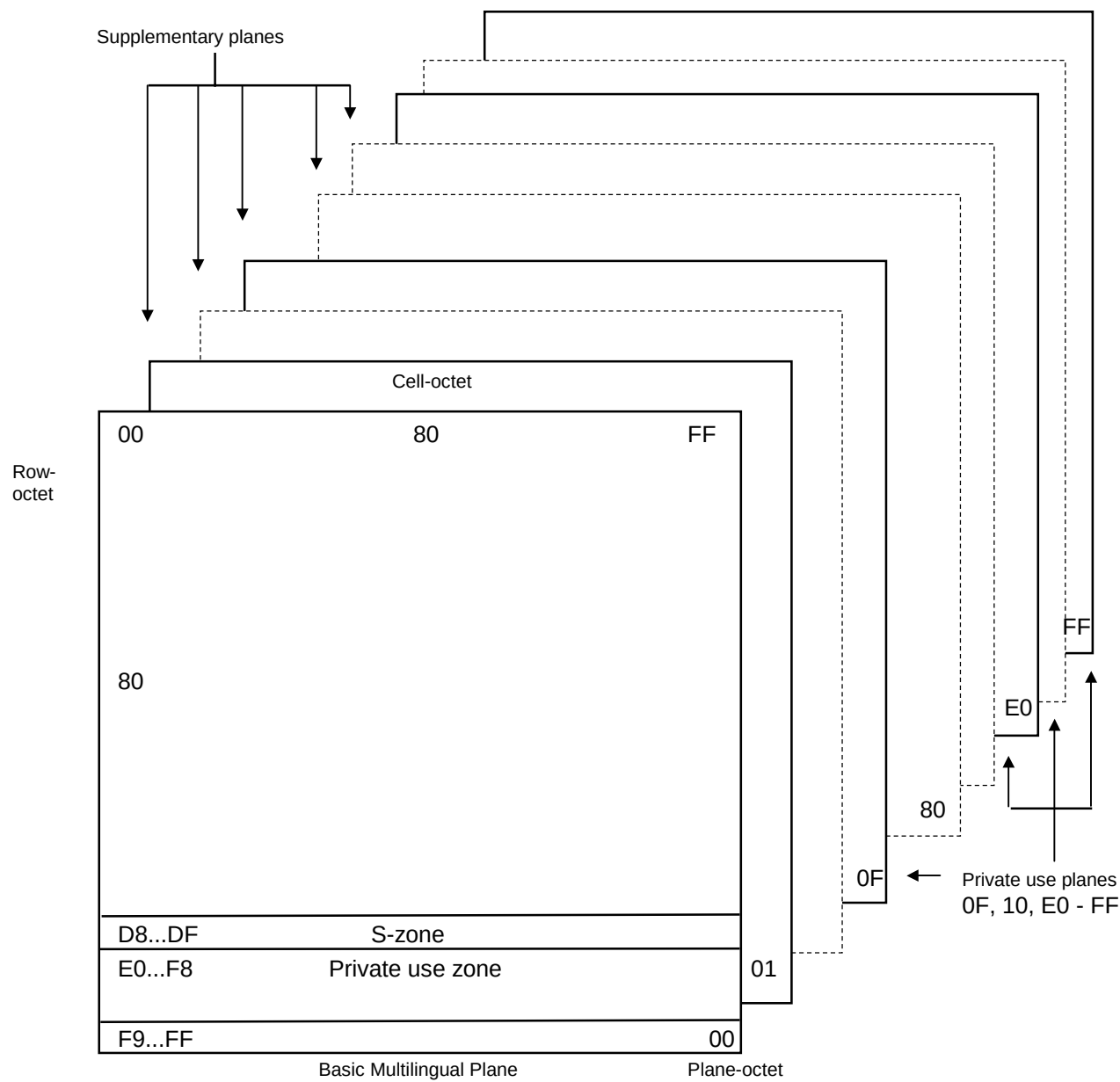


Figure 1 - Entire coding space of the Universal Multiple-Octet Coded Character Set



NOTE - Labels “S-zone” and “Private use zone” are specified in clause 8.

Figure 2 - Group 00 of the Universal Multiple-Octet Coded Character Set

6.2 Coding of characters

In the canonical form of the coded character set, each character within the entire coded character set shall be represented by a sequence of four octets. The most significant octet of this sequence shall be the group-octet. The least significant octet of this sequence shall be the cell-octet. Thus this sequence may be represented as

m.s.		l.s.	
Group-octet	Plane-octet	Row-octet	Cell-octet

where m.s. means the most significant octet, and l.s. means the least significant octet.

For brevity, the octets may be termed

m.s.		l.s.	
G-octet	P-octet	R-octet	C-octet

Where appropriate, these may be further abbreviated to G, P, R, and C.

The value of any octet shall be represented by two hexadecimal digits, for example: 31 or FE. When a single character is to be identified in terms of the values of its group, plane, row, and cell, this shall be represented such as:

0000 0030 for DIGIT ZERO

0000 0041 for LATIN CAPITAL LETTER A

When referring to characters within an identified plane, the leading four digits (for G-octet and P-octet) may be omitted. For example, within plane 00, 0030 may be used to refer to DIGIT ZERO.

6.3 Octet order

The sequence of the octets that represent a character, and the most significant and least significant ends of it, shall be maintained as shown above. When serialized as octets, a more significant octet shall precede less significant octets. When not serialized as octets, the order of octets may be specified by agreement between sender and recipient (see 16.1 and annex H).

6.4 Naming of characters

ISO/IEC 10646 assigns a unique name to each character. The name of a character either:

- a. denotes the customary meaning of the character, or
- b. describes the shape of the corresponding graphic symbol, or
- c. follows the rule given in clause 27 for Chinese/Japanese/Korean (CJK) unified ideographs.

Guidelines to be used for constructing the names of characters in cases a. and b. are given in annex L.

6.5 Short identifiers for characters

ISO/IEC 10646 defines a short identifier for each character. The short identifier for any character is distinct from the short identifier for any other character.

NOTE - These short identifiers are independent of the language in which this standard is written, and are thus retained in all translations of the text.

The following alternative forms of notation of a short identifier are defined here.

- a. The eight-digit form of short identifier shall consist of the sequence of eight hexadecimal digits that represents the code position of the character (see 6.2).
- b. The four-digit form of short identifier shall consist of the last four digits of the eight-digit form. It is not defined if the first four digits of the eight-digit form are not all zeroes; that is, for characters allocated outside the Basic Multilingual Plane.
- c. The character “-” (HYPHEN-MINUS) may, as an option, precede the 8-digit form of short identifier.
- d. The character “+” (PLUS SIGN) may, as an option, precede the 4-digit form of short identifier.
- e. The prefix letter “U” (LATIN CAPITAL LETTER U) may, as an option, precede any of the four forms of short identifier defined in a. to d. above.

The capital letters A to F, and U that appear within short identifiers may be replaced by the corresponding small letters.

The full syntax of the notation of a short identifier, in Backus-Naur form, is:

$$\{ U \mid u \} [\{ + \}xxxx \mid \{ - \}xxxxxxxxxx]$$

where “x” represents one hexadecimal digit (0 to 9, A to F, or a to f), for example:

-hhhhhhhhh +kkkk

Uhhhhhhhhh U+kkkk

where hhhhhhhh indicates the eight-digit form and kkkk indicates the four-digit form.

NOTE 1 - As an example the short identifier for LATIN SMALL LETTER LONG S (see tables for Row 01 in clause 26) may be notated in any of the following forms:

0000017F	-0000017F	U0000017F	U-0000017F
017F	+017F	U017F	U+017F

Any of the capital letters may be replaced by the corresponding small letter.

NOTE 2 - Two special prefixed forms of notation have also been used, in which the letter T (LATIN CAPITAL LETTER T or LATIN SMALL LETTER T) replaces the

letter U in the corresponding prefixed forms. The forms of notation that included the prefix letter T indicated that the short identifier refers to a character in ISO/IEC 10646-1 First Edition (before the application of any Amendments), whereas the forms of notation that include the prefix letter U always indicate that the short identifier refers to a character in ISO/IEC 10646 at the most recent state of amendment. Corresponding short identifiers of the form T-xxxxxxx and U-xxxxxxx refer to the same character except when xxxxxxxx lies in the range 00003400 to 00004DFF inclusive. Forms of notation that include no prefix letter always indicate a reference to the most recent state of amendment of ISO/IEC 10646, unless otherwise qualified.

7 General requirements for the UCS

The following requirements apply to the entire coded character set.

- a) The values of P-, and R-, and C-octets used for representing graphic characters shall be in the range 00 to FF. The values of G-octets used for representation of graphic characters shall be in the range 00 to 7F. On any plane, code positions FFFE and FFFF shall not be used.

NOTE - Code position FFFE is reserved for "signature" (see annex H). Code position FFFF can be used for internal processing uses requiring a numeric value that is guaranteed not to be a coded character, such as in terminating tables, or signaling end-of-text. Since it is the largest two-octet value, it may also be used as the final value in binary or sequential searching index.

- b) Code positions to which a character is not allocated, except for the positions reserved for private use characters or for transformation formats, are reserved for future standardization and shall not be used for any other purpose. Future editions of ISO/IEC 10646 will not allocate any characters to code positions reserved for private use characters or for transformation formats.
- c) The same graphic character shall not be allocated to more than one code position. There are graphic characters with similar shapes in the coded character set; they are used for different purposes and have different character names.

8 The Basic Multilingual Plane

Plane 00 of Group 00 shall be the Basic Multilingual Plane (BMP). The BMP can be used as a two-octet coded character set in which case it shall be called UCS-2 (see 13.1).

Code positions 0000 0000 to 0000 001F in the BMP are reserved for control characters, and code position 0000 007F is reserved for the character

DELETE (see clause 15). Code positions 0000 0080 to 0000 009F are reserved for control characters.

Code positions 0000 D800 to 0000 DFFF are reserved for the use of UTF-16 (see annex C). These positions are known as the S-zone.

Code positions 0000 E000 to 0000 F8FF are reserved for private use (see clause 10). These positions are known as the private use zone.

Code positions 0000 FFFE and 0000 FFFF are reserved.

9 Other planes

9.1 Planes reserved for future standardization

Planes 11 to DF in Group 00 and Planes 00 to FF in Groups 01 to 5F are reserved for future standardization, and thus those code positions shall not be used for any other purpose.

9.2 Planes accessible by UTF-16

Each code position in Planes 01 to 10 of Group 00 has a unique mapping to a four-octet sequence in accordance with the UTF-16 form of coded representation (see annex C). This form is compatible with the two-octet BMP form of UCS-2 (see 13.1).

Code positions in Planes 11 to FF of Group 00, or in Planes 00 to FF of other groups, do not have a mapping to the UTF-16 form.

10 Private use groups, planes, and zones

10.1 Private use characters

Private use characters are not restrained in any way by ISO/IEC 10646. Private use characters can be used to provide user-defined characters. For example, this is a common requirement for users of ideographic scripts.

NOTE 1 - For meaningful interchange of private use characters, an agreement, independent of ISO/IEC 10646, is necessary between sender and recipient.

Private use characters can be used for dynamically-redefinable character applications.

NOTE 2 - For meaningful interchange of dynamically-redefinable characters, an agreement, independent of ISO/IEC 10646 is necessary between sender and recipient. ISO/IEC 10646 does not specify the techniques for defining or setting up dynamically-redefinable characters.

10.2 Code positions for private use characters

The code positions of the 32 groups from Group 60 to Group 7F shall be for private use.

The code positions of Plane 0F and Plane 10, and of the 32 planes from Plane E0 to Plane FF, of Group 00 shall be for private use.

The 6400 code positions E000 to F8FF of the Basic Multilingual Plane shall be for private use.

The contents of these code positions are not specified in ISO/IEC 10646 (see 10.1).

11 Revision and updating of the UCS

The revision and updating of this coded character set will be carried out by ISO/IEC JTC1/SC2.

NOTE - It is intended that in future editions of ISO/IEC 10646, the names and allocation of the characters in this edition will remain unchanged.

12 Subsets

ISO/IEC 10646 provides the specification of subsets of coded graphic characters for use in interchange, by originating devices, and by receiving devices.

There are two alternatives for the specification of subsets: limited subset and selected subset. An adopted subset may comprise either of them, or a combination of the two.

12.1 Limited subset

A limited subset consists of a list of graphic characters in the specified subset. This specification allows applications and devices that were developed using other codes to interwork with this coded character set.

A claim of conformance referring to a limited subset shall list the graphic characters in the subset by the names of graphic characters or code positions as defined in ISO/IEC 10646.

12.2 Selected subset

A selected subset consists of a list of collections of graphic characters as defined in ISO/IEC 10646. The collections from which the selection may be made are listed in an annex of each part of ISO/IEC 10646 (see annex A). A selected subset shall always automatically include the Cells 20 to 7E of Row 00 of Plane 00 of Group 00.

A claim of conformance referring to a selected subset shall list the collections chosen as defined in ISO/IEC 10646.

13 Coded representation forms of the UCS

ISO/IEC 10646 provides four alternative forms of coded representation of characters. Two of these forms are specified in this clause, and two others,

UTF-16 and UTF-8, are specified in annexes C and D respectively.

NOTE - The characters from the ISO/IEC 646 IRV repertoire are coded by simple zero extensions to their coded representations in ISO/IEC 646 IRV. Therefore, their coded representations have the same integer values when represented as 8-bit, 16-bit, or 32-bit integers. For implementations sensitive to a zero-valued octet (e.g. for use as a string terminator), use of 8-bit based array data type should be avoided as any zero-valued octet may be interpreted incorrectly. Use of data types at least 16-bits wide is more suitable for UCS-2, and use of data types at least 32-bits wide is more suitable for UCS-4.

13.1 Two-octet BMP form

This coded representation form permits the use of characters from the Basic Multilingual Plane with each character represented by two octets.

Within a CC-data-element conforming to the two-octet BMP form, a character from the Basic Multilingual Plane shall be represented by two octets comprising the R-octet and the C-octet as specified in 6.2 (i.e. its RC-element).

NOTE - A coded graphic character using the two-octet BMP form may be implemented by a 16-bit integer for processing.

13.2 Four-octet canonical form

The canonical form permits the use of all the characters of ISO/IEC 10646, with each character represented by four octets.

Within a CC-data-element conforming to the four-octet canonical form, every character shall be represented by four octets comprising the G-octet, the P-octet, the R-octet, and the C-octet as specified in 6.2.

NOTE - A coded graphic character using the four-octet canonical form may be implemented by a 32-bit integer for processing.

14 Implementation levels

ISO/IEC 10646 specifies three levels of implementation. Combining characters are described in 24 and listed in annex B.

14.1 Implementation level 1

When implementation level 1 is used, a CC-data-element shall not contain coded representations of combining characters (see clause B.1) nor of characters from HANGUL JAMO block (see clause 25). When implementation level 1 is used the unique-spelling rule shall apply (25.2).

14.2 Implementation level 2

When implementation level 2 is used, a CC-data-element shall not contain coded representations of characters listed in clause B.2. When implementation level 2 is used the unique-spelling rule shall apply (25.2).

14.3 Implementation level 3

When implementation level 3 is used, a CC-data-element may contain coded representations of any characters.

15 Use of control functions with the UCS

This coded character set provides for use of control functions encoded according to ISO/IEC 6429 or similarly structured standards for control functions, and standards derived from these. A set or subset of such coded control functions may be used in conjunction with this coded character set. These standards encode a control function as a sequence of one or more octets.

When a control character of ISO/IEC 6429 is used with this coded character set, its coded representation as specified in ISO/IEC 6429 shall be padded to correspond with the number of octets in the adopted form (see clause 13 and annexes C and D). Thus, the least significant octet shall be the bit combination specified in ISO/IEC 6429, and the more significant octet(s) shall be zeros.

For example, the control character FORM FEED is represented by "000C" in the two-octet form, and "0000 000C" in the four-octet form.

For escape sequences, control sequences, and control strings (see ISO/IEC 6429) consisting of a coded control character followed by additional bit combinations in the range 20 to 7F, each bit combination shall be padded by octet(s) with value 00.

For example, the escape sequence "ESC 02/00 04/00" is represented by "001B 0020 0040" in the two-octet form, and "0000 001B 0000 0020 0000 0040" in the four-octet form.

NOTE - The term "character" appears in the definition of many of the control functions specified in ISO/IEC 6429, to identify the elements on which the control functions will act. When such control functions are applied to coded characters according to ISO/IEC 10646 the action of those control functions will depend on the type of element from ISO/IEC 10646 that has been chosen, by the application, to be the element (or character) on which the control functions act. These elements may be chosen to be characters (non-combining characters and/or combining characters) or may be chosen in other ways (such as composite sequences) when applicable.

Code extension control functions for the ISO/IEC 2022 code extension techniques (such as designation escape sequences, single shift, and locking shift) shall not be used with this coded character set.

16 Declaration of identification of features

16.1 Purpose and context of identification

CC-data-elements conforming to ISO/IEC 10646 are intended to form all or part of a composite unit of coded information that is interchanged between an originator and a recipient. The identification of ISO/IEC 10646 (including the form), the implementation level, and any subset of the coding space that have been adopted by the originator must also be available to the recipient. The route by which such identification is communicated to the recipient is outside the scope of ISO/IEC 10646.

However, some standards for interchange of coded information may permit, or require, that the coded representation of the identification applicable to the CC-data-element forms a part of the interchanged information. This clause specifies a coded representation for the identification of UCS with an implementation level and a subset of ISO/IEC 10646, and also of a C0 and a C1 set of control functions from ISO/IEC 6429 for use in conjunction with ISO/IEC 10646. Such coded representations provide all or part of an identification data element, which may be included in information interchange in accordance with the relevant standard.

If two or more of the identifications are present, the order of those identifications shall follow the order as specified in this clause.

NOTE - An alternative method of identification is described in annex N.

16.2 Identification of UCS coded representation form with implementation level

When the escape sequences from ISO/IEC 2022 are used, the identification of a coded representation form of UCS (see clause 13) and an implementation level (see clause 14) specified by ISO/IEC 10646 shall be by a designation sequence chosen from the following list:

- ESC 02/05 02/15 04/00
UCS-2 with implementation level 1
- ESC 02/05 02/15 04/01
UCS-4 with implementation level 1
- ESC 02/05 02/15 04/03
UCS-2 with implementation level 2
- ESC 02/05 02/15 04/04
UCS-4 with implementation level 2
- ESC 02/05 02/15 04/05
UCS-2 with implementation level 3
- ESC 02/05 02/15 04/06
UCS-4 with implementation level 3

or from the lists in C.5 and D.6.

