

# The Unicode Standard

## Version 8.0 – Core Specification

To learn about the latest version of the Unicode Standard, see <http://www.unicode.org/versions/latest/>.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

The authors and publisher have taken care in the preparation of this specification, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

Copyright © 1991–2015 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction. For information regarding permissions, inquire at <http://www.unicode.org/reporting.html>. For information about the Unicode terms of use, please see <http://www.unicode.org/copyright.html>.

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 8.0

Includes bibliographical references and index.

ISBN 978-1-936213-10-8 (<http://www.unicode.org/versions/Unicode8.0.0/>)

1. Unicode (Computer character set) I. Allen, Julie D. II. Unicode Consortium.

QA268.U545 2015

ISBN 978-1-936213-10-8

Published in Mountain View, CA

August 2015

## Appendix D

# *Changes from Previous Versions*

This appendix provides version history of the standard and summarizes updates that have been made to conformance specifications, character content, and data files in the Unicode Character Database since the publication of *The Unicode Standard, Version 6.0*.

## D.1 Versions of the Unicode Standard

The Unicode Technical Committee updates the Unicode Standard to respond to the needs of implementers and users while maintaining consistency with ISO/IEC 10646. The relationship between these versions of Unicode and ISO/IEC 10646 is shown in *Table D-1*. For more detail on the relationship of Unicode and ISO/IEC 10646, see *Appendix C, Relationship to ISO/IEC 10646*.

**Table D-1.** Versions of Unicode and ISO/IEC 10646

Year	Version	Published	ISO/IEC 10646
1991	Unicode 1.0	Vol. 1, Addison-Wesley	Basis for Committee Draft 2 of 10646-1
1992	Unicode 1.0.1	Vol. 1, 2, Addison-Wesley	Interim merger version
1993	Unicode 1.1	UTR #4	Matches 10646-1
1996	Unicode 2.0	Addison-Wesley	Matches 10646-1 plus amendments
1998	Unicode 2.1	UTR #8	Matches 10646-1 plus amendments
2000	Unicode 3.0	Addison-Wesley	Matches 10646-1 second edition
2001	Unicode 3.1	UAX #27	Matches 10646-1 second edition plus two characters, 10646-2 first edition
2002	Unicode 3.2	UAX #28	Matches 10646-1 second edition plus amendment, 10646-2 first edition
2003	Unicode 4.0	Addison-Wesley	Matches 10646:2003, third version
2005	Unicode 4.1	Web publication	Matches 10646:2003, third version, plus Amd. 1
2006	Unicode 5.0	Addison-Wesley (2007)	Matches 10646:2003, third version, plus Amd. 1, Amd. 2, and four characters from Amd. 3
2008	Unicode 5.1	Web publication	Matches 10646:2003, third version, plus Amd. 1 through Amd. 4
2009	Unicode 5.2	Web publication	Matches 10646:2003, third version, plus Amd. 1 through Amd. 6
2010	Unicode 6.0	Web publication	Matches 10646:2011, second edition
2012	Unicode 6.1	Web publication	Matches 10646:2012, third edition
2012	Unicode 6.2	Web publication	Matches 10646:2012, third edition
2013	Unicode 6.3	Web publication	Matches 10646:2012, third edition
2014	Unicode 7.0	Web publication	Matches 10646:2012, third edition, plus Amd. 1 and Amd. 2
2015	Unicode 8.0	Web publication	Matches 10646:2014, fourth edition, plus Amd. 1, and 51 characters from Amd. 2

*Table D-2*, *Table D-3*, *Table D-4*, and *Table D-5* document the number of code points allocated in the different versions of the Unicode Standard. For an explanation of the types of characters which count as *graphic* or *format* in these summary statistics, see *Table 2-3, Types of Code Points*. In each table, the row labeled “Graphic + Format” represents the traditional count of Unicode characters and is the typical answer to the question, “How many charac-

ters are in the Unicode Standard?” The numbers cited in the row labeled “Han Compatibility” in each table include the 12 unified CJK ideographs encoded in the CJK Compatibility Ideographs block.

*Table D-2* lists the allocation of code points by type for the earliest, historic versions of the Unicode Standard up to the publication of Version 3.0. In some cases the values in *Table D-2* differ slightly from summary statistics published in those very early versions of the standard, primarily due to a refined accounting of the allocations in Unicode 1.0.

**Table D-2.** Allocation of Code Points by Type (Versions 1.0.0 to 3.0)

	V1.0.0	V1.0.1	V1.1	V2.0	V2.1	V3.0
Alphabets, Symbols	4,734	4,728	6,290	6,491	6,493	10,210
Han (URO)		20,902	20,902	20,902	20,902	20,902
Han Extension A						6,582
Han Compatibility		302	302	302	302	302
Subtotal Han		21,204	21,204	21,204	21,204	27,786
Hangul Syllables	2,350	2,350	6,656	11,172	11,172	11,172
Graphic Characters	7,084	28,282	34,150	38,867	38,869	49,168
Format Characters	12	12	18	18	18	26
Graphic + Format	7,096	28,294	34,168	38,885	38,887	49,194
Controls	65	65	65	65	65	65
Private Use	5,632	6,144	6,400	137,468	137,468	137,468
Total Assigned	12,793	34,503	40,633	176,418	176,420	186,727
Surrogate Code Points				2,048	2,048	2,048
Noncharacters	2	2	2	34	34	34
Total Designated	12,795	34,505	40,635	178,500	178,502	188,809
Reserved Code Points	52,741	31,031	24,901	935,612	935,610	925,303

Table D-3 lists the allocation of code points by type for intermediate versions of the Unicode Standard, up to the publication of Version 5.1.