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 2022, it shall consist only of the sequences of bit combinations as shown above.

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 10646, it shall be padded in accordance with clause 15.

16.3 Identification of subsets of graphic characters

When the control sequences of ISO/IEC 6429 are used, the identification of subsets (see clause 12) specified by ISO/IEC 10646 shall be by a control sequence IDENTIFY UNIVERSAL CHARACTER SUBSET (IUCS) as shown below.

CSI Ps... 02/00 06/13

Ps... means that there can be any number of selective parameters. The parameters are to be taken from the subset collection numbers as shown in annex A of each part of ISO/IEC 10646. When there is more than one parameter, each parameter value is separated by an octet with value 03/11.

Parameter values are represented by digits where octet values 03/00 to 03/09 represent digits 0 to 9.

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 2022, it shall consist only of the sequences of bit combinations as shown above.

If such a control sequence appears within a CC-data-element conforming to ISO/IEC 10646, it shall be padded in accordance with clause 15.

16.4 Identification of control function set

When the escape sequences from ISO/IEC 2022 are used, the identification of each set of control functions (see clause 15) of ISO/IEC 6429 to be used in conjunction with ISO/IEC 10646 shall be an identifier sequence of the type shown below.

ESC 02/01 04/00 identifies the full C0 set of ISO/IEC 6429

ESC 02/02 04/03 identifies the full C1 set of ISO/IEC 6429

For a subset of C0 or C1 sets, the final octet F shall be obtained from the International Register of Coded Character Sets. The identifier sequences for these sets shall be:

ESC 02/01 F identifies a C0 set

ESC 02/02 F identifies a C1 set

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 2022, it shall

consist only of the sequences of bit combinations as shown above.

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 10646, it shall be padded in accordance with clause 15.

16.5 Identification of the coding system of ISO/IEC 2022

When the escape sequences from ISO/IEC 2022 are used, the identification of a return, or transfer, from UCS to the coding system of ISO/IEC 2022 shall be by the escape sequence ESC 02/05 04/00. If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 10646, it shall be padded in accordance with clause 15.

If such an escape sequence appears within a CC-data-element conforming to ISO/IEC 2022, it shall consist only of the sequence of bit combinations as shown above.

NOTE - Escape sequence ESC 02/05 04/00 is normally used for return to the restored state of ISO/IEC 2022. The escape sequence ESC 02/05 04/00 specified here is sometimes not exactly as specified in ISO/IEC 2022 due to the presence of padding octets. For this reason the escape sequences in 16.2 for the identification of UCS include the octet 02/15 to indicate that the return does not always conform to that standard.

17 Structure of the code tables and lists

The clauses 26 and 27 set out the detailed code tables and the lists of character names for the graphic characters. Together, these specify graphic characters, their coded representation, and the character name for each character.

The graphic symbols are to be regarded as typical visual representations of the characters. ISO/IEC 10646 does not attempt to prescribe the exact shape of each character. The shape is affected by the design of the font employed, which is outside the scope of ISO/IEC 10646.

Graphic characters specified in ISO/IEC 10646 are uniquely identified by their names. This does not imply that the graphic symbols by which they are commonly imaged are always different. Examples of graphic characters with similar graphic symbols are LATIN CAPITAL LETTER A, GREEK CAPITAL LETTER ALPHA, and CYRILLIC CAPITAL LETTER A.

The meaning attributed to any character is not specified by ISO/IEC 10646; it may differ from country to country, or from one application to another.

For the alphabetic scripts, the general principle has been to arrange the characters within any row in approximate alphabetic sequence; where the script has capital and small letters, these are arranged in

pairs. However, this general principle has been overridden in some cases. For example, for those scripts for which a relevant standard exists, the characters are allocated according to that standard. This arrangement within the code tables will aid conversion between the existing standards and this coded character set. In general, however, it is anticipated that conversion between this coded character set and any other coded character set will use a table lookup technique.

It is not intended, nor will it often be the case, that the characters needed by any one user will be found all grouped together in one part of the code table.

Furthermore, the user of any script will find that needed characters may have been coded elsewhere in this coded character set. This especially applies to the digits, to the symbols, and to the use of Latin letters in dual-script applications.

Therefore, in using this coded character set, the reader is advised to refer first to the block names list in annex A.2 or an overview of the BMP in figures 3 and 4, and then to turn to the specific code table rows for the relevant script and for symbols and digits. In addition, annex G contains an alphabetically sorted list of character names.

18 Block names

Named blocks of contiguous code positions are specified within a plane for the purpose of allocation of characters sharing some common characteristic, such as script. The blocks specified within the BMP are listed in A.2 of annex A, and are illustrated in figures 3 and 4.

19 Characters in bi-directional context

A class of left/right handed pairs of characters have special significance in the context of bi-directional text. In this context the terms LEFT or RIGHT in the character name are also intended to imply “opening” or “closing” forms of character shape, rather than a strict left-hand or right-hand form. These characters are listed below.

<u>Code</u> <u>Position</u>	<u>Name</u>
0028	LEFT PARENTHESIS
0029	RIGHT PARENTHESIS
005B	LEFT SQUARE BRACKET
005D	RIGHT SQUARE BRACKET
007B	LEFT CURLY BRACKET
007D	RIGHT CURLY BRACKET
2045	LEFT SQUARE BRACKET WITH QUILL
2046	RIGHT SQUARE BRACKET WITH QUILL
207D	SUPERSCRIFT LEFT PARENTHESIS
207E	SUPERSCRIFT RIGHT PARENTHESIS
208D	SUBSCRIPT LEFT PARENTHESIS

208E	SUBSCRIPT RIGHT PARENTHESIS
2329	LEFT-POINTING ANGLE BRACKET
232A	RIGHT-POINTING ANGLE BRACKET
3008	LEFT ANGLE BRACKET
3009	RIGHT ANGLE BRACKET
300A	LEFT DOUBLE ANGLE BRACKET
300B	RIGHT DOUBLE ANGLE BRACKET
300C	LEFT CORNER BRACKET
300D	RIGHT CORNER BRACKET
300E	LEFT WHITE CORNER BRACKET
300F	RIGHT WHITE CORNER BRACKET
3010	LEFT BLACK LENTICULAR BRACKET
3011	RIGHT BLACK LENTICULAR BRACKET
3014	LEFT TORTOISE SHELL BRACKET
3015	RIGHT TORTOISE SHELL BRACKET
3016	LEFT WHITE LENTICULAR BRACKET
3017	RIGHT WHITE LENTICULAR BRACKET
3018	LEFT WHITE TORTOISE SHELL BRACKET
3019	RIGHT WHITE TORTOISE SHELL BRACKET
301A	LEFT WHITE SQUARE BRACKET
301B	RIGHT WHITE SQUARE BRACKET

The interpretation and rendering of any of these characters depend on the state related to the symmetric swapping characters (see F.2.2) and on the direction of the character being rendered that are in effect at the point in the CC-data-element where the coded representation of the character appears.

For example, if the character

ACTIVATE SYMMETRIC SWAPPING

occurs and if the direction of the character is from right to left, the character shall be interpreted as if the term LEFT or RIGHT in its name had been replaced by the term RIGHT or LEFT, respectively.

NOTE - In the context of Arabic bi-directional text, certain mathematical symbols may also have special significance (see annex E).

20 Special characters

There are some characters that do not have printable graphic symbols. These characters include space characters. They are

<u>Code</u> <u>Position</u>	<u>Name</u>
0020	SPACE
00A0	NO-BREAK SPACE
2000	EN QUAD
2001	EM QUAD
2002	EN SPACE
2003	EM SPACE
2004	THREE-PER-EM SPACE
2005	FOUR-PER-EM SPACE
2006	SIX-PER-EM SPACE
2007	FIGURE SPACE
2008	PUNCTUATION SPACE
2009	THIN SPACE
200A	HAIR SPACE
3000	IDEOGRAPHIC SPACE

Currency symbols in ISO/IEC 10646 do not necessarily identify the currency of a country. For example,

YEN SIGN can be used for Japanese yen and Chinese yuan. Also, DOLLAR SIGN is used in numerous countries including the United States of America.

There is a special class of characters called Alternate Format Characters which are included for compatibility with some industry practices. These are described in annex F.

21 Presentation forms of characters

Each presentation form of a character provides an alternative form, for use in a particular context, to the nominal form of the character or sequence of characters from the other zones of graphic characters. The transformation from the nominal form to the presentation forms may involve substitution, superimposition, or combination.

The rules for the superimposition, choice of differently shaped characters, or combination into ligatures, or conjuncts, which are often of extreme complexity, are not specified in ISO/IEC 10646.

In general, presentation forms are not intended to be used as a substitute for the nominal forms of the graphic characters specified elsewhere within this coded character set. However, specific applications may encode these presentation forms instead of the nominal forms for specific reasons among which is compatibility with existing devices. The rules for searching, sorting, and other processing operations on presentation forms are outside the scope of ISO/IEC 10646.

Within the BMP these characters are mostly allocated to positions in rows FB to FF.

22 Compatibility characters

Compatibility characters are included in ISO/IEC 10646 primarily for compatibility with existing coded character sets to allow two-way code conversion without loss of information.

Within the BMP many of these characters are allocated to positions within rows F9, FA, FE, and FF, and within rows 31 and 33. Some compatibility characters are also allocated within other rows.

23 Order of characters

Usually, coded characters appear in a CC-data-element in logical order (logical or backing store order corresponds approximately to the order in which characters are entered from the keyboard, after corrections such as insertions, deletions, and overtyping have taken place). This applies even when characters of different dominant direction are

mixed: left-to-right (Greek, Latin, Thai) with right-to-left (Arabic, Hebrew), or with vertical (Mongolian) script.

Some characters may not appear linearly in final rendered text. For example, the medial form of DEVANAGARI VOWEL SIGN I is displayed before the character that it logically follows in the CC-data-element.

24 Combining characters

This clause specifies the use of combining characters. A list of combining characters is shown in clause B.1. A list of combining characters not allowed in implementation level 2 is shown in clause B.2.

NOTE - The names of many script-independent combining characters contain the word "COMBINING".

24.1 Order of combining characters

Coded representations of combining characters shall follow that of the graphic character with which they are associated (for example, coded representations of LATIN SMALL LETTER A followed by

COMBINING TILDE represent a composite sequence for Latin "ã").

If a combining character is to be regarded as a composite sequence in its own right, it shall be coded as a composite sequence by association with the character SPACE. For example, grave accent can be composed as SPACE followed by COMBINING GRAVE ACCENT.

NOTE - Indic matras form a special category of combining characters, since the presentation can depend on more than one of the surrounding characters. Thus it might not be desirable to associate Indic matra with the character SPACE.

24.2 Appearance in code tables

Combining characters intended to be positioned relative to the associated character are depicted within the character code tables above, below, to the right of, to the left of, in, around, or through a dotted circle. In presentation, these characters are intended to be positioned relative to the preceding base character in some manner, and not to stand alone or function as base characters. This is the motivation for the term "combining". Diacritics are the principal class of combining characters used in European alphabets.

In the code tables for some scripts, such as Hebrew, Arabic, and the scripts of India and South East Asia, combining characters are indicated in relation to dotted circles to show their position relative to the base character. Many of these combining characters encode vowel letters; as such they are not generally referred to as "diacritical marks".

24.3 Multiple combining characters

There are instances where more than one combining character is applied to a single graphic character. ISO/IEC 10646 does not restrict the number of combining characters that can follow a base character. The following rules shall apply:

a) If the combining characters can interact in presentation (for example, COMBINING MACRON and COMBINING DIAERESIS), then the position of the combining characters in the resulting graphic display is determined by the order of the coded representation of the combining characters. The presentations of combining characters are to be positioned from the base character outward. For example, combining characters placed above a base character are stacked vertically, starting with the first encountered in the sequence of coded representations and continuing for as many marks above as are required by the coded combining characters following the coded base character. For combining characters placed below a base character, the situation is inverted, with the combining characters starting from the base character and stacking downward.

An example of multiple combining characters above the base character is found in Thai, where a consonant letter can have above it one of the vowels 0000 0E34 to 0000 0E37 and, above that, one of four tone marks 0000 0E48 to 0000 0E4B. The order of the coded representation is: base consonant, followed by a vowel, followed by a tone mark.

b) Some specific combining characters override the default stacking behaviour by being positioned horizontally rather than stacking, or by forming a ligature with an adjacent combining character. When positioned horizontally, the order of coded representations is reflected by positioning in the dominant order of the script with which they are used. For example, horizontal accents in a left-to-right script are coded left-to-right.

Prominent characters that show such override behaviour are associated with specific scripts or alphabets. For example, the COMBINING GREEK KORONIS (0000 0343) requires that, together with a following acute or grave accent, they be rendered side-by-side above a letter, rather than the accent marks being stacked above the COMBINING GREEK KORONIS. The order of the coded representations is: the letter itself, followed by that of the breathing mark, followed by that of the accent marks. Two Vietnamese tone marks which have the same graphic appearance as the Latin acute and grave accent marks do not stack above the three Vietnamese vowel letters which already contain the circumflex diacritic (â, ê, ô). Instead, they form liga-

tures with the circumflex component of the vowel letters.

c) If the combining characters do not interact in presentation (for example, when one combining character is above a graphic character and another is below), the resultant graphic symbol from the base character and combining characters in different orders may appear the same. For example, the coded representations of LATIN SMALL LETTER A, followed by COMBINING CARON, followed by COMBINING OGONEK may result in the same graphic symbol as the coded representations of LATIN SMALL LETTER A, followed by COMBINING OGONEK, followed by COMBINING CARON.

Combining characters in Hebrew or Arabic scripts do not normally interact. Therefore, the sequence of their coded representations in a composite sequence does not affect its graphic symbol. The rules for forming the combined graphic symbol are beyond the scope of ISO/IEC 10646.

NOTE - Where combining characters are used for the generation of composite sequences in implementation level 3, this facility may be used to provide an alternative coded representation of text. For example, in implementation level 3 the French word "là" may be represented by the characters LATIN SMALL LETTER L followed by LATIN SMALL LETTER A WITH GRAVE, or may be represented by the characters LATIN SMALL LETTER L followed by LATIN SMALL LETTER A followed by COMBINING GRAVE ACCENT.

24.4 Collections containing combining characters

In some collections of characters listed in annex A, such as collections 14 (BASIC ARABIC) or 25 (THAI), both combining characters and non-combining characters are included.

When implementation level 1 or 2 is adopted, a CC-data-element shall not contain the coded representations of combining characters listed in annex B, even though the adopted subset may include them.

Other collections of characters listed in annex A comprise only combining characters, for example collection 7 (COMBINING DIACRITICAL MARKS). Such a collection shall not be included in the adopted subset when implementation level 1 is adopted.

25 Special features of individual scripts

25.1 Hangul syllable composition method

In rendering, a sequence of Hangul Jamo (from HANGUL JAMO block: 1100 to 11FF) are displayed as a series of syllable blocks. Jamo can be classified into three classes: Choseong (syllable-initial character), Jungseong (syllable-peak character), and Jongseong (syllable-final character). A complete

syllable block is composed of a Choseong and a Jungseong, and optionally a Jongseong.

An incomplete syllable is a string of one or more characters which does not constitute a complete syllable (for example, a Choseong alone, a Jungseong alone, a Jongseong alone, or a Jungseong followed by a Jongseong). An incomplete syllable which starts with a Jungseong or a Jongseong shall be preceded by a CHOSEONG FILLER (0000 115F). An incomplete syllable composed of a Choseong alone shall be followed by a JUNGSEONG FILLER (0000 1160).

The implementation level 3 shall be used for the Hangul syllable composition method.

NOTE 1 - Hangul Jamo are not combining characters.

NOTE 2 - When a combining character such as HANGUL SINGLE DOT TONE MARK (0000 302E) is intended to apply to a sequence of Hangul Jamo it should be placed at the end of the sequence, after the Hangul Jamo character which completes the syllable block.

25.2 Features of Indic alphabetic scripts

In the tables for Rows 09 to 0D and 0F, and for the MYANMAR block in Row 10, of the BMP (see clause 26) the graphic symbols shown for some characters appear to be formed as compounds of the graphic symbols for two other characters in the same table.

Examples:

Row 0B Tamil. The graphic symbol for 0B94 TAMIL LETTER AU appears as if it is constructed from the graphic symbols for:

0B93 TAMIL LETTER OO and 0BD7 TAMIL AU LENGTH MARK

Row 0D Malayalam. The graphic symbol for 0D4A MALAYALAM VOWEL SIGN O appears as if it is constructed from the graphic symbols for:

0D46 MALAYALAM VOWEL SIGN E and 0D3E MALAYALAM VOWEL SIGN AA

In such cases a single coded character may appear to the user to be equivalent to the sequence of two coded characters whose graphic symbols, when combined, are visually similar to the graphic symbol of that single character, as in a composite sequence (4.14).

A “unique-spelling” rule is defined as follows. According to this rule, no coded character from a table for Rows 09 to 0D or 0F, or for the MYANMAR block in Row 10, shall be regarded as equivalent to a sequence of two or more other coded characters taken from the same table.

This “unique-spelling” rule shall apply in Levels 1 and 2.

NOTE - In Levels 1 and 2, if such a sequence occurs in a CC-data-element it is always made available to the user as

two distinct characters in accordance with their respective character names.

26 Code tables and lists of character names

26.1 General

An overview of the Basic Multilingual Plane is shown in figure 3. Detailed code tables and lists of character names for the Basic Multilingual Plane are shown on the following pages.

Guidelines to be used for constructing names of characters are given in annex L for information. In some cases, a name of a character is followed by additional explanatory statements not part of the name. These statements are in parentheses and not in capital letters except for the initials of the word, where required.

26.2 Character names and annotations for Hangul syllables

Names for the Hangul syllable characters in code positions 0000 AC00 - 0000 D7A3 are derived from their code position numbers by the numerical procedure described below. Lists of names for these characters are not provided opposite the code tables.

1. Obtain the code position number of the Hangul syllable character. It is of the form 0000 $h_1h_2h_3h_4$ where h_1 , h_2 , h_3 , and h_4 are hexadecimal digits; h_1h_2 is the Row number within the BMP and h_3h_4 is the cell number within the row. The number $h_1h_2h_3h_4$ lies within the range AC00 to D7A3.

2. Derive the decimal numbers d_1 , d_2 , d_3 , d_4 that are numerically equal to the hexadecimal digits h_1 , h_2 , h_3 , h_4 respectively.

3. Calculate the character index C from the formula:

$$C = 4096 \times (d_1 - 10) + 256 \times (d_2 - 12) + 16 \times d_3 + d_4$$

Note: If $C < 0$ or $> 11,171$ then the character is not in the HANGUL SYLLABLES block.

4. Calculate the syllable component indices I , P , F from the following formulae:

$$I = C / 588 \quad (\text{Note: } 0 \leq I \leq 18)$$

$$P = (C \% 588) / 28 \quad (\text{Note: } 0 \leq P \leq 20)$$

$$F = C \% 28 \quad (\text{Note: } 0 \leq F \leq 27)$$

where “/” indicates integer division (i.e. x / y is the integer quotient of the division), and “%” indicates the modulo operation (i.e. $x \% y$ is the remainder after the integer division x / y).

5. Obtain the Latin character strings that correspond to the three indices I , P , F from columns 2, 3, and

4 respectively of table 1 below (for $I = 11$ and for $F = 0$ the corresponding strings are null). Concatenate these three strings in left-to-right order to make a single string, the syllable-name.

6. The character name for the character at position 0000 $h_1h_2h_3h_4$ is then:

HANGUL SYLLABLE $s-n$

where “ $s-n$ ” indicates the syllable-name string derived in step 5.

Example.

For the character in code position D4DE:

$d_1 = 13$, $d_2 = 4$, $d_3 = 13$, $d_4 = 14$.

$C = 10462$

$I = 17$, $P = 16$, $F = 18$.

The corresponding Latin character strings are:

P, WI, BS.

The syllable-name is PWIBS, and the character name is:

HANGUL SYLLABLE PWIBS

For convenience a list of the syllable-names is provided in annex R.

For each Hangul syllable character a short annotation is defined. This annotation consists of an altern-

ative transliteration of the Hangul syllable into Latin characters.

Annotations for the Hangul syllable characters in code positions 0000 AC00 - 0000 D7A3 are also derived from their code position numbers by a similar numerical procedure described below.

7. Carry out steps 1 to 4 as described above.

8. Obtain the Latin character strings that correspond to the three indices I , P , F from columns 5, 6, and 7 respectively of Table 1 below (for $I = 11$ and for $F = 0$ the corresponding strings are null). Concatenate these three strings in left-to-right order to make a single string, and enclose it within parentheses to form the annotation.

Example.

For the character in code position D4DE:

$d_1 = 13$, $d_2 = 4$, $d_3 = 13$, $d_4 = 14$.

$C = 10462$

$I = 17$, $P = 16$, $F = 18$.

The corresponding Latin character strings are:

ph, wi, ps,

and the annotation is (phwips).

Table 1: Elements of Hangul syllable names and annotations

Index number	Syllable name elements			Annotation elements		
	<i>I</i> string	<i>P</i> string	<i>F</i> string	<i>I</i> string	<i>P</i> string	<i>F</i> string
0	G	A		k	a	
1	GG	AE	G	kk	ae	k
2	N	YA	GG	n	ya	kk
3	D	YAE	GS	t	yae	ks
4	DD	EO	N	tt	eo	n
5	R	E	NJ	r	e	nc
6	M	YEO	NH	m	yeo	nh
7	B	YE	D	p	ye	t
8	BB	O	L	pp	o	l
9	S	WA	LG	s	wa	lk
10	SS	WAE	LM	ss	wae	lm
11		OE	LB		oe	lp
12	J	YO	LS	c	yo	ls
13	JJ	U	LT	cc	u	lth
14	C	WEO	LP	ch	weo	lph
15	K	WE	LH	kh	we	lh
16	T	WI	M	th	wi	m
17	P	YU	B	ph	yu	p
18	H	EU	BS	h	eu	ps
19		YI	S		yi	s
20		I	SS		i	ss
21			NG			ng
22			J			c
23			C			ch
24			K			kh
25			T			th
26			P			ph
27			H			h

Row-octet

00	Rows 00 to 33 (see figure 4)					
..						
..						
..						
..						
..						
33						
34						
..						
4D						
4E	CJK Unified Ideographs Extension A					
..						
..						
..						
..						
..						
..						
..						
9F						
A0..						
A3	Yi Syllables				Yi Radicals	
A4						
A5..						
AB						
AC						
..	Hangul Syllables					
..						
..						
..						
D7						
D8..						
DF						
E0	S-zone (for use in UTF-16 only)					
..						
..						
F8						
F9	Private Use Zone					
FA						
FB						
FC						
FD	CJK Compatibility Ideographs					
FE						
FF						
	Alphabetic Presentation Forms					
	Arabic Presentation Forms-A					
	Comb. Half M'ks CJK Compat. F'ms Small Form Variants Arabic Presentation Forms-B					
	Halfwidth And Fullwidth Forms Special					

= not graphic characters

= reserved for future standardization

NOTE - Vertical boundaries within rows are indicated in approximate positions only.

Figure 3 - Overview of the Basic Multilingual Plane

Row-octet

00		Basic Latin		Latin-1 Supplement
01		Latin Extended-A		Latin Extended-B
02	Latin Extended-B	IPA (Intl. Phonetic Alph.) Extensions		Spacing Modifier Letters
03	Combining Diacritical Marks		Greek and Coptic	
04		Cyrillic		
05		Armenian		Hebrew
06		Arabic		
07	Syriac		Thaana	
08				
09	Devanagari		Bengali	
0A	Gurmukhi		Gujarati	
0B	Oriya		Tamil	
0C	Telugu		Kannada	
0D	Malayalam		Sinhala	
0E	Thai		Lao	
0F		Tibetan		
10	Myanmar		Georgian	
11		Hangul Jamo		
12		Ethiopic		
13			Cherokee	
14		Unified Canadian Aboriginal Syllabics		
16		Ogham	Runic	
17			Khmer	
18	Mongolian			
19				
..				
1D				
1E		Latin Extended Additional		
1F		Greek Extended		
20	General Punctuation	Super-/Subscripts	Currency Symbols	Comb. Mks. Symb.
21	Letterlike Symbols	Number Forms	Arrows	
22		Mathematical Operators		
23		Miscellaneous Technical		
24	Control Pictures	O.C.R.	Enclosed Alphanumerics	
25	Box Drawing	Block Elements	Geometric Shapes	
26		Miscellaneous Symbols		
27	Dingbats			
28		Braille Patterns		
29				
..				
2D				
2E			CJK Radicals Supplement	
2F		Kangxi Radicals		Ideog. Descr.
30	CJK Symbols And Punctuation	Hiragana	Katakana	
31	Bopomofo	Hangul Compatibility Jamo	Kanbun	Bopomofo Ext.
32		Enclosed CJK Letters And Months		
33		CJK Compatibility		

 = not graphic characters

 = reserved for future standardization

NOTE - Vertical boundaries within rows are indicated in approximate positions only.

Figure 4 - Overview of Rows 00 to 33 of the Basic Multilingual Plane

Table 2 - Row 00: Basic Latin

	000	001	002	003	004	005	006	007
0			SP 0020	0 0030	@ 0040	P 0050	` 0060	p 0070
1			! 0021	1 0031	A 0041	Q 0051	a 0061	q 0071
2			" 0022	2 0032	B 0042	R 0052	b 0062	r 0072
3			# 0023	3 0033	C 0043	S 0053	c 0063	s 0073
4			\$ 0024	4 0034	D 0044	T 0054	d 0064	t 0074
5			% 0025	5 0035	E 0045	U 0055	e 0065	u 0075
6			& 0026	6 0036	F 0046	V 0056	f 0066	v 0076
7			' 0027	7 0037	G 0047	W 0057	g 0067	w 0077
8			(0028	8 0038	H 0048	X 0058	h 0068	x 0078
9			) 0029	9 0039	I 0049	Y 0059	i 0069	y 0079
A			* 002A	: 003A	J 004A	Z 005A	j 006A	z 007A
B			+ 002B	; 003B	K 004B	[005B	k 006B	{ 007B
C			, 002C	< 003C	L 004C	\ 005C	l 006C	 007C
D			- 002D	= 003D	M 004D	] 005D	m 006D	} 007D
E			. 002E	> 003E	N 004E	^ 005E	n 006E	~ 007E
F			/ 002F	? 003F	O 004F	_ 005F	o 006F	

G = 00
P = 00

Table 2 - Row 00: Basic Latin

HEX	NAME	HEX	NAME
0020	SPACE	007B	LEFT CURLY BRACKET
0021	EXCLAMATION MARK	007C	VERTICAL LINE
0022	QUOTATION MARK	007D	RIGHT CURLY BRACKET
0023	NUMBER SIGN	007E	TILDE
0024	DOLLAR SIGN		
0025	PERCENT SIGN		
0026	AMPERSAND		
0027	APOSTROPHE		
0028	LEFT PARENTHESIS		
0029	RIGHT PARENTHESIS		
002A	ASTERISK		
002B	PLUS SIGN		
002C	COMMA		
002D	HYPHEN-MINUS		
002E	FULL STOP		
002F	SOLIDUS		
0030	DIGIT ZERO		
0031	DIGIT ONE		
0032	DIGIT TWO		
0033	DIGIT THREE		
0034	DIGIT FOUR		
0035	DIGIT FIVE		
0036	DIGIT SIX		
0037	DIGIT SEVEN		
0038	DIGIT EIGHT		
0039	DIGIT NINE		
003A	COLON		
003B	SEMICOLON		
003C	LESS-THAN SIGN		
003D	EQUALS SIGN		
003E	GREATER-THAN SIGN		
003F	QUESTION MARK		
0040	COMMERCIAL AT		
0041	LATIN CAPITAL LETTER A		
0042	LATIN CAPITAL LETTER B		
0043	LATIN CAPITAL LETTER C		
0044	LATIN CAPITAL LETTER D		
0045	LATIN CAPITAL LETTER E		
0046	LATIN CAPITAL LETTER F		
0047	LATIN CAPITAL LETTER G		
0048	LATIN CAPITAL LETTER H		
0049	LATIN CAPITAL LETTER I		
004A	LATIN CAPITAL LETTER J		
004B	LATIN CAPITAL LETTER K		
004C	LATIN CAPITAL LETTER L		
004D	LATIN CAPITAL LETTER M		
004E	LATIN CAPITAL LETTER N		
004F	LATIN CAPITAL LETTER O		
0050	LATIN CAPITAL LETTER P		
0051	LATIN CAPITAL LETTER Q		
0052	LATIN CAPITAL LETTER R		
0053	LATIN CAPITAL LETTER S		
0054	LATIN CAPITAL LETTER T		
0055	LATIN CAPITAL LETTER U		
0056	LATIN CAPITAL LETTER V		
0057	LATIN CAPITAL LETTER W		
0058	LATIN CAPITAL LETTER X		
0059	LATIN CAPITAL LETTER Y		
005A	LATIN CAPITAL LETTER Z		
005B	LEFT SQUARE BRACKET		
005C	REVERSE SOLIDUS		
005D	RIGHT SQUARE BRACKET		
005E	CIRCUMFLEX ACCENT		
005F	LOW LINE		
0060	GRAVE ACCENT		
0061	LATIN SMALL LETTER A		
0062	LATIN SMALL LETTER B		
0063	LATIN SMALL LETTER C		
0064	LATIN SMALL LETTER D		
0065	LATIN SMALL LETTER E		
0066	LATIN SMALL LETTER F		
0067	LATIN SMALL LETTER G		
0068	LATIN SMALL LETTER H		
0069	LATIN SMALL LETTER I		
006A	LATIN SMALL LETTER J		
006B	LATIN SMALL LETTER K		
006C	LATIN SMALL LETTER L		
006D	LATIN SMALL LETTER M		
006E	LATIN SMALL LETTER N		
006F	LATIN SMALL LETTER O		
0070	LATIN SMALL LETTER P		
0071	LATIN SMALL LETTER Q		
0072	LATIN SMALL LETTER R		
0073	LATIN SMALL LETTER S		
0074	LATIN SMALL LETTER T		
0075	LATIN SMALL LETTER U		
0076	LATIN SMALL LETTER V		
0077	LATIN SMALL LETTER W		
0078	LATIN SMALL LETTER X		
0079	LATIN SMALL LETTER Y		
007A	LATIN SMALL LETTER Z		

Group 00

Plane 00

Table 3 - Row 00: Latin-1 Supplement

	008	009	00A	00B	00C	00D	00E	00F
0			 00A0	° 00B0	À 00C0	Ð 00D0	à 00E0	ð 00F0
1			¡ 00A1	± 00B1	Á 00C1	Ñ 00D1	á 00E1	ñ 00F1
2			¢ 00A2	² 00B2	Â 00C2	Ò 00D2	â 00E2	ò 00F2
3			£ 00A3	³ 00B3	Ã 00C3	Ó 00D3	ã 00E3	ó 00F3
4			¤ 00A4	´ 00B4	Ä 00C4	Ô 00D4	ä 00E4	ô 00F4
5			¥ 00A5	µ 00B5	Å 00C5	Õ 00D5	å 00E5	õ 00F5
6			¦ 00A6	¶ 00B6	Æ 00C6	Ö 00D6	æ 00E6	ö 00F6
7			§ 00A7	· 00B7	Ç 00C7	× 00D7	ç 00E7	÷ 00F7
8			¨ 00A8	¸ 00B8	È 00C8	Ø 00D8	è 00E8	ø 00F8
9			© 00A9	¹ 00B9	É 00C9	Ù 00D9	é 00E9	ù 00F9
A			ª 00AA	º 00BA	Ê 00CA	Ú 00DA	ê 00EA	ú 00FA
B			« 00AB	» 00BB	Ë 00CB	Û 00DB	ë 00EB	û 00FB
C			¬ 00AC	¼ 00BC	Ì 00CC	Ü 00DC	ì 00EC	ü 00FC
D			 00AD	½ 00BD	Í 00CD	Ý 00DD	í 00ED	ý 00FD
E			® 00AE	¾ 00BE	Î 00CE	Þ 00DE	î 00EE	þ 00FE
F			¯ 00AF	¿ 00BF	Ï 00CF	ß 00DF	ï 00EF	ÿ 00FF

G = 00
P = 00

Table 3 - Row 00: Latin-1 Supplement

HEX	NAME	HEX	NAME
00A0	NO-BREAK SPACE	00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX
00A1	INVERTED EXCLAMATION MARK	00FC	LATIN SMALL LETTER U WITH DIAERESIS
00A2	CENT SIGN	00FD	LATIN SMALL LETTER Y WITH ACUTE
00A3	POUND SIGN	00FE	LATIN SMALL LETTER THORN (Icelandic)
00A4	CURRENCY SIGN	00FF	LATIN SMALL LETTER Y WITH DIAERESIS
00A5	YEN SIGN		
00A6	BROKEN BAR		
00A7	SECTION SIGN		
00A8	DIAERESIS		
00A9	COPYRIGHT SIGN		
00AA	FEMININE ORDINAL INDICATOR		
00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK *		
00AC	NOT SIGN		
00AD	SOFT HYPHEN		
00AE	REGISTERED SIGN		
00AF	MACRON		
00B0	DEGREE SIGN		
00B1	PLUS-MINUS SIGN		
00B2	SUPERSCRIT TWO		
00B3	SUPERSCRIT THREE		
00B4	ACUTE ACCENT		
00B5	MICRO SIGN		
00B6	PILCROW SIGN		
00B7	MIDDLE DOT		
00B8	CEDILLA		
00B9	SUPERSCRIT ONE		
00BA	MASCULINE ORDINAL INDICATOR		
00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK *		
00BC	VULGAR FRACTION ONE QUARTER		
00BD	VULGAR FRACTION ONE HALF		
00BE	VULGAR FRACTION THREE QUARTERS		
00BF	INVERTED QUESTION MARK		
00C0	LATIN CAPITAL LETTER A WITH GRAVE		
00C1	LATIN CAPITAL LETTER A WITH ACUTE		
00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX		
00C3	LATIN CAPITAL LETTER A WITH TILDE		
00C4	LATIN CAPITAL LETTER A WITH DIAERESIS		
00C5	LATIN CAPITAL LETTER A WITH RING ABOVE		
00C6	LATIN CAPITAL LETTER AE (ash) *		
00C7	LATIN CAPITAL LETTER C WITH CEDILLA		
00C8	LATIN CAPITAL LETTER E WITH GRAVE		
00C9	LATIN CAPITAL LETTER E WITH ACUTE		
00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX		
00CB	LATIN CAPITAL LETTER E WITH DIAERESIS		
00CC	LATIN CAPITAL LETTER I WITH GRAVE		
00CD	LATIN CAPITAL LETTER I WITH ACUTE		
00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX		
00CF	LATIN CAPITAL LETTER I WITH DIAERESIS		
00D0	LATIN CAPITAL LETTER ETH (Icelandic)		
00D1	LATIN CAPITAL LETTER N WITH TILDE		
00D2	LATIN CAPITAL LETTER O WITH GRAVE		
00D3	LATIN CAPITAL LETTER O WITH ACUTE		
00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX		
00D5	LATIN CAPITAL LETTER O WITH TILDE		
00D6	LATIN CAPITAL LETTER O WITH DIAERESIS		
00D7	MULTIPLICATION SIGN		
00D8	LATIN CAPITAL LETTER O WITH STROKE		
00D9	LATIN CAPITAL LETTER U WITH GRAVE		
00DA	LATIN CAPITAL LETTER U WITH ACUTE		
00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX		
00DC	LATIN CAPITAL LETTER U WITH DIAERESIS		
00DD	LATIN CAPITAL LETTER Y WITH ACUTE		
00DE	LATIN CAPITAL LETTER THORN (Icelandic)		
00DF	LATIN SMALL LETTER SHARP S (German)		
00E0	LATIN SMALL LETTER A WITH GRAVE		
00E1	LATIN SMALL LETTER A WITH ACUTE		
00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX		
00E3	LATIN SMALL LETTER A WITH TILDE		
00E4	LATIN SMALL LETTER A WITH DIAERESIS		
00E5	LATIN SMALL LETTER A WITH RING ABOVE		
00E6	LATIN SMALL LETTER AE (ash) *		
00E7	LATIN SMALL LETTER C WITH CEDILLA		
00E8	LATIN SMALL LETTER E WITH GRAVE		
00E9	LATIN SMALL LETTER E WITH ACUTE		
00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX		
00EB	LATIN SMALL LETTER E WITH DIAERESIS		
00EC	LATIN SMALL LETTER I WITH GRAVE		
00ED	LATIN SMALL LETTER I WITH ACUTE		
00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX		
00EF	LATIN SMALL LETTER I WITH DIAERESIS		
00F0	LATIN SMALL LETTER ETH (Icelandic)		
00F1	LATIN SMALL LETTER N WITH TILDE		
00F2	LATIN SMALL LETTER O WITH GRAVE		
00F3	LATIN SMALL LETTER O WITH ACUTE		
00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX		
00F5	LATIN SMALL LETTER O WITH TILDE		
00F6	LATIN SMALL LETTER O WITH DIAERESIS		
00F7	DIVISION SIGN		
00F8	LATIN SMALL LETTER O WITH STROKE		
00F9	LATIN SMALL LETTER U WITH GRAVE		
00FA	LATIN SMALL LETTER U WITH ACUTE		