**Table D-3.** Allocation of Code Points by Type (Versions 3.1 to 5.1)

	V3.1	V3.2	V4.0	V4.1	V5.0	V5.1
Alphabets, Symbols	11,798	12,753	13,973	15,117	16,486	18,101
Han (URO)	20,902	20,902	20,902	20,902	20,902	20,902
Han (URO Extension)				22	22	30
Han Extension A	6,582	6,582	6,582	6,582	6,582	6,582
Han Extension B	42,711	42,711	42,711	42,711	42,711	42,711
Han Compatibility	844	903	903	1,009	1,009	1,009
Subtotal Han	71,039	71,098	71,098	71,226	71,226	71,234
Hangul Syllables	11,172	11,172	11,172	11,172	11,172	11,172
Graphic Characters	94,009	95,023	96,243	97,515	98,884	100,507
Format Characters	131	133	139	140	140	141
Graphic + Format	94,140	95,156	96,382	97,655	99,024	100,648
Controls	65	65	65	65	65	65
Private Use	137,468	137,468	137,468	137,468	137,468	137,468
Total Assigned	231,673	232,689	233,915	235,188	236,557	238,181
Surrogate Code Points	2,048	2,048	2,048	2,048	2,048	2,048
Noncharacters	66	66	66	66	66	66
Total Designated	233,787	234,803	236,029	237,302	238,671	240,295
Reserved Code Points	880,325	879,309	878,083	876,810	875,441	873,817

Table D-4 lists the allocation of code points for more recent versions of the Unicode Standard, up to the publication of Version 7.0.

**Table D-4.** Allocation of Code Points by Type (Versions 5.2 to 7.0)

	V5.2	V6.0	V6.1	V6.2	V6.3	V7.0
Alphabets, Symbols	20,588	22,454	23,184	23,185	23,184	26,013
Han (URO)	20,902	20,902	20,902	20,902	20,902	20,902
Han (URO Extension)	38	38	39	39	39	39
Han Extension A	6,582	6,582	6,582	6,582	6,582	6,582
Han Extension B	42,711	42,711	42,711	42,711	42,711	42,711
Han Extension C	4,149	4,149	4,149	4,149	4,149	4,149
Han Extension D		222	222	222	222	222
Han Compatibility	1,012	1,012	1,014	1,014	1,014	1,014
Subtotal Han	75,394	75,616	75,619	75,619	75,619	75,619
Hangul Syllables	11,172	11,172	11,172	11,172	11,172	11,172
Graphic Characters	107,154	109,242	109,975	109,976	109,975	112,804
Format Characters	142	142	141	141	147	152
Graphic + Format	107,296	109,384	110,116	110,117	110,122	112,956
Controls	65	65	65	65	65	65
Private Use	137,468	137,468	137,468	137,468	137,468	137,468
Total Assigned	244,829	246,917	247,649	247,650	247,655	250,489
Surrogate Code Points	2,048	2,048	2,048	2,048	2,048	2,048
Noncharacters	66	66	66	66	66	66
Total Designated	246,943	249,031	249,763	249,764	249,769	252,603
Reserved Code Points	867,169	865,081	864,349	864,348	864,343	861,509

Table D-5 lists the allocation of code points for Version 8.0 of the Unicode Standard.

**Table D-5. Allocation of Code Points by Type (Version 8.0)**

	<b>V8.0</b>
Alphabets, Symbols	27,958
Han (URO)	20,902
Han (URO Extension)	48
Han Extension A	6,582
Han Extension B	42,711
Han Extension C	4,149
Han Extension D	222
Han Extension E	5,762
Han Compatibility	1,014
Subtotal Han	81,390
Hangul Syllables	11,172
Graphic Characters	120,520
Format Characters	152
Graphic + Format	120,672
Controls	65
Private Use	137,468
Total Assigned	258,205
Surrogate Code Points	2,048
Noncharacters	66
Total Designated	260,319
Reserved Code Points	853,793

## D.2 Clause and Definition Updates

Version 7.0 of the Unicode Standard clarified a number of definitions subsequent to the publication of Version 6.0.

*Table D-6* provides a list of all clauses and definitions that were clarified, changed, or newly added in Version 6.1.

**Table D-6.** Version 6.1 Clause and Definition Updates

Number	Clause or Definition	Type of Update
D58	Grapheme base	Clarification
D59	Grapheme extender	Clarification
D60	Grapheme cluster	Clarification
D107	Starter	Clarification
D110	Singleton decomposition	Clarification

There are no changes to any clauses and no updates to definitions in Version 6.2.

*Table D-6* provides a list of all clauses and definitions that were clarified, changed, or newly added in Version 6.3.

**Table D-7.** Version 6.3 Clause and Definition Updates

Number	Clause or Definition	Type of Update
D14	Noncharacter	Clarification
D136	Case-ignorable	Clarification

The clarifications to definitions D14 and D136 that were made in Version 6.3 were incorporated into the core specification in Version 7.0.