Group 00

Plane 00

Row 00



Documents

Document Sharing Hub

Table 4 - Row 01: Latin Extended-A

	010	011	012	013	014	015	016	017
0	Ā 0100	Đ 0110	Ġ 0120	İ 0130	ı· 0140	Ŏ 0150	Š 0160	Ű 0170
1	ā 0101	đ 0111	ġ 0121	ı 0131	Ł 0141	ő 0151	š 0161	ű 0171
2	Ǻ 0102	Ē 0112	Ģ 0122	IJ 0132	ł 0142	Œ 0152	Ŧ 0162	Ț 0172
3	ǻ 0103	ē 0113	ġ 0123	ij 0133	Ŋ 0143	œ 0153	ţ 0163	ț 0173
4	AꞤ 0104	Ě 0114	Ĥ 0124	Ĵ 0134	ń 0144	Ŕ 0154	Ť 0164	Ŵ 0174
5	ą 0105	ě 0115	ĥ 0125	ĵ 0135	Ņ 0145	ŕ 0155	ť 0165	ŵ 0175
6	Ć 0106	Ė 0116	Ħ 0126	Ꞥ 0136	ņ 0146	Ŗ 0156	Ŧ 0166	Ŷ 0176
7	ć 0107	ė 0117	ħ 0127	ꞥ 0137	Ň 0147	ŗ 0157	ţ 0167	ŷ 0177
8	Ĉ 0108	Ę 0118	Ĩ 0128	κ 0138	ň 0148	Ř 0158	Ũ 0168	Ÿ 0178
9	ĉ 0109	ę 0119	ĩ 0129	Ł 0139	’n 0149	ř 0159	ũ 0169	Ž 0179
A	Ċ 010A	Ė 011A	Ī 012A	Í 013A	Ņ 014A	Ś 015A	Ū 016A	ž 017A
B	ċ 010B	ė 011B	ī 012B	Ł 013B	ņ 014B	ś 015B	ū 016B	Ž 017B
C	Č 010C	Ĝ 011C	Ĭ 012C	ı 013C	Ō 014C	Ŝ 015C	Ů 016C	ž 017C
D	č 010D	ĝ 011D	ĭ 012D	Ł 013D	ō 014D	ŝ 015D	ů 016D	Ž 017D
E	Ǻ 010E	Ǻ 011E	Ǻ 012E	Ǻ 013E	Ǻ 014E	Ǻ 015E	Ǻ 016E	Ǻ 017E
F	d’ 010F	Ǻ 011F	Ǻ 012F	L· 013F	Ǻ 014F	Ǻ 015F	Ǻ 016F	f 017F

G = 00
P = 00

Table 4 - Row 01: Latin Extended-A

HEX	NAME	HEX	NAME
0100	LATIN CAPITAL LETTER A WITH MACRON	015B	LATIN SMALL LETTER S WITH ACUTE
0101	LATIN SMALL LETTER A WITH MACRON	015C	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
0102	LATIN CAPITAL LETTER A WITH BREVE	015D	LATIN SMALL LETTER S WITH CIRCUMFLEX
0103	LATIN SMALL LETTER A WITH BREVE	015E	LATIN CAPITAL LETTER S WITH CEDILLA *
0104	LATIN CAPITAL LETTER A WITH OGONEK	015F	LATIN SMALL LETTER S WITH CEDILLA *
0105	LATIN SMALL LETTER A WITH OGONEK	0160	LATIN CAPITAL LETTER S WITH CARON
0106	LATIN CAPITAL LETTER C WITH ACUTE	0161	LATIN SMALL LETTER S WITH CARON
0107	LATIN SMALL LETTER C WITH ACUTE	0162	LATIN CAPITAL LETTER T WITH CEDILLA *
0108	LATIN CAPITAL LETTER C WITH CIRCUMFLEX	0163	LATIN SMALL LETTER T WITH CEDILLA *
0109	LATIN SMALL LETTER C WITH CIRCUMFLEX	0164	LATIN CAPITAL LETTER T WITH CARON
010A	LATIN CAPITAL LETTER C WITH DOT ABOVE	0165	LATIN SMALL LETTER T WITH CARON
010B	LATIN SMALL LETTER C WITH DOT ABOVE	0166	LATIN CAPITAL LETTER T WITH STROKE
010C	LATIN CAPITAL LETTER C WITH CARON	0167	LATIN SMALL LETTER T WITH STROKE
010D	LATIN SMALL LETTER C WITH CARON	0168	LATIN CAPITAL LETTER U WITH TILDE
010E	LATIN CAPITAL LETTER D WITH CARON	0169	LATIN SMALL LETTER U WITH TILDE
010F	LATIN SMALL LETTER D WITH CARON	016A	LATIN CAPITAL LETTER U WITH MACRON
0110	LATIN CAPITAL LETTER D WITH STROKE	016B	LATIN SMALL LETTER U WITH MACRON
0111	LATIN SMALL LETTER D WITH STROKE	016C	LATIN CAPITAL LETTER U WITH BREVE
0112	LATIN CAPITAL LETTER E WITH MACRON	016D	LATIN SMALL LETTER U WITH BREVE
0113	LATIN SMALL LETTER E WITH MACRON	016E	LATIN CAPITAL LETTER U WITH RING ABOVE
0114	LATIN CAPITAL LETTER E WITH BREVE	016F	LATIN SMALL LETTER U WITH RING ABOVE
0115	LATIN SMALL LETTER E WITH BREVE	0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
0116	LATIN CAPITAL LETTER E WITH DOT ABOVE	0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE
0117	LATIN SMALL LETTER E WITH DOT ABOVE	0172	LATIN CAPITAL LETTER U WITH OGONEK
0118	LATIN CAPITAL LETTER E WITH OGONEK	0173	LATIN SMALL LETTER U WITH OGONEK
0119	LATIN SMALL LETTER E WITH OGONEK	0174	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
011A	LATIN CAPITAL LETTER E WITH CARON	0175	LATIN SMALL LETTER W WITH CIRCUMFLEX
011B	LATIN SMALL LETTER E WITH CARON	0176	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
011C	LATIN CAPITAL LETTER G WITH CIRCUMFLEX	0177	LATIN SMALL LETTER Y WITH CIRCUMFLEX
011D	LATIN SMALL LETTER G WITH CIRCUMFLEX	0178	LATIN CAPITAL LETTER Y WITH DIAERESIS
011E	LATIN CAPITAL LETTER G WITH BREVE	0179	LATIN CAPITAL LETTER Z WITH ACUTE
011F	LATIN SMALL LETTER G WITH BREVE	017A	LATIN SMALL LETTER Z WITH ACUTE
0120	LATIN CAPITAL LETTER G WITH DOT ABOVE	017B	LATIN CAPITAL LETTER Z WITH DOT ABOVE
0121	LATIN SMALL LETTER G WITH DOT ABOVE	017C	LATIN SMALL LETTER Z WITH DOT ABOVE
0122	LATIN CAPITAL LETTER G WITH CEDILLA	017D	LATIN CAPITAL LETTER Z WITH CARON
0123	LATIN SMALL LETTER G WITH CEDILLA	017E	LATIN SMALL LETTER Z WITH CARON
0124	LATIN CAPITAL LETTER H WITH CIRCUMFLEX	017F	LATIN SMALL LETTER LONG S
0125	LATIN SMALL LETTER H WITH CIRCUMFLEX		
0126	LATIN CAPITAL LETTER H WITH STROKE		
0127	LATIN SMALL LETTER H WITH STROKE		
0128	LATIN CAPITAL LETTER I WITH TILDE		
0129	LATIN SMALL LETTER I WITH TILDE		
012A	LATIN CAPITAL LETTER I WITH MACRON		
012B	LATIN SMALL LETTER I WITH MACRON		
012C	LATIN CAPITAL LETTER I WITH BREVE		
012D	LATIN SMALL LETTER I WITH BREVE		
012E	LATIN CAPITAL LETTER I WITH OGONEK		
012F	LATIN SMALL LETTER I WITH OGONEK		
0130	LATIN CAPITAL LETTER I WITH DOT ABOVE		
0131	LATIN SMALL LETTER DOTLESS I		
0132	LATIN CAPITAL LIGATURE IJ		
0133	LATIN SMALL LIGATURE IJ		
0134	LATIN CAPITAL LETTER J WITH CIRCUMFLEX		
0135	LATIN SMALL LETTER J WITH CIRCUMFLEX		
0136	LATIN CAPITAL LETTER K WITH CEDILLA		
0137	LATIN SMALL LETTER K WITH CEDILLA		
0138	LATIN SMALL LETTER KRA (Greenlandic)		
0139	LATIN CAPITAL LETTER L WITH ACUTE		
013A	LATIN SMALL LETTER L WITH ACUTE		
013B	LATIN CAPITAL LETTER L WITH CEDILLA		
013C	LATIN SMALL LETTER L WITH CEDILLA		
013D	LATIN CAPITAL LETTER L WITH CARON		
013E	LATIN SMALL LETTER L WITH CARON		
013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT		
0140	LATIN SMALL LETTER L WITH MIDDLE DOT		
0141	LATIN CAPITAL LETTER L WITH STROKE		
0142	LATIN SMALL LETTER L WITH STROKE		
0143	LATIN CAPITAL LETTER N WITH ACUTE		
0144	LATIN SMALL LETTER N WITH ACUTE		
0145	LATIN CAPITAL LETTER N WITH CEDILLA		
0146	LATIN SMALL LETTER N WITH CEDILLA		
0147	LATIN CAPITAL LETTER N WITH CARON		
0148	LATIN SMALL LETTER N WITH CARON		
0149	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE		
014A	LATIN CAPITAL LETTER ENG (Sami)		
014B	LATIN SMALL LETTER ENG (Sami)		
014C	LATIN CAPITAL LETTER O WITH MACRON		
014D	LATIN SMALL LETTER O WITH MACRON		
014E	LATIN CAPITAL LETTER O WITH BREVE		
014F	LATIN SMALL LETTER O WITH BREVE		
0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE		
0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE		
0152	LATIN CAPITAL LIGATURE OE		
0153	LATIN SMALL LIGATURE OE		
0154	LATIN CAPITAL LETTER R WITH ACUTE		
0155	LATIN SMALL LETTER R WITH ACUTE		
0156	LATIN CAPITAL LETTER R WITH CEDILLA		
0157	LATIN SMALL LETTER R WITH CEDILLA		
0158	LATIN CAPITAL LETTER R WITH CARON		
0159	LATIN SMALL LETTER R WITH CARON		
015A	LATIN CAPITAL LETTER S WITH ACUTE		

Group 00

Plane 00

Table 5 - Row 01: Latin Extended-B

	018	019	01A	01B	01C	01D	01E
0	Ƨ 0180	Ǝ 0190	Ɔ 01A0	Ƨ 01B0	Ƨ 01C0	Ƨ 01D0	Ƨ 01E0
1	Ƨ 0181	Ǝ 0191	Ɔ 01A1	Ƨ 01B1	Ƨ 01C1	Ƨ 01D1	Ƨ 01E1
2	Ƨ 0182	Ǝ 0192	Ɔ 01A2	Ƨ 01B2	Ƨ 01C2	Ƨ 01D2	Ƨ 01E2
3	Ƨ 0183	Ǝ 0193	Ɔ 01A3	Ƨ 01B3	Ƨ 01C3	Ƨ 01D3	Ƨ 01E3
4	Ƨ 0184	Ǝ 0194	Ɔ 01A4	Ƨ 01B4	Ƨ 01C4	Ƨ 01D4	Ƨ 01E4
5	Ƨ 0185	Ǝ 0195	Ɔ 01A5	Ƨ 01B5	Ƨ 01C5	Ƨ 01D5	Ƨ 01E5
6	Ƨ 0186	Ǝ 0196	Ɔ 01A6	Ƨ 01B6	Ƨ 01C6	Ƨ 01D6	Ƨ 01E6
7	Ƨ 0187	Ǝ 0197	Ɔ 01A7	Ƨ 01B7	Ƨ 01C7	Ƨ 01D7	Ƨ 01E7
8	Ƨ 0188	Ǝ 0198	Ɔ 01A8	Ƨ 01B8	Ƨ 01C8	Ƨ 01D8	Ƨ 01E8
9	Ƨ 0189	Ǝ 0199	Ɔ 01A9	Ƨ 01B9	Ƨ 01C9	Ƨ 01D9	Ƨ 01E9
A	Ƨ 018A	Ǝ 019A	Ɔ 01AA	Ƨ 01BA	Ƨ 01CA	Ƨ 01DA	Ƨ 01EA
B	Ƨ 018B	Ǝ 019B	Ɔ 01AB	Ƨ 01BB	Ƨ 01CB	Ƨ 01DB	Ƨ 01EB
C	Ƨ 018C	Ǝ 019C	Ɔ 01AC	Ƨ 01BC	Ƨ 01CC	Ƨ 01DC	Ƨ 01EC
D	Ƨ 018D	Ǝ 019D	Ɔ 01AD	Ƨ 01BD	Ƨ 01CD	Ƨ 01DD	Ƨ 01ED
E	Ƨ 018E	Ǝ 019E	Ɔ 01AE	Ƨ 01BE	Ƨ 01CE	Ƨ 01DE	Ƨ 01EE
F	Ƨ 018F	Ǝ 019F	Ɔ 01AF	Ƨ 01BF	Ƨ 01CF	Ƨ 01DF	Ƨ 01EF

G = 00
P = 00

Table 5 - Row 01: Latin Extended-B

HEX	NAME	HEX	NAME
0180	LATIN SMALL LETTER B WITH STROKE	01DB	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE
0181	LATIN CAPITAL LETTER B WITH HOOK	01DC	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
0182	LATIN CAPITAL LETTER B WITH TOPBAR	01DD	LATIN SMALL LETTER TURNED E
0183	LATIN SMALL LETTER B WITH TOPBAR	01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON
0184	LATIN CAPITAL LETTER TONE SIX	01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
0185	LATIN SMALL LETTER TONE SIX	01E0	LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON
0186	LATIN CAPITAL LETTER OPEN O	01E1	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
0187	LATIN CAPITAL LETTER C WITH HOOK	01E2	LATIN CAPITAL LETTER AE WITH MACRON (ash) *
0188	LATIN SMALL LETTER C WITH HOOK	01E3	LATIN SMALL LETTER AE WITH MACRON (ash) *
0189	LATIN CAPITAL LETTER AFRICAN D *	01E4	LATIN CAPITAL LETTER G WITH STROKE
018A	LATIN CAPITAL LETTER D WITH HOOK	01E5	LATIN SMALL LETTER G WITH STROKE
018B	LATIN CAPITAL LETTER D WITH TOPBAR	01E6	LATIN CAPITAL LETTER G WITH CARON
018C	LATIN SMALL LETTER D WITH TOPBAR	01E7	LATIN SMALL LETTER G WITH CARON
018D	LATIN SMALL LETTER TURNED DELTA	01E8	LATIN CAPITAL LETTER K WITH CARON
018E	LATIN CAPITAL LETTER REVERSED E	01E9	LATIN SMALL LETTER K WITH CARON
018F	LATIN CAPITAL LETTER SCHWA	01EA	LATIN CAPITAL LETTER O WITH OGONEK
0190	LATIN CAPITAL LETTER OPEN E	01EB	LATIN SMALL LETTER O WITH OGONEK
0191	LATIN CAPITAL LETTER F WITH HOOK	01EC	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON
0192	LATIN SMALL LETTER F WITH HOOK	01ED	LATIN SMALL LETTER O WITH OGONEK AND MACRON
0193	LATIN CAPITAL LETTER G WITH HOOK	01EE	LATIN CAPITAL LETTER EZH WITH CARON
0194	LATIN CAPITAL LETTER GAMMA	01EF	LATIN SMALL LETTER EZH WITH CARON
0195	LATIN SMALL LETTER HV (hwair)		
0196	LATIN CAPITAL LETTER IOTA		
0197	LATIN CAPITAL LETTER I WITH STROKE		
0198	LATIN CAPITAL LETTER K WITH HOOK		
0199	LATIN SMALL LETTER K WITH HOOK		
019A	LATIN SMALL LETTER L WITH BAR		
019B	LATIN SMALL LETTER LAMBDA WITH STROKE		
019C	LATIN CAPITAL LETTER TURNED M		
019D	LATIN CAPITAL LETTER N WITH LEFT HOOK		
019E	LATIN SMALL LETTER N WITH LONG RIGHT LEG		
019F	LATIN CAPITAL LETTER O WITH MIDDLE TILDE *		
01A0	LATIN CAPITAL LETTER O WITH HORN		
01A1	LATIN SMALL LETTER O WITH HORN		
01A2	LATIN CAPITAL LETTER OI (gha)		
01A3	LATIN SMALL LETTER OI (gha)		
01A4	LATIN CAPITAL LETTER P WITH HOOK		
01A5	LATIN SMALL LETTER P WITH HOOK		
01A6	LATIN LETTER YR		
01A7	LATIN CAPITAL LETTER TONE TWO		
01A8	LATIN SMALL LETTER TONE TWO		
01A9	LATIN CAPITAL LETTER ESH		
01AA	LATIN LETTER REVERSED ESH LOOP		
01AB	LATIN SMALL LETTER T WITH PALATAL HOOK		
01AC	LATIN CAPITAL LETTER T WITH HOOK		
01AD	LATIN SMALL LETTER T WITH HOOK		
01AE	LATIN CAPITAL LETTER T WITH RETROFLEX HOOK		
01AF	LATIN CAPITAL LETTER U WITH HORN		
01B0	LATIN SMALL LETTER U WITH HORN		
01B1	LATIN CAPITAL LETTER UPSILON		
01B2	LATIN CAPITAL LETTER V WITH HOOK		
01B3	LATIN CAPITAL LETTER Y WITH HOOK		
01B4	LATIN SMALL LETTER Y WITH HOOK		
01B5	LATIN CAPITAL LETTER Z WITH STROKE		
01B6	LATIN SMALL LETTER Z WITH STROKE		
01B7	LATIN CAPITAL LETTER EZH		
01B8	LATIN CAPITAL LETTER EZH REVERSED		
01B9	LATIN SMALL LETTER EZH REVERSED		
01BA	LATIN SMALL LETTER EZH WITH TAIL		
01BB	LATIN LETTER TWO WITH STROKE		
01BC	LATIN CAPITAL LETTER TONE FIVE		
01BD	LATIN SMALL LETTER TONE FIVE		
01BE	LATIN LETTER INVERTED GLOTTAL STOP WITH STROKE		
01BF	LATIN LETTER WYNN		
01C0	LATIN LETTER DENTAL CLICK		
01C1	LATIN LETTER LATERAL CLICK		
01C2	LATIN LETTER ALVEOLAR CLICK		
01C3	LATIN LETTER RETROFLEX CLICK		
01C4	LATIN CAPITAL LETTER DZ WITH CARON		
01C5	LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON		
01C6	LATIN SMALL LETTER DZ WITH CARON		
01C7	LATIN CAPITAL LETTER LJ		
01C8	LATIN CAPITAL LETTER L WITH SMALL LETTER J		
01C9	LATIN SMALL LETTER LJ		
01CA	LATIN CAPITAL LETTER NJ		
01CB	LATIN CAPITAL LETTER N WITH SMALL LETTER J		
01CC	LATIN SMALL LETTER NJ		
01CD	LATIN CAPITAL LETTER A WITH CARON		
01CE	LATIN SMALL LETTER A WITH CARON		
01CF	LATIN CAPITAL LETTER I WITH CARON		
01D0	LATIN SMALL LETTER I WITH CARON		
01D1	LATIN CAPITAL LETTER O WITH CARON		
01D2	LATIN SMALL LETTER O WITH CARON		
01D3	LATIN CAPITAL LETTER U WITH CARON		
01D4	LATIN SMALL LETTER U WITH CARON		
01D5	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON		
01D6	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON		
01D7	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE		
01D8	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE		
01D9	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON		
01DA	LATIN SMALL LETTER U WITH DIAERESIS AND CARON		

Group 00

Plane 00

Table 6 - Rows 01-02: Latin Extended-B

	01F	020	021	022	023	024
0	Ĳ 01F0	Ă 0200	Ř 0210		Ŏ 0230	
1	DZ 01F1	ă 0201	ř 0211		ō 0231	
2	Dz 01F2	Â 0202	Ŕ 0212	8 0222	Ȳ 0232	
3	dz 01F3	â 0203	ŕ 0213	8 0223	ȳ 0233	
4	Ć 01F4	Ě 0204	Ů 0214	Ź 0224		
5	ć 01F5	ě 0205	ů 0215	z 0225		
6	H 01F6	Ê 0206	Û 0216	À 0226		
7	h 01F7	ê 0207	û 0217	à 0227		
8	Ń 01F8	Ĭ 0208	Ş 0218	Ț 0228		
9	ń 01F9	ĭ 0209	ş 0219	ţ 0229		
A	Ą 01FA	Î 020A	Ț 021A	Ȫ 022A		
B	ą 01FB	î 020B	ț 021B	ȫ 022B		
C	Æ 01FC	Ŏ 020C	3 021C	Ȫ 022C		
D	æ 01FD	ò 020D	3 021D	Ȫ 022D		
E	Ø 01FE	Ô 020E	Ĥ 021E	Ò 022E		
F	ø 01FF	ô 020F	ĥ 021F	ò 022F		

G = 00
P = 00

Table 6 - Rows 01-02: Latin Extended-B

HEX	NAME	HEX	NAME
01F0	LATIN SMALL LETTER J WITH CARON	024B	(This position shall not be used)
01F1	LATIN CAPITAL LETTER DZ	024C	(This position shall not be used)
01F2	LATIN CAPITAL LETTER D WITH SMALL LETTER Z	024D	(This position shall not be used)
01F3	LATIN SMALL LETTER DZ	024E	(This position shall not be used)
01F4	LATIN CAPITAL LETTER G WITH ACUTE	024F	(This position shall not be used)
01F5	LATIN SMALL LETTER G WITH ACUTE		
01F6	LATIN CAPITAL LETTER HWAIR		
01F7	LATIN CAPITAL LETTER WYNN		
01F8	LATIN CAPITAL LETTER N WITH GRAVE		
01F9	LATIN SMALL LETTER N WITH GRAVE		
01FA	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE		
01FB	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE		
01FC	LATIN CAPITAL LETTER AE WITH ACUTE (ash) *		
01FD	LATIN SMALL LETTER AE WITH ACUTE (ash) *		
01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE		
01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE		
0200	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE		
0201	LATIN SMALL LETTER A WITH DOUBLE GRAVE		
0202	LATIN CAPITAL LETTER A WITH INVERTED BREVE		
0203	LATIN SMALL LETTER A WITH INVERTED BREVE		
0204	LATIN CAPITAL LETTER E WITH DOUBLE GRAVE		
0205	LATIN SMALL LETTER E WITH DOUBLE GRAVE		
0206	LATIN CAPITAL LETTER E WITH INVERTED BREVE		
0207	LATIN SMALL LETTER E WITH INVERTED BREVE		
0208	LATIN CAPITAL LETTER I WITH DOUBLE GRAVE		
0209	LATIN SMALL LETTER I WITH DOUBLE GRAVE		
020A	LATIN CAPITAL LETTER I WITH INVERTED BREVE		
020B	LATIN SMALL LETTER I WITH INVERTED BREVE		
020C	LATIN CAPITAL LETTER O WITH DOUBLE GRAVE		
020D	LATIN SMALL LETTER O WITH DOUBLE GRAVE		
020E	LATIN CAPITAL LETTER O WITH INVERTED BREVE		
020F	LATIN SMALL LETTER O WITH INVERTED BREVE		
0210	LATIN CAPITAL LETTER R WITH DOUBLE GRAVE		
0211	LATIN SMALL LETTER R WITH DOUBLE GRAVE		
0212	LATIN CAPITAL LETTER R WITH INVERTED BREVE		
0213	LATIN SMALL LETTER R WITH INVERTED BREVE		
0214	LATIN CAPITAL LETTER U WITH DOUBLE GRAVE		
0215	LATIN SMALL LETTER U WITH DOUBLE GRAVE		
0216	LATIN CAPITAL LETTER U WITH INVERTED BREVE		
0217	LATIN SMALL LETTER U WITH INVERTED BREVE		
0218	LATIN CAPITAL LETTER S WITH COMMA BELOW *		
0219	LATIN SMALL LETTER S WITH COMMA BELOW *		
021A	LATIN CAPITAL LETTER T WITH COMMA BELOW *		
021B	LATIN SMALL LETTER T WITH COMMA BELOW *		
021C	LATIN CAPITAL LETTER YOGH		
021D	LATIN SMALL LETTER YOGH		
021E	LATIN CAPITAL LETTER H WITH CARON		
021F	LATIN SMALL LETTER H WITH CARON		
0220	(This position shall not be used)		
0221	(This position shall not be used)		
0222	LATIN CAPITAL LETTER OU		
0223	LATIN SMALL LETTER OU		
0224	LATIN CAPITAL LETTER Z WITH HOOK		
0225	LATIN SMALL LETTER Z WITH HOOK		
0226	LATIN CAPITAL LETTER A WITH DOT ABOVE		
0227	LATIN SMALL LETTER A WITH DOT ABOVE		
0228	LATIN CAPITAL LETTER E WITH CEDILLA		
0229	LATIN SMALL LETTER E WITH CEDILLA		
022A	LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON		
022B	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON		
022C	LATIN CAPITAL LETTER O WITH TILDE AND MACRON		
022D	LATIN SMALL LETTER O WITH TILDE AND MACRON		
022E	LATIN CAPITAL LETTER O WITH DOT ABOVE		
022F	LATIN SMALL LETTER O WITH DOT ABOVE		
0230	LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON		
0231	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON		
0232	LATIN CAPITAL LETTER Y WITH MACRON		
0233	LATIN SMALL LETTER Y WITH MACRON		
0234	(This position shall not be used)		
0235	(This position shall not be used)		
0236	(This position shall not be used)		
0237	(This position shall not be used)		
0238	(This position shall not be used)		
0239	(This position shall not be used)		
023A	(This position shall not be used)		
023B	(This position shall not be used)		
023C	(This position shall not be used)		
023D	(This position shall not be used)		
023E	(This position shall not be used)		
023F	(This position shall not be used)		
0240	(This position shall not be used)		
0241	(This position shall not be used)		
0242	(This position shall not be used)		
0243	(This position shall not be used)		
0244	(This position shall not be used)		
0245	(This position shall not be used)		
0246	(This position shall not be used)		
0247	(This position shall not be used)		
0248	(This position shall not be used)		
0249	(This position shall not be used)		
024A	(This position shall not be used)		

Group 00

Plane 00

Row 01



Documents

Document Sharing Hub

Table 7 - Row 02: IPA Extensions

	025	026	027	028	029	02A
0	ɐ 0250	ɟ 0260	ɥ 0270	ʀ 0280	ʐ 0290	ɕ 02A0
1	ɑ 0251	ɡ 0261	ɱ 0271	ʁ 0281	ʑ 0291	ʑ̥ 02A1
2	ɒ 0252	ɠ 0262	ɲ 0272	ʂ 0282	ʑ̥ 0292	ɕ̥ 02A2
3	ɓ 0253	ɣ 0263	ɳ 0273	ʃ 0283	ʑ̥ 0293	ɕ̥ 02A3
4	ɔ̥ 0254	ɣ̥ 0264	ɴ 0274	ʃ̥ 0284	ʑ̥ 0294	ɕ̥ 02A4
5	ɕ 0255	ɥ̥ 0265	ɵ 0275	ɺ 0285	ɕ̥ 0295	ɕ̥ 02A5
6	ɖ 0256	ɦ 0266	æ̥ 0276	ɹ̥ 0286	ɺ̥ 0296	ɕ̥ 02A6
7	ɖ̥ 0257	ɦ̥ 0267	ω̥ 0277	ɺ̥ 0287	ɕ̥ 0297	ɕ̥ 02A7
8	ə̥ 0258	ɨ̥ 0268	ϕ̥ 0278	ɺ̥ 0288	ɕ̥ 0298	ɕ̥ 02A8
9	ə̥ 0259	ɺ̥ 0269	ɺ̥ 0279	ɺ̥ 0289	ɺ̥ 0299	ɺ̥ 02A9
A	ɺ̥ 025A	ɺ̥ 026A	ɺ̥ 027A	ɺ̥ 028A	ɺ̥ 029A	ɺ̥ 02AA
B	ɺ̥ 025B	ɺ̥ 026B	ɺ̥ 027B	ɺ̥ 028B	ɺ̥ 029B	ɺ̥ 02AB
C	ɺ̥ 025C	ɺ̥ 026C	ɺ̥ 027C	ɺ̥ 028C	ɺ̥ 029C	ɺ̥ 02AC
D	ɺ̥ 025D	ɺ̥ 026D	ɺ̥ 027D	ɺ̥ 028D	ɺ̥ 029D	ɺ̥ 02AD
E	ɺ̥ 025E	ɺ̥ 026E	ɺ̥ 027E	ɺ̥ 028E	ɺ̥ 029E	
F	ɺ̥ 025F	ɺ̥ 026F	ɺ̥ 027F	ɺ̥ 028F	ɺ̥ 029F	

G = 00
P = 00

Table 7 - Row 02: IPA Extensions

HEX	NAME	HEX	NAME
0250	LATIN SMALL LETTER TURNED A	02AB	LATIN SMALL LETTER LZ DIGRAPH
0251	LATIN SMALL LETTER ALPHA	02AC	LATIN LETTER BILABIAL PERCUSSIVE
0252	LATIN SMALL LETTER TURNED ALPHA	02AD	LATIN LETTER BIDENTAL PERCUSSIVE
0253	LATIN SMALL LETTER B WITH HOOK	02AE	(This position shall not be used)
0254	LATIN SMALL LETTER OPEN O	02AF	(This position shall not be used)
0255	LATIN SMALL LETTER C WITH CURL		
0256	LATIN SMALL LETTER D WITH TAIL		
0257	LATIN SMALL LETTER D WITH HOOK		
0258	LATIN SMALL LETTER REVERSED E		
0259	LATIN SMALL LETTER SCHWA		
025A	LATIN SMALL LETTER SCHWA WITH HOOK		
025B	LATIN SMALL LETTER OPEN E		
025C	LATIN SMALL LETTER REVERSED OPEN E		
025D	LATIN SMALL LETTER REVERSED OPEN E WITH HOOK		
025E	LATIN SMALL LETTER CLOSED REVERSED OPEN E		
025F	LATIN SMALL LETTER DOTLESS J WITH STROKE		
0260	LATIN SMALL LETTER G WITH HOOK		
0261	LATIN SMALL LETTER SCRIPT G		
0262	LATIN LETTER SMALL CAPITAL G		
0263	LATIN SMALL LETTER GAMMA		
0264	LATIN SMALL LETTER RAMS HORN		
0265	LATIN SMALL LETTER TURNED H		
0266	LATIN SMALL LETTER H WITH HOOK		
0267	LATIN SMALL LETTER HENG WITH HOOK		
0268	LATIN SMALL LETTER I WITH STROKE		
0269	LATIN SMALL LETTER IOTA		
026A	LATIN LETTER SMALL CAPITAL I		
026B	LATIN SMALL LETTER L WITH MIDDLE TILDE		
026C	LATIN SMALL LETTER L WITH BELT		
026D	LATIN SMALL LETTER L WITH RETROFLEX HOOK		
026E	LATIN SMALL LETTER LEZH		
026F	LATIN SMALL LETTER TURNED M		
0270	LATIN SMALL LETTER TURNED M WITH LONG LEG		
0271	LATIN SMALL LETTER M WITH HOOK		
0272	LATIN SMALL LETTER N WITH LEFT HOOK		
0273	LATIN SMALL LETTER N WITH RETROFLEX HOOK		
0274	LATIN LETTER SMALL CAPITAL N		
0275	LATIN SMALL LETTER BARRED O		
0276	LATIN LETTER SMALL CAPITAL OE		
0277	LATIN SMALL LETTER CLOSED OMEGA		
0278	LATIN SMALL LETTER PHI		
0279	LATIN SMALL LETTER TURNED R		
027A	LATIN SMALL LETTER TURNED R WITH LONG LEG		
027B	LATIN SMALL LETTER TURNED R WITH HOOK		
027C	LATIN SMALL LETTER R WITH LONG LEG		
027D	LATIN SMALL LETTER R WITH TAIL		
027E	LATIN SMALL LETTER R WITH FISHHOOK		
027F	LATIN SMALL LETTER REVERSED R WITH FISHHOOK		
0280	LATIN LETTER SMALL CAPITAL R		
0281	LATIN LETTER SMALL CAPITAL INVERTED R		
0282	LATIN SMALL LETTER S WITH HOOK		
0283	LATIN SMALL LETTER ESH		
0284	LATIN SMALL LETTER DOTLESS J WITH STROKE AND HOOK		
0285	LATIN SMALL LETTER SQUAT REVERSED ESH		
0286	LATIN SMALL LETTER ESH WITH CURL		
0287	LATIN SMALL LETTER TURNED T		
0288	LATIN SMALL LETTER T WITH RETROFLEX HOOK		
0289	LATIN SMALL LETTER U BAR		
028A	LATIN SMALL LETTER UPSILON		
028B	LATIN SMALL LETTER V WITH HOOK		
028C	LATIN SMALL LETTER TURNED V		
028D	LATIN SMALL LETTER TURNED W		
028E	LATIN SMALL LETTER TURNED Y		
028F	LATIN LETTER SMALL CAPITAL Y		
0290	LATIN SMALL LETTER Z WITH RETROFLEX HOOK		
0291	LATIN SMALL LETTER Z WITH CURL		
0292	LATIN SMALL LETTER EZH		
0293	LATIN SMALL LETTER EZH WITH CURL		
0294	LATIN LETTER GLOTTAL STOP		
0295	LATIN LETTER PHARYNGEAL VOICED FRICATIVE		
0296	LATIN LETTER INVERTED GLOTTAL STOP		
0297	LATIN LETTER STRETCHED C		
0298	LATIN LETTER BILABIAL CLICK		
0299	LATIN LETTER SMALL CAPITAL B		
029A	LATIN SMALL LETTER CLOSED OPEN E		
029B	LATIN LETTER SMALL CAPITAL G WITH HOOK		
029C	LATIN LETTER SMALL CAPITAL H		
029D	LATIN SMALL LETTER J WITH CROSSED-TAIL		
029E	LATIN SMALL LETTER TURNED K		
029F	LATIN LETTER SMALL CAPITAL L		
02A0	LATIN SMALL LETTER Q WITH HOOK		
02A1	LATIN LETTER GLOTTAL STOP WITH STROKE		
02A2	LATIN LETTER REVERSED GLOTTAL STOP WITH STROKE		
02A3	LATIN SMALL LETTER DZ DIGRAPH		
02A4	LATIN SMALL LETTER DEZH DIGRAPH		
02A5	LATIN SMALL LETTER DZ DIGRAPH WITH CURL		
02A6	LATIN SMALL LETTER TS DIGRAPH		
02A7	LATIN SMALL LETTER TESH DIGRAPH		
02A8	LATIN SMALL LETTER TC DIGRAPH WITH CURL		
02A9	LATIN SMALL LETTER FENG DIGRAPH		
02AA	LATIN SMALL LETTER LS DIGRAPH		

Group 00

Plane 00

Row 02



Documents

Document Sharing Hub

Table 8 - Row 02: Spacing Modifier Letters

	02B	02C	02D	02E	02F
0	h 02B0	ʔ 02C0	˘ 02D0	Ÿ 02E0	
1	ĥ 02B1	ƚ 02C1	˙ 02D1	ı 02E1	
2	ĵ 02B2	< 02C2	ˊ 02D2	š 02E2	
3	ŗ 02B3	> 02C3	ˋ 02D3	Ẋ 02E3	
4	Ĳ 02B4	Λ 02C4	⊥ 02D4	Ϛ 02E4	
5	Ĵ 02B5	∨ 02C5	⌞ 02D5	⌏ 02E5	
6	Ɓ 02B6	^ 02C6	+ 02D6	⌐ 02E6	
7	Ƶ 02B7	˘ 02C7	- 02D7	⌑ 02E7	
8	Ʒ 02B8	˙ 02C8	˘ 02D8	⌒ 02E8	
9	/ 02B9	- 02C9	˙ 02D9	⌓ 02E9	
A	// 02BA	/ 02CA	◦ 02DA	⌔ 02EA	
B	‘ 02BB	˘ 02CB	˘ 02DB	⌕ 02EB	
C	, 02BC	˘ 02CC	˘ 02DC	˘ 02EC	
D	‘ 02BD	- 02CD	// 02DD	= 02ED	
E	˘ 02BE	˘ 02CE	˘ 02DE	” 02EE	
F	˘ 02BF	˘ 02CF	× 02DF		

G = 00
P = 00

Table 8 - Row 02: Spacing Modifier Letters

HEX	NAME	HEX	NAME
02B0	MODIFIER LETTER SMALL H		
02B1	MODIFIER LETTER SMALL H WITH HOOK		
02B2	MODIFIER LETTER SMALL J		
02B3	MODIFIER LETTER SMALL R		
02B4	MODIFIER LETTER SMALL TURNED R		
02B5	MODIFIER LETTER SMALL TURNED R WITH HOOK		
02B6	MODIFIER LETTER SMALL CAPITAL INVERTED R		
02B7	MODIFIER LETTER SMALL W		
02B8	MODIFIER LETTER SMALL Y		
02B9	MODIFIER LETTER PRIME		
02BA	MODIFIER LETTER DOUBLE PRIME		
02BB	MODIFIER LETTER TURNED COMMA		
02BC	MODIFIER LETTER APOSTROPHE		
02BD	MODIFIER LETTER REVERSED COMMA		
02BE	MODIFIER LETTER RIGHT HALF RING		
02BF	MODIFIER LETTER LEFT HALF RING		
02C0	MODIFIER LETTER GLOTTAL STOP		
02C1	MODIFIER LETTER REVERSED GLOTTAL STOP		
02C2	MODIFIER LETTER LEFT ARROWHEAD		
02C3	MODIFIER LETTER RIGHT ARROWHEAD		
02C4	MODIFIER LETTER UP ARROWHEAD		
02C5	MODIFIER LETTER DOWN ARROWHEAD		
02C6	MODIFIER LETTER CIRCUMFLEX ACCENT		
02C7	CARON (Mandarin Chinese third tone)		
02C8	MODIFIER LETTER VERTICAL LINE		
02C9	MODIFIER LETTER MACRON (Mandarin Chinese first tone)		
02CA	MODIFIER LETTER ACUTE ACCENT (Mandarin Chinese second tone)		
02CB	MODIFIER LETTER GRAVE ACCENT (Mandarin Chinese fourth tone)		
02CC	MODIFIER LETTER LOW VERTICAL LINE		
02CD	MODIFIER LETTER LOW MACRON		
02CE	MODIFIER LETTER LOW GRAVE ACCENT		
02CF	MODIFIER LETTER LOW ACUTE ACCENT		
02D0	MODIFIER LETTER TRIANGULAR COLON		
02D1	MODIFIER LETTER HALF TRIANGULAR COLON		
02D2	MODIFIER LETTER CENTRED RIGHT HALF RING		
02D3	MODIFIER LETTER CENTRED LEFT HALF RING		
02D4	MODIFIER LETTER UP TACK		
02D5	MODIFIER LETTER DOWN TACK		
02D6	MODIFIER LETTER PLUS SIGN		
02D7	MODIFIER LETTER MINUS SIGN		
02D8	BREVE		
02D9	DOT ABOVE (Mandarin Chinese light tone)		
02DA	RING ABOVE		
02DB	OGONEK		
02DC	SMALL TILDE		
02DD	DOUBLE ACUTE ACCENT		
02DE	MODIFIER LETTER RHOTIC HOOK		
02DF	MODIFIER LETTER CROSS ACCENT		
02E0	MODIFIER LETTER SMALL GAMMA		
02E1	MODIFIER LETTER SMALL L		
02E2	MODIFIER LETTER SMALL S		
02E3	MODIFIER LETTER SMALL X		
02E4	MODIFIER LETTER SMALL REVERSED GLOTTAL STOP		
02E5	MODIFIER LETTER EXTRA-HIGH TONE BAR		
02E6	MODIFIER LETTER HIGH TONE BAR		
02E7	MODIFIER LETTER MID TONE BAR		
02E8	MODIFIER LETTER LOW TONE BAR		
02E9	MODIFIER LETTER EXTRA-LOW TONE BAR		
02EA	MODIFIER LETTER YIN DEPARTING TONE MARK		
02EB	MODIFIER LETTER YANG DEPARTING TONE MARK		
02EC	MODIFIER LETTER VOICING		
02ED	MODIFIER LETTER UNASPIRATED		
02EE	MODIFIER LETTER DOUBLE APOSTROPHE		
02EF	(This position shall not be used)		
02F0	(This position shall not be used)		
02F1	(This position shall not be used)		
02F2	(This position shall not be used)		
02F3	(This position shall not be used)		
02F4	(This position shall not be used)		
02F5	(This position shall not be used)		
02F6	(This position shall not be used)		
02F7	(This position shall not be used)		
02F8	(This position shall not be used)		
02F9	(This position shall not be used)		
02FA	(This position shall not be used)		
02FB	(This position shall not be used)		
02FC	(This position shall not be used)		
02FD	(This position shall not be used)		
02FE	(This position shall not be used)		
02FF	(This position shall not be used)		

Group 00

Plane 00

Row 02



Documents

Document Sharing Hub

Table 9 - Row 03: Combining Diacritical Marks

	030	031	032	033	034	035	036
0	 0300	 0310	 0320	 0330	 0340		 0360
1	 0301	 0311	 0321	 0331	 0341		 0361
2	 0302	 0312	 0322	 0332	 0342		 0362
3	 0303	 0313	 0323	 0333	 0343		
4	 0304	 0314	 0324	 0334	 0344		
5	 0305	 0315	 0325	 0335	 0345		
6	 0306	 0316	 0326	 0336	 0346		
7	 0307	 0317	 0327	 0337	 0347		
8	 0308	 0318	 0328	 0338	 0348		
9	 0309	 0319	 0329	 0339	 0349		
A	 030A	 031A	 032A	 033A	 034A		
B	 030B	 031B	 032B	 033B	 034B		
C	 030C	 031C	 032C	 033C	 034C		
D	 030D	 031D	 032D	 033D	 034D		
E	 030E	 031E	 032E	 033E	 034E		
F	 030F	 031F	 032F	 033F			

G = 00
P = 00

Table 9 - Row 03: Combining Diacritical Marks

HEX	NAME	HEX	NAME
0300	COMBINING GRAVE ACCENT (Varia)	035B	(This position shall not be used)
0301	COMBINING ACUTE ACCENT (Oxia)	035C	(This position shall not be used)
0302	COMBINING CIRCUMFLEX ACCENT	035D	(This position shall not be used)
0303	COMBINING TILDE	035E	(This position shall not be used)
0304	COMBINING MACRON	035F	(This position shall not be used)
0305	COMBINING OVERLINE	0360	COMBINING DOUBLE TILDE
0306	COMBINING BREVE (Vrachy)	0361	COMBINING DOUBLE INVERTED BREVE
0307	COMBINING DOT ABOVE	0362	COMBINING DOUBLE RIGHTWARDS ARROW BELOW
0308	COMBINING DIAERESIS (Dialytika)	0363	(This position shall not be used)
0309	COMBINING HOOK ABOVE	0364	(This position shall not be used)
030A	COMBINING RING ABOVE	0365	(This position shall not be used)
030B	COMBINING DOUBLE ACUTE ACCENT	0366	(This position shall not be used)
030C	COMBINING CARON	0367	(This position shall not be used)
030D	COMBINING VERTICAL LINE ABOVE (Tonos)	0368	(This position shall not be used)
030E	COMBINING DOUBLE VERTICAL LINE ABOVE	0369	(This position shall not be used)
030F	COMBINING DOUBLE GRAVE ACCENT	036A	(This position shall not be used)
0310	COMBINING CANDRABINDU	036B	(This position shall not be used)
0311	COMBINING INVERTED BREVE	036C	(This position shall not be used)
0312	COMBINING TURNED COMMA ABOVE	036D	(This position shall not be used)
0313	COMBINING COMMA ABOVE (Psi)	036E	(This position shall not be used)
0314	COMBINING REVERSED COMMA ABOVE (Dasia)	036F	(This position shall not be used)
0315	COMBINING COMMA ABOVE RIGHT		
0316	COMBINING GRAVE ACCENT BELOW		
0317	COMBINING ACUTE ACCENT BELOW		
0318	COMBINING LEFT TACK BELOW		
0319	COMBINING RIGHT TACK BELOW		
031A	COMBINING LEFT ANGLE ABOVE		
031B	COMBINING HORN		
031C	COMBINING LEFT HALF RING BELOW		
031D	COMBINING UP TACK BELOW		
031E	COMBINING DOWN TACK BELOW		
031F	COMBINING PLUS SIGN BELOW		
0320	COMBINING MINUS SIGN BELOW		
0321	COMBINING PALATALIZED HOOK BELOW		
0322	COMBINING RETROFLEX HOOK BELOW		
0323	COMBINING DOT BELOW		
0324	COMBINING DIAERESIS BELOW		
0325	COMBINING RING BELOW		
0326	COMBINING COMMA BELOW		
0327	COMBINING CEDILLA		
0328	COMBINING OGONEK		
0329	COMBINING VERTICAL LINE BELOW		
032A	COMBINING BRIDGE BELOW		
032B	COMBINING INVERTED DOUBLE ARCH BELOW		
032C	COMBINING CARON BELOW		
032D	COMBINING CIRCUMFLEX ACCENT BELOW		
032E	COMBINING BREVE BELOW		
032F	COMBINING INVERTED BREVE BELOW		
0330	COMBINING TILDE BELOW		
0331	COMBINING MACRON BELOW		
0332	COMBINING LOW LINE		
0333	COMBINING DOUBLE LOW LINE		
0334	COMBINING TILDE OVERLAY		
0335	COMBINING SHORT STROKE OVERLAY		
0336	COMBINING LONG STROKE OVERLAY		
0337	COMBINING SHORT SOLIDUS OVERLAY		
0338	COMBINING LONG SOLIDUS OVERLAY		
0339	COMBINING RIGHT HALF RING BELOW		
033A	COMBINING INVERTED BRIDGE BELOW		
033B	COMBINING SQUARE BELOW		
033C	COMBINING SEAGULL BELOW		
033D	COMBINING X ABOVE		
033E	COMBINING VERTICAL TILDE		
033F	COMBINING DOUBLE OVERLINE		
0340	COMBINING GRAVE TONE MARK (Vietnamese)		
0341	COMBINING ACUTE TONE MARK (Vietnamese)		
0342	COMBINING GREEK PERISPOMENI		
0343	COMBINING GREEK KORONIS		
0344	COMBINING GREEK DIALYTIKA TONOS		
0345	COMBINING GREEK YPOGEGRAMMENI		
0346	COMBINING BRIDGE ABOVE		
0347	COMBINING EQUALS SIGN BELOW		
0348	COMBINING DOUBLE VERTICAL LINE BELOW		
0349	COMBINING LEFT ANGLE BELOW		
034A	COMBINING NOT TILDE ABOVE		
034B	COMBINING HOMOTHETIC ABOVE		
034C	COMBINING ALMOST EQUAL TO ABOVE		
034D	COMBINING LEFT RIGHT ARROW BELOW		
034E	COMBINING UPWARDS ARROW BELOW		
034F	(This position shall not be used)		
0350	(This position shall not be used)		
0351	(This position shall not be used)		
0352	(This position shall not be used)		
0353	(This position shall not be used)		
0354	(This position shall not be used)		
0355	(This position shall not be used)		
0356	(This position shall not be used)		
0357	(This position shall not be used)		
0358	(This position shall not be used)		
0359	(This position shall not be used)		
035A	(This position shall not be used)		

Group 00

Plane 00

Table 10 - Row 03: Greek and Coptic

	037	038	039	03A	03B	03C	03D	03E	03F
0			ι̇ 0390	Π 03A0	ϒ̇ 03B0	π 03C0	β 03D0	Ϡ 03E0	κ 03F0
1			Α 0391	Ρ 03A1	α 03B1	ρ 03C1	ϑ 03D1	ϡ 03E1	ϙ 03F1
2			Β 0392		β 03B2	ς 03C2	Υ 03D2	Ϡ 03E2	ϙ 03F2
3			Γ 0393	Σ 03A3	γ 03B3	σ 03C3	Υ̇ 03D3	Ϡ 03E3	ϙ 03F3
4	΄ 0374	΄ 0384	Δ 0394	Τ 03A4	δ 03B4	τ 03C4	Ϛ̇ 03D4	ϣ 03E4	
5	΄ 0375	΄̇ 0385	Ε 0395	Υ 03A5	ε 03B5	υ 03C5	φ 03D5	ϣ 03E5	
6		Α̇ 0386	Ζ 0396	Φ 03A6	ζ 03B6	φ 03C6	ω 03D6	ϣ̇ 03E6	
7		· 0387	Η 0397	Χ 03A7	η 03B7	χ 03C7	ϝ 03D7	ϣ̇ 03E7	
8		Ε̇ 0388	Θ 0398	Ψ 03A8	θ 03B8	ψ 03C8		ϣ̇ 03E8	
9		Η̇ 0389	Ι 0399	Ω 03A9	ι 03B9	ω 03C9		ϣ̇ 03E9	
A	ι̇ 037A	Ι̇ 038A	Κ 039A	Ι̇ 03AA	κ 03BA	ϊ 03CA	Σ 03DA	Χ̇ 03EA	
B			Λ 039B	Υ̇ 03AB	λ 03BB	ϒ̇ 03CB	ς̇ 03DB	ϣ̇ 03EB	
C		Ο̇ 038C	Μ 039C	ά 03AC	μ 03BC	ό 03CC	Φ 03DC	ϣ̇ 03EC	
D			Ν 039D	έ 03AD	ν 03BD	ύ 03CD	Φ̇ 03DD	ϣ̇ 03ED	
E	;	Υ̇ 038E	Ξ 039E	ή 03AE	ξ 03BE	ώ 03CE	ϣ̇ 03DE	ϣ̇ 03EE	
F		Ω̇ 038F	Ο 039F	ί 03AF	ο 03BF		ϣ̇ 03DF	ϣ̇ 03EF	

G = 00
P = 00

Table 10 - Row 03: Greek and Coptic

HEX	NAME	HEX	NAME
0370	(This position shall not be used)	03CB	GREEK SMALL LETTER UPSILON WITH DIALYTIKA
0371	(This position shall not be used)	03CC	GREEK SMALL LETTER OMICRON WITH TONOS
0372	(This position shall not be used)	03CD	GREEK SMALL LETTER UPSILON WITH TONOS
0373	(This position shall not be used)	03CE	GREEK SMALL LETTER OMEGA WITH TONOS
0374	GREEK NUMERAL SIGN (Dexia keraia)	03CF	(This position shall not be used)
0375	GREEK LOWER NUMERAL SIGN (Aristeri keraia)	03D0	GREEK BETA SYMBOL
0376	(This position shall not be used)	03D1	GREEK THETA SYMBOL
0377	(This position shall not be used)	03D2	GREEK UPSILON WITH HOOK SYMBOL
0378	(This position shall not be used)	03D3	GREEK UPSILON WITH ACUTE AND HOOK SYMBOL
0379	(This position shall not be used)	03D4	GREEK UPSILON WITH DIAERESIS AND HOOK SYMBOL
037A	GREEK YPOGEGRAMMENI	03D5	GREEK PHI SYMBOL
037B	(This position shall not be used)	03D6	GREEK PI SYMBOL
037C	(This position shall not be used)	03D7	GREEK KAI SYMBOL
037D	(This position shall not be used)	03D8	(This position shall not be used)
037E	GREEK QUESTION MARK (Eroimatiko)	03D9	(This position shall not be used)
037F	(This position shall not be used)	03DA	GREEK LETTER STIGMA
0380	(This position shall not be used)	03DB	GREEK SMALL LETTER STIGMA
0381	(This position shall not be used)	03DC	GREEK LETTER DIGAMMA
0382	(This position shall not be used)	03DD	GREEK SMALL LETTER DIGAMMA
0383	(This position shall not be used)	03DE	GREEK LETTER KOPPA
0384	GREEK TONOS	03DF	GREEK SMALL LETTER KOPPA
0385	GREEK DIALYTIKA TONOS	03E0	GREEK LETTER SAMPI
0386	GREEK CAPITAL LETTER ALPHA WITH TONOS	03E1	GREEK SMALL LETTER SAMPI
0387	GREEK ANO TELEIA	03E2	COPTIC CAPITAL LETTER SHEI
0388	GREEK CAPITAL LETTER EPSILON WITH TONOS	03E3	COPTIC SMALL LETTER SHEI
0389	GREEK CAPITAL LETTER ETA WITH TONOS	03E4	COPTIC CAPITAL LETTER FEI
038A	GREEK CAPITAL LETTER IOTA WITH TONOS	03E5	COPTIC SMALL LETTER FEI
038B	(This position shall not be used)	03E6	COPTIC CAPITAL LETTER KHEI
038C	GREEK CAPITAL LETTER OMICRON WITH TONOS	03E7	COPTIC SMALL LETTER KHEI
038D	(This position shall not be used)	03E8	COPTIC CAPITAL LETTER HORI
038E	GREEK CAPITAL LETTER UPSILON WITH TONOS	03E9	COPTIC SMALL LETTER HORI
038F	GREEK CAPITAL LETTER OMEGA WITH TONOS	03EA	COPTIC CAPITAL LETTER GANGIA
0390	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS	03EB	COPTIC SMALL LETTER GANGIA
0391	GREEK CAPITAL LETTER ALPHA	03EC	COPTIC CAPITAL LETTER SHIMA
0392	GREEK CAPITAL LETTER BETA	03ED	COPTIC SMALL LETTER SHIMA
0393	GREEK CAPITAL LETTER GAMMA	03EE	COPTIC CAPITAL LETTER DEI
0394	GREEK CAPITAL LETTER DELTA	03EF	COPTIC SMALL LETTER DEI
0395	GREEK CAPITAL LETTER EPSILON	03F0	GREEK KAPPA SYMBOL
0396	GREEK CAPITAL LETTER ZETA	03F1	GREEK RHO SYMBOL
0397	GREEK CAPITAL LETTER ETA	03F2	GREEK LUNATE SIGMA SYMBOL
0398	GREEK CAPITAL LETTER THETA	03F3	GREEK LETTER YOT
0399	GREEK CAPITAL LETTER IOTA	03F4	(This position shall not be used)
039A	GREEK CAPITAL LETTER KAPPA	03F5	(This position shall not be used)
039B	GREEK CAPITAL LETTER LAMDA	03F6	(This position shall not be used)
039C	GREEK CAPITAL LETTER MU	03F7	(This position shall not be used)
039D	GREEK CAPITAL LETTER NU	03F8	(This position shall not be used)
039E	GREEK CAPITAL LETTER XI	03F9	(This position shall not be used)
039F	GREEK CAPITAL LETTER OMICRON	03FA	(This position shall not be used)
03A0	GREEK CAPITAL LETTER PI	03FB	(This position shall not be used)
03A1	GREEK CAPITAL LETTER RHO	03FC	(This position shall not be used)
03A2	(This position shall not be used)	03FD	(This position shall not be used)
03A3	GREEK CAPITAL LETTER SIGMA	03FE	(This position shall not be used)
03A4	GREEK CAPITAL LETTER TAU	03FF	(This position shall not be used)
03A5	GREEK CAPITAL LETTER UPSILON		
03A6	GREEK CAPITAL LETTER PHI		
03A7	GREEK CAPITAL LETTER CHI		
03A8	GREEK CAPITAL LETTER PSI		
03A9	GREEK CAPITAL LETTER OMEGA		
03AA	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA		
03AB	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA		
03AC	GREEK SMALL LETTER ALPHA WITH TONOS		
03AD	GREEK SMALL LETTER EPSILON WITH TONOS		
03AE	GREEK SMALL LETTER ETA WITH TONOS		
03AF	GREEK SMALL LETTER IOTA WITH TONOS		
03B0	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS		
03B1	GREEK SMALL LETTER ALPHA		
03B2	GREEK SMALL LETTER BETA		
03B3	GREEK SMALL LETTER GAMMA		
03B4	GREEK SMALL LETTER DELTA		
03B5	GREEK SMALL LETTER EPSILON		
03B6	GREEK SMALL LETTER ZETA		
03B7	GREEK SMALL LETTER ETA		
03B8	GREEK SMALL LETTER THETA		
03B9	GREEK SMALL LETTER IOTA		
03BA	GREEK SMALL LETTER KAPPA		
03BB	GREEK SMALL LETTER LAMDA		
03BC	GREEK SMALL LETTER MU		
03BD	GREEK SMALL LETTER NU		
03BE	GREEK SMALL LETTER XI		
03BF	GREEK SMALL LETTER OMICRON		
03C0	GREEK SMALL LETTER PI		
03C1	GREEK SMALL LETTER RHO		
03C2	GREEK SMALL LETTER FINAL SIGMA		
03C3	GREEK SMALL LETTER SIGMA		
03C4	GREEK SMALL LETTER TAU		
03C5	GREEK SMALL LETTER UPSILON		
03C6	GREEK SMALL LETTER PHI		
03C7	GREEK SMALL LETTER CHI		
03C8	GREEK SMALL LETTER PSI		
03C9	GREEK SMALL LETTER OMEGA		
03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA		

Group 00

Plane 00

Row 03



Documents

Document Sharing Hub

Table 11 - Row 04: Cyrillic

	040	041	042	043	044	045	046	047
0	È 0400	А 0410	Р 0420	а 0430	р 0440	è 0450	Პ 0460	Ψ 0470
1	Ë 0401	Б 0411	С 0421	б 0431	с 0441	ë 0451	Ჟ 0461	ψ 0471
2	Ლ 0402	В 0412	Т 0422	в 0432	т 0442	Მ 0452	Რ 0462	Θ 0472
3	Ѓ 0403	Г 0413	У 0423	г 0433	у 0443	ѓ 0453	Ѣ 0463	ϑ 0473
4	Є 0404	Д 0414	Ф 0424	д 0434	ф 0444	ё 0454	Ј 0464	ϒ 0474
5	Ѕ 0405	Е 0415	Х 0425	е 0435	х 0445	ѕ 0455	Ѹ 0465	Ϛ 0475
6	Ї 0406	Ж 0416	Ц 0426	ж 0436	ц 0446	ї 0456	Ს 0466	Უ 0476
7	İ 0407	З 0417	Ч 0427	з 0437	ч 0447	ï 0457	Ტ 0467	Ფ 0477
8	Ј 0408	И 0418	Ш 0428	и 0438	ш 0448	ј 0458	Უ 0468	Ქ 0478
9	Љ 0409	Й 0419	Щ 0429	љ 0439	щ 0449	љ 0459	Ფ 0469	Ღ 0479
A	Њ 040A	К 041A	Ъ 042A	к 043A	ъ 044A	њ 045A	Ქ 046A	Ყ 047A
B	Ლ 040B	Л 041B	Ы 042B	л 043B	ы 044B	Მ 045B	Ყ 046B	Ყ 047B
C	Ќ 040C	М 041C	Ь 042C	м 043C	ь 044C	ќ 045C	Ყ 046C	Ყ 047C
D	Й 040D	Н 041D	Э 042D	н 043D	э 044D	й 045D	Ყ 046D	Ყ 047D
E	Ў 040E	О 041E	Ю 042E	о 043E	ю 044E	ў 045E	Ყ 046E	Ყ 047E
F	Ѹ 040F	П 041F	Я 042F	п 043F	я 044F	Ѹ 045F	Ყ 046F	Ყ 047F

G = 00
P = 00

Table 11 - Row 04: Cyrillic

HEX	NAME	HEX	NAME
0400	CYRILLIC CAPITAL LETTER IE WITH GRAVE	045B	CYRILLIC SMALL LETTER TSHE (Serbocroatian)
0401	CYRILLIC CAPITAL LETTER IO	045C	CYRILLIC SMALL LETTER KJE
0402	CYRILLIC CAPITAL LETTER DJE (Serbocroatian)	045D	CYRILLIC SMALL LETTER I WITH GRAVE
0403	CYRILLIC CAPITAL LETTER GJE	045E	CYRILLIC SMALL LETTER SHORT U (Byelorussian)
0404	CYRILLIC CAPITAL LETTER UKRAINIAN IE	045F	CYRILLIC SMALL LETTER DZHE
0405	CYRILLIC CAPITAL LETTER DZE	0460	CYRILLIC CAPITAL LETTER OMEGA
0406	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I	0461	CYRILLIC SMALL LETTER OMEGA
0407	CYRILLIC CAPITAL LETTER YI (Ukrainian)	0462	CYRILLIC CAPITAL LETTER YAT
0408	CYRILLIC CAPITAL LETTER JE	0463	CYRILLIC SMALL LETTER YAT
0409	CYRILLIC CAPITAL LETTER LJE	0464	CYRILLIC CAPITAL LETTER IOTIFIED E
040A	CYRILLIC CAPITAL LETTER NJE	0465	CYRILLIC SMALL LETTER IOTIFIED E
040B	CYRILLIC CAPITAL LETTER TSHE (Serbocroatian)	0466	CYRILLIC CAPITAL LETTER LITTLE YUS
040C	CYRILLIC CAPITAL LETTER KJE	0467	CYRILLIC SMALL LETTER LITTLE YUS
040D	CYRILLIC CAPITAL LETTER I WITH GRAVE	0468	CYRILLIC CAPITAL LETTER IOTIFIED LITTLE YUS
040E	CYRILLIC CAPITAL LETTER SHORT U (Byelorussian)	0469	CYRILLIC SMALL LETTER IOTIFIED LITTLE YUS
040F	CYRILLIC CAPITAL LETTER DZHE	046A	CYRILLIC CAPITAL LETTER BIG YUS
0410	CYRILLIC CAPITAL LETTER A	046B	CYRILLIC SMALL LETTER BIG YUS
0411	CYRILLIC CAPITAL LETTER BE	046C	CYRILLIC CAPITAL LETTER IOTIFIED BIG YUS
0412	CYRILLIC CAPITAL LETTER VE	046D	CYRILLIC SMALL LETTER IOTIFIED BIG YUS
0413	CYRILLIC CAPITAL LETTER GHE	046E	CYRILLIC CAPITAL LETTER KSI
0414	CYRILLIC CAPITAL LETTER DE	046F	CYRILLIC SMALL LETTER KSI
0415	CYRILLIC CAPITAL LETTER IE	0470	CYRILLIC CAPITAL LETTER PSI
0416	CYRILLIC CAPITAL LETTER ZHE	0471	CYRILLIC SMALL LETTER PSI
0417	CYRILLIC CAPITAL LETTER ZE	0472	CYRILLIC CAPITAL LETTER FITA
0418	CYRILLIC CAPITAL LETTER I	0473	CYRILLIC SMALL LETTER FITA
0419	CYRILLIC CAPITAL LETTER SHORT I	0474	CYRILLIC CAPITAL LETTER IZHITSА
041A	CYRILLIC CAPITAL LETTER KA	0475	CYRILLIC SMALL LETTER IZHITSА
041B	CYRILLIC CAPITAL LETTER EL	0476	CYRILLIC CAPITAL LETTER IZHITSА WITH DOUBLE GRAVE ACCENT
041C	CYRILLIC CAPITAL LETTER EM	0477	CYRILLIC SMALL LETTER IZHITSА WITH DOUBLE GRAVE ACCENT
041D	CYRILLIC CAPITAL LETTER EN	0478	CYRILLIC CAPITAL LETTER UK
041E	CYRILLIC CAPITAL LETTER O	0479	CYRILLIC SMALL LETTER UK
041F	CYRILLIC CAPITAL LETTER PE	047A	CYRILLIC CAPITAL LETTER ROUND OMEGA
0420	CYRILLIC CAPITAL LETTER ER	047B	CYRILLIC SMALL LETTER ROUND OMEGA
0421	CYRILLIC CAPITAL LETTER ES	047C	CYRILLIC CAPITAL LETTER OMEGA WITH TITLO
0422	CYRILLIC CAPITAL LETTER TE	047D	CYRILLIC SMALL LETTER OMEGA WITH TITLO
0423	CYRILLIC CAPITAL LETTER U	047E	CYRILLIC CAPITAL LETTER OT
0424	CYRILLIC CAPITAL LETTER EF	047F	CYRILLIC SMALL LETTER OT
0425	CYRILLIC CAPITAL LETTER HA		
0426	CYRILLIC CAPITAL LETTER TSE		
0427	CYRILLIC CAPITAL LETTER CHE		
0428	CYRILLIC CAPITAL LETTER SHA		
0429	CYRILLIC CAPITAL LETTER SHCHA		
042A	CYRILLIC CAPITAL LETTER HARD SIGN		
042B	CYRILLIC CAPITAL LETTER YERU		
042C	CYRILLIC CAPITAL LETTER SOFT SIGN		
042D	CYRILLIC CAPITAL LETTER E		
042E	CYRILLIC CAPITAL LETTER YU		
042F	CYRILLIC CAPITAL LETTER YA		
0430	CYRILLIC SMALL LETTER A		
0431	CYRILLIC SMALL LETTER BE		
0432	CYRILLIC SMALL LETTER VE		
0433	CYRILLIC SMALL LETTER GHE		
0434	CYRILLIC SMALL LETTER DE		
0435	CYRILLIC SMALL LETTER IE		
0436	CYRILLIC SMALL LETTER ZHE		
0437	CYRILLIC SMALL LETTER ZE		
0438	CYRILLIC SMALL LETTER I		
0439	CYRILLIC SMALL LETTER SHORT I		
043A	CYRILLIC SMALL LETTER KA		
043B	CYRILLIC SMALL LETTER EL		
043C	CYRILLIC SMALL LETTER EM		
043D	CYRILLIC SMALL LETTER EN		
043E	CYRILLIC SMALL LETTER O		
043F	CYRILLIC SMALL LETTER PE		
0440	CYRILLIC SMALL LETTER ER		
0441	CYRILLIC SMALL LETTER ES		
0442	CYRILLIC SMALL LETTER TE		
0443	CYRILLIC SMALL LETTER U		
0444	CYRILLIC SMALL LETTER EF		
0445	CYRILLIC SMALL LETTER HA		
0446	CYRILLIC SMALL LETTER TSE		
0447	CYRILLIC SMALL LETTER CHE		
0448	CYRILLIC SMALL LETTER SHA		
0449	CYRILLIC SMALL LETTER SHCHA		
044A	CYRILLIC SMALL LETTER HARD SIGN		
044B	CYRILLIC SMALL LETTER YERU		
044C	CYRILLIC SMALL LETTER SOFT SIGN		
044D	CYRILLIC SMALL LETTER E		
044E	CYRILLIC SMALL LETTER YU		
044F	CYRILLIC SMALL LETTER YA		
0450	CYRILLIC SMALL LETTER IE WITH GRAVE		
0451	CYRILLIC SMALL LETTER IO		
0452	CYRILLIC SMALL LETTER DJE (Serbocroatian)		
0453	CYRILLIC SMALL LETTER GJE		
0454	CYRILLIC SMALL LETTER UKRAINIAN IE		
0455	CYRILLIC SMALL LETTER DZE		
0456	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I		
0457	CYRILLIC SMALL LETTER YI (Ukrainian)		
0458	CYRILLIC SMALL LETTER JE		
0459	CYRILLIC SMALL LETTER LJE		
045A	CYRILLIC SMALL LETTER NJE		

Group 00

Plane 00

Table 12 - Row 04: Cyrillic

	048	049	04A	04B	04C	04D	04E	04F
0	Ѐ 0480	Ґ 0490	Ҁ 04A0	ҁ 04B0	І 04C0	Ӑ 04D0	Ӓ 04E0	Ӕ 04F0
1	҃ 0481	҄ 0491	҅ 04A1	҆ 04B1	҇ 04C1	҈ 04D1	҉ 04E1	Ҋ 04F1
2	ҋ 0482	Ҍ 0492	ҍ 04A2	Ҏ 04B2	ҏ 04C2	Ґ 04D2	ґ 04E2	ҕ 04F2
3	ґ 0483	Җ 0493	җ 04A3	Ҙ 04B3	Ҝ 04C3	Ҟ 04D3	ҟ 04E3	Ҡ 04F3
4	Ҡ 0484	Ң 0494	Ҥ 04A4	ҥ 04B4	Ҧ 04C4	Ҩ 04D4	ҩ 04E4	Ҫ 04F4
5	Ҫ 0485	Ҭ 0495	ҭ 04A5	Ү 04B5		Ұ 04D5	Ҳ 04E5	Ҵ 04F5
6	Ҵ 0486	Ҹ 0496	ҹ 04A6	Һ 04B6		Ҽ 04D6	Ҿ 04E6	
7		Ҽ 0497	ҽ 04A7	ҿ 04B7	Ң 04C7	Ҭ 04D7	Ү 04E7	
8	Ң 0488	Ұ 0498	Ҷ 04A8	Ҹ 04B8	Ҩ 04C8	Ҵ 04D8	Ҷ 04E8	Ҹ 04F8
9	Ҹ 0489	Ҷ 0499	Ҹ 04A9	Ҹ 04B9		Ҷ 04D9	Ҷ 04E9	Ҹ 04F9
A		Ҹ 049A	Ҹ 04AA	Ҹ 04BA		Ҷ 04DA	Ҷ 04EA	
B		Ҹ 049B	Ҹ 04AB	Ҹ 04BB	Ҹ 04CB	Ҷ 04DB	Ҷ 04EB	
C	Ҹ 048C	Ҹ 049C	Ҹ 04AC	Ҹ 04BC	Ҹ 04CC	Ҷ 04DC	Ҷ 04EC	
D	Ҹ 048D	Ҹ 049D	Ҹ 04AD	Ҹ 04BD		Ҷ 04DD	Ҷ 04ED	
E	Ҹ 048E	Ҹ 049E	Ҹ 04AE	Ҹ 04BE		Ҷ 04DE	Ҷ 04EE	
F	Ҹ 048F	Ҹ 049F	Ҹ 04AF	Ҹ 04BF		Ҷ 04DF	Ҷ 04EF	

G = 00
P = 00

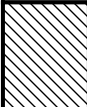
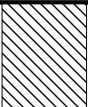
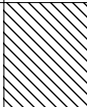
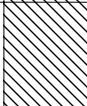
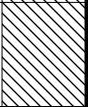
Table 12 - Row 04: Cyrillic

HEX	NAME	HEX	NAME
0480	CYRILLIC CAPITAL LETTER KOPPA	04DB	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS
0481	CYRILLIC SMALL LETTER KOPPA	04DC	CYRILLIC CAPITAL LETTER ZHE WITH DIAERESIS
0482	CYRILLIC THOUSANDS SIGN	04DD	CYRILLIC SMALL LETTER ZHE WITH DIAERESIS
0483	COMBINING CYRILLIC TITLO	04DE	CYRILLIC CAPITAL LETTER ZE WITH DIAERESIS
0484	COMBINING CYRILLIC PALATALIZATION	04DF	CYRILLIC SMALL LETTER ZE WITH DIAERESIS
0485	COMBINING CYRILLIC DASIA PNEUMATA	04E0	CYRILLIC CAPITAL LETTER ABKHASIAN DZE
0486	COMBINING CYRILLIC PSILI PNEUMATA	04E1	CYRILLIC SMALL LETTER ABKHASIAN DZE
0487	(This position shall not be used)	04E2	CYRILLIC CAPITAL LETTER I WITH MACRON
0488	COMBINING CYRILLIC HUNDRED THOUSANDS SIGN	04E3	CYRILLIC SMALL LETTER I WITH MACRON
0489	COMBINING CYRILLIC MILLIONS SIGN	04E4	CYRILLIC CAPITAL LETTER I WITH DIAERESIS
048A	(This position shall not be used)	04E5	CYRILLIC SMALL LETTER I WITH DIAERESIS
048B	(This position shall not be used)	04E6	CYRILLIC CAPITAL LETTER O WITH DIAERESIS
048C	CYRILLIC CAPITAL LETTER SEMISOFT SIGN	04E7	CYRILLIC SMALL LETTER O WITH DIAERESIS
048D	CYRILLIC SMALL LETTER SEMISOFT SIGN	04E8	CYRILLIC CAPITAL LETTER BARRED O
048E	CYRILLIC CAPITAL LETTER ER WITH TICK	04E9	CYRILLIC SMALL LETTER BARRED O
048F	CYRILLIC SMALL LETTER ER WITH TICK	04EA	CYRILLIC CAPITAL LETTER BARRED O WITH DIAERESIS
0490	CYRILLIC CAPITAL LETTER GHE WITH UPTURN	04EB	CYRILLIC SMALL LETTER BARRED O WITH DIAERESIS
0491	CYRILLIC SMALL LETTER GHE WITH UPTURN	04EC	CYRILLIC CAPITAL LETTER E WITH DIAERESIS
0492	CYRILLIC CAPITAL LETTER GHE WITH STROKE	04ED	CYRILLIC SMALL LETTER E WITH DIAERESIS
0493	CYRILLIC SMALL LETTER GHE WITH STROKE	04EE	CYRILLIC CAPITAL LETTER U WITH MACRON
0494	CYRILLIC CAPITAL LETTER GHE WITH MIDDLE HOOK	04EF	CYRILLIC SMALL LETTER U WITH MACRON
0495	CYRILLIC SMALL LETTER GHE WITH MIDDLE HOOK	04F0	CYRILLIC CAPITAL LETTER U WITH DIAERESIS
0496	CYRILLIC CAPITAL LETTER ZHE WITH DESCENDER	04F1	CYRILLIC SMALL LETTER U WITH DIAERESIS
0497	CYRILLIC SMALL LETTER ZHE WITH DESCENDER	04F2	CYRILLIC CAPITAL LETTER U WITH DOUBLE ACUTE
0498	CYRILLIC CAPITAL LETTER ZE WITH DESCENDER	04F3	CYRILLIC SMALL LETTER U WITH DOUBLE ACUTE
0499	CYRILLIC SMALL LETTER ZE WITH DESCENDER	04F4	CYRILLIC CAPITAL LETTER CHE WITH DIAERESIS
049A	CYRILLIC CAPITAL LETTER KA WITH DESCENDER	04F5	CYRILLIC SMALL LETTER CHE WITH DIAERESIS
049B	CYRILLIC SMALL LETTER KA WITH DESCENDER	04F6	(This position shall not be used)
049C	CYRILLIC CAPITAL LETTER KA WITH VERTICAL STROKE	04F7	(This position shall not be used)
049D	CYRILLIC SMALL LETTER KA WITH VERTICAL STROKE	04F8	CYRILLIC CAPITAL LETTER YERU WITH DIAERESIS
049E	CYRILLIC CAPITAL LETTER KA WITH STROKE	04F9	CYRILLIC SMALL LETTER YERU WITH DIAERESIS
049F	CYRILLIC SMALL LETTER KA WITH STROKE	04FA	(This position shall not be used)
04A0	CYRILLIC CAPITAL LETTER BASHKIR KA	04FB	(This position shall not be used)
04A1	CYRILLIC SMALL LETTER BASHKIR KA	04FC	(This position shall not be used)
04A2	CYRILLIC CAPITAL LETTER EN WITH DESCENDER	04FD	(This position shall not be used)
04A3	CYRILLIC SMALL LETTER EN WITH DESCENDER	04FE	(This position shall not be used)
04A4	CYRILLIC CAPITAL LIGATURE EN GHE	04FF	(This position shall not be used)
04A5	CYRILLIC SMALL LIGATURE EN GHE		
04A6	CYRILLIC CAPITAL LETTER PE WITH MIDDLE HOOK (Abkhasian)		
04A7	CYRILLIC SMALL LETTER PE WITH MIDDLE HOOK (Abkhasian)		
04A8	CYRILLIC CAPITAL LETTER ABKHASIAN HA		
04A9	CYRILLIC SMALL LETTER ABKHASIAN HA		
04AA	CYRILLIC CAPITAL LETTER ES WITH DESCENDER		
04AB	CYRILLIC SMALL LETTER ES WITH DESCENDER		
04AC	CYRILLIC CAPITAL LETTER TE WITH DESCENDER		
04AD	CYRILLIC SMALL LETTER TE WITH DESCENDER		
04AE	CYRILLIC CAPITAL LETTER STRAIGHT U		
04AF	CYRILLIC SMALL LETTER STRAIGHT U		
04B0	CYRILLIC CAPITAL LETTER STRAIGHT U WITH STROKE		
04B1	CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE		
04B2	CYRILLIC CAPITAL LETTER HA WITH DESCENDER		
04B3	CYRILLIC SMALL LETTER HA WITH DESCENDER		
04B4	CYRILLIC CAPITAL LIGATURE TE TSE (Abkhasian)		
04B5	CYRILLIC SMALL LIGATURE TE TSE (Abkhasian)		
04B6	CYRILLIC CAPITAL LETTER CHE WITH DESCENDER		
04B7	CYRILLIC SMALL LETTER CHE WITH DESCENDER		
04B8	CYRILLIC CAPITAL LETTER CHE WITH VERTICAL STROKE		
04B9	CYRILLIC SMALL LETTER CHE WITH VERTICAL STROKE		
04BA	CYRILLIC CAPITAL LETTER SHHA		
04BB	CYRILLIC SMALL LETTER SHHA		
04BC	CYRILLIC CAPITAL LETTER ABKHASIAN CHE		
04BD	CYRILLIC SMALL LETTER ABKHASIAN CHE		
04BE	CYRILLIC CAPITAL LETTER ABKHASIAN CHE WITH DESCENDER		
04BF	CYRILLIC SMALL LETTER ABKHASIAN CHE WITH DESCENDER		
04C0	CYRILLIC LETTER PALOCHKA		
04C1	CYRILLIC CAPITAL LETTER ZHE WITH BREVE		
04C2	CYRILLIC SMALL LETTER ZHE WITH BREVE		
04C3	CYRILLIC CAPITAL LETTER KA WITH HOOK		
04C4	CYRILLIC SMALL LETTER KA WITH HOOK		
04C5	(This position shall not be used)		
04C6	(This position shall not be used)		
04C7	CYRILLIC CAPITAL LETTER EN WITH HOOK		
04C8	CYRILLIC SMALL LETTER EN WITH HOOK		
04C9	(This position shall not be used)		
04CA	(This position shall not be used)		
04CB	CYRILLIC CAPITAL LETTER KHAKASSIAN CHE		
04CC	CYRILLIC SMALL LETTER KHAKASSIAN CHE		
04CD	(This position shall not be used)		
04CE	(This position shall not be used)		
04CF	(This position shall not be used)		
04D0	CYRILLIC CAPITAL LETTER A WITH BREVE		
04D1	CYRILLIC SMALL LETTER A WITH BREVE		
04D2	CYRILLIC CAPITAL LETTER A WITH DIAERESIS		
04D3	CYRILLIC SMALL LETTER A WITH DIAERESIS		
04D4	CYRILLIC CAPITAL LIGATURE A IE		
04D5	CYRILLIC SMALL LIGATURE A IE		
04D6	CYRILLIC CAPITAL LETTER IE WITH BREVE		
04D7	CYRILLIC SMALL LETTER IE WITH BREVE		
04D8	CYRILLIC CAPITAL LETTER SCHWA		
04D9	CYRILLIC SMALL LETTER SCHWA		
04DA	CYRILLIC CAPITAL LETTER SCHWA WITH DIAERESIS		

Group 00

Plane 00

Table 13 - Row 05: Armenian

	053	054	055	056	057	058
0		Հ 0540	Ր 0550		Հ 0570	Ր 0580
1	Ա 0531	Զ 0541	Յ 0551	Մ 0561	Ճ 0571	Գ 0581
2	Բ 0532	Ղ 0542	Ի 0552	Ը 0562	Ղ 0572	Լ 0582
3	Գ 0533	Ճ 0543	Փ 0553	Դ 0563	Ճ 0573	Վ 0583
4	Դ 0534	Մ 0544	Ք 0554	Դ 0564	Մ 0574	Ք 0584
5	Ե 0535	Յ 0545	Օ 0555	Է 0565	Ե 0575	Օ 0585
6	Զ 0536	Ն 0546	Ֆ 0556	Դ 0566	Ն 0576	Ֆ 0586
7	Է 0537	Շ 0547		Է 0567	Ն 0577	Լ 0587
8	Ը 0538	Ո 0548		Ը 0568	Ո 0578	
9	Թ 0539	Չ 0549	՛ 0559	Թ 0569	Ն 0579	՝ 0589
A	Ժ 053A	Պ 054A	՛ 055A	Ժ 056A	Մ 057A	՛ 058A
B	Ի 053B	Ջ 054B	՛ 055B	Ի 056B	Ջ 057B	
C	Լ 053C	Խ 054C	~ 055C	Լ 056C	Խ 057C	
D	Խ 053D	Ս 054D	՛ 055D	Խ 056D	Ս 057D	
E	Շ 053E	Վ 054E	՛ 055E	Շ 056E	Վ 057E	
F	Կ 053F	Տ 054F	՛ 055F	Կ 056F	Տ 057F	

G = 00
P = 00

Table 13 - Row 05: Armenian

HEX	NAME	HEX	NAME
0530	(This position shall not be used)	058B	(This position shall not be used)
0531	ARMENIAN CAPITAL LETTER AYB	058C	(This position shall not be used)
0532	ARMENIAN CAPITAL LETTER BEN	058D	(This position shall not be used)
0533	ARMENIAN CAPITAL LETTER GIM	058E	(This position shall not be used)
0534	ARMENIAN CAPITAL LETTER DA	058F	(This position shall not be used)
0535	ARMENIAN CAPITAL LETTER ECH		
0536	ARMENIAN CAPITAL LETTER ZA		
0537	ARMENIAN CAPITAL LETTER EH		
0538	ARMENIAN CAPITAL LETTER ET		
0539	ARMENIAN CAPITAL LETTER TO		
053A	ARMENIAN CAPITAL LETTER ZHE		
053B	ARMENIAN CAPITAL LETTER INI		
053C	ARMENIAN CAPITAL LETTER LIWN		
053D	ARMENIAN CAPITAL LETTER XEH		
053E	ARMENIAN CAPITAL LETTER CA		
053F	ARMENIAN CAPITAL LETTER KEN		
0540	ARMENIAN CAPITAL LETTER HO		
0541	ARMENIAN CAPITAL LETTER JA		
0542	ARMENIAN CAPITAL LETTER GHAD		
0543	ARMENIAN CAPITAL LETTER CHEH		
0544	ARMENIAN CAPITAL LETTER MEN		
0545	ARMENIAN CAPITAL LETTER YI		
0546	ARMENIAN CAPITAL LETTER NOW		
0547	ARMENIAN CAPITAL LETTER SHA		
0548	ARMENIAN CAPITAL LETTER VO		
0549	ARMENIAN CAPITAL LETTER CHA		
054A	ARMENIAN CAPITAL LETTER PEH		
054B	ARMENIAN CAPITAL LETTER JHEH		
054C	ARMENIAN CAPITAL LETTER RA		
054D	ARMENIAN CAPITAL LETTER SEH		
054E	ARMENIAN CAPITAL LETTER VEW		
054F	ARMENIAN CAPITAL LETTER TIWN		
0550	ARMENIAN CAPITAL LETTER REH		
0551	ARMENIAN CAPITAL LETTER CO		
0552	ARMENIAN CAPITAL LETTER YIWN		
0553	ARMENIAN CAPITAL LETTER PIWR		
0554	ARMENIAN CAPITAL LETTER KEH		
0555	ARMENIAN CAPITAL LETTER OH		
0556	ARMENIAN CAPITAL LETTER FEH		
0557	(This position shall not be used)		
0558	(This position shall not be used)		
0559	ARMENIAN MODIFIER LETTER LEFT HALF RING		
055A	ARMENIAN APOSTROPHE		
055B	ARMENIAN EMPHASIS MARK		
055C	ARMENIAN EXCLAMATION MARK		
055D	ARMENIAN COMMA		
055E	ARMENIAN QUESTION MARK		
055F	ARMENIAN ABBREVIATION MARK		
0560	(This position shall not be used)		
0561	ARMENIAN SMALL LETTER AYB		
0562	ARMENIAN SMALL LETTER BEN		
0563	ARMENIAN SMALL LETTER GIM		
0564	ARMENIAN SMALL LETTER DA		
0565	ARMENIAN SMALL LETTER ECH		
0566	ARMENIAN SMALL LETTER ZA		
0567	ARMENIAN SMALL LETTER EH		
0568	ARMENIAN SMALL LETTER ET		
0569	ARMENIAN SMALL LETTER TO		
056A	ARMENIAN SMALL LETTER ZHE		
056B	ARMENIAN SMALL LETTER INI		
056C	ARMENIAN SMALL LETTER LIWN		
056D	ARMENIAN SMALL LETTER XEH		
056E	ARMENIAN SMALL LETTER CA		
056F	ARMENIAN SMALL LETTER KEN		
0570	ARMENIAN SMALL LETTER HO		
0571	ARMENIAN SMALL LETTER JA		
0572	ARMENIAN SMALL LETTER GHAD		
0573	ARMENIAN SMALL LETTER CHEH		
0574	ARMENIAN SMALL LETTER MEN		
0575	ARMENIAN SMALL LETTER YI		
0576	ARMENIAN SMALL LETTER NOW		
0577	ARMENIAN SMALL LETTER SHA		
0578	ARMENIAN SMALL LETTER VO		
0579	ARMENIAN SMALL LETTER CHA		
057A	ARMENIAN SMALL LETTER PEH		
057B	ARMENIAN SMALL LETTER JHEH		
057C	ARMENIAN SMALL LETTER RA		
057D	ARMENIAN SMALL LETTER SEH		
057E	ARMENIAN SMALL LETTER VEW		
057F	ARMENIAN SMALL LETTER TIWN		
0580	ARMENIAN SMALL LETTER REH		
0581	ARMENIAN SMALL LETTER CO		
0582	ARMENIAN SMALL LETTER YIWN		
0583	ARMENIAN SMALL LETTER PIWR		
0584	ARMENIAN SMALL LETTER KEH		
0585	ARMENIAN SMALL LETTER OH		
0586	ARMENIAN SMALL LETTER FEH		
0587	ARMENIAN SMALL LIGATURE ECH YIWN		
0588	(This position shall not be used)		
0589	ARMENIAN FULL STOP		
058A	ARMENIAN HYPHEN		

Group 00

Plane 00

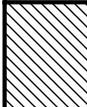
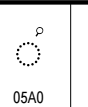
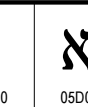
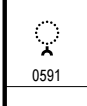
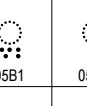
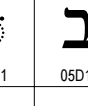
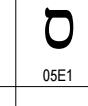
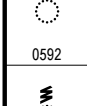
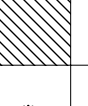
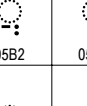
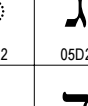
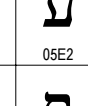
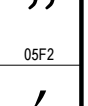
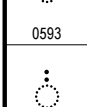
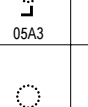
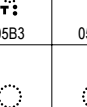
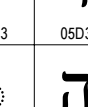
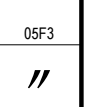
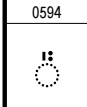
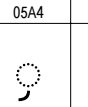
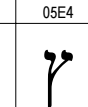
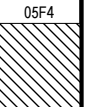
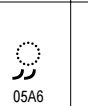
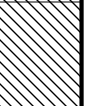
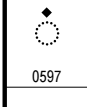
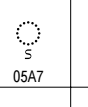
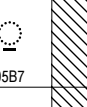
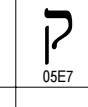
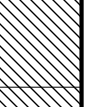
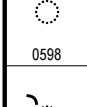
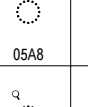
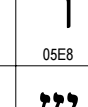
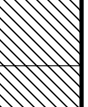
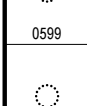
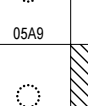
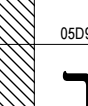
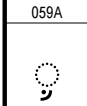
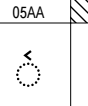
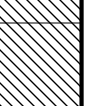
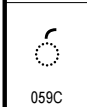
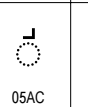
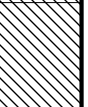
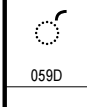
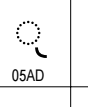
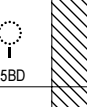
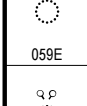
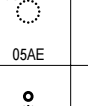
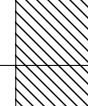
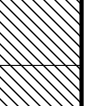
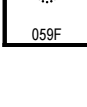
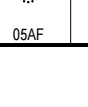
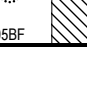
Row 05



Documents

Document Sharing Hub

Table 14 - Row 05: Hebrew

	059	05A	05B	05C	05D	05E	05F
0							
1							
2							
3							
4							
5							
6							
7							
8							
9							
A							
B							
C							
D							
E							
F							

G = 00
P = 00