

ISO/IEC 10646:2011-03 (E)

Information technology - Universal Coded Character Set (UCS)

Contents		Page
Foreword		7
Introduction		8
1	Scope	9
2	Conformance	9
2.1	General	9
2.2	Conformance of information interchange	9
2.3	Conformance of devices	10
3	Normative references	10
4	Terms and definitions	11
5	General structure of the UCS	17
6	Basic structure and nomenclature	17
6.1	Structure	17
6.2	Coding of characters	19
6.3	Types of code points	19
6.4	Naming of characters	20
6.5	Short identifiers for code points (UIDs)	20
6.6	UCS Sequence Identifiers	21
6.7	Octet sequence identifiers	21
7	Revision and updating of the UCS	22
8	Subsets	22
8.1	Limited subset	22
8.2	Selected subset	22
9	UCS encoding forms	22
9.1	UTF-8	22
9.2	UTF-16	23
9.3	UTF-32 (UCS-4)	24
10	UCS Encoding schemes	24
10.1	UTF-8	24
10.2	UTF-16BE	24
10.3	UTF-16LE	24
10.4	UTF-16	24
10.5	UTF-32BE	25
10.6	UTF-32LE	25
10.7	UTF-32	25
11	Use of control functions with the UCS	25
12	Declaration of identification of features	26
12.1	Purpose and context of identification	26
12.2	Identification of a UCS encoding form	27
12.3	Identification of subsets of graphic characters	27
12.4	Identification of control function set	27

13	Structure of the code charts and lists	28
14	Block and collection names	29
14.1	Block names	29
14.2	Collection names	29
15	Mirrored characters in bidirectional context	29
15.1	Mirrored characters	29
15.2	Directionality of bidirectional text	29
16	Special characters	30
16.1	Space characters	30
16.2	Currency symbols	30
16.3	Format Characters	30
16.4	Ideographic description characters	31
16.5	Variation selectors and variation sequences	31
17	Presentation forms of characters	34
18	Compatibility characters	34
19	Order of characters	34
20	Combining characters	35
20.1	Order of combining characters	35
20.2	Combining class and canonical ordering	35
20.3	Appearance in code charts	35
20.4	Alternate coded representations	35
20.5	Multiple combining characters	35
20.6	Collections containing combining characters	36
20.7	Combining Grapheme Joiner	36
21	Normalization forms	36
22	Special features of individual scripts and symbol repertoires	37
22.1	Hangul syllable composition method	37
22.2	Features of scripts used in India and some other South Asian countries	37
22.3	Byzantine musical symbols	38
22.4	Source references for pictographic symbols	38
23	Source references for CJK Ideographs	38
23.1	List of source references	38
23.2	Source references for CJK Unified Ideographs	41
23.3	Source reference presentation for CJK Unified Ideographs	42
23.4	Source references for CJK Compatibility Ideographs	43
23.5	Source references presentation for CJK Compatibility Ideographs	44
24	Character names and annotations	44
24.1	Entity names	44
24.2	Name formation	44
24.3	Single name	45
24.4	Name immutability	45
24.5	Name uniqueness	45
24.6	Character names for CJK Ideographs	46
24.7	Character names for Hangul syllables	46

25	Named UCS Sequence Identifiers	48
26	Structure of the Basic Multilingual Plane	50
27	Structure of the Supplementary Multilingual Plane for scripts and symbols (SMP)	52
28	Structure of the Supplementary Ideographic Plane (SIP)	54
29	Structure of the Tertiary Ideographic Plane (TIP)	54
30	Structure of the Supplementary Special-purpose Plane (SSP)	54
31	Code charts and lists of character names	55
31.1	Code chart	55
31.2	Character names list	55
31.3	Pointers to code charts and lists of character names	56
Annex A (normative) Collections of graphic characters for subsets		1599
A.1	Collections of coded graphic characters	1599
A.2	Blocks lists	1604
A.3	Fixed collections of the whole UCS (except Unicode collections)	1606
A.4	CJK collections	1609
A.5	Other collections	1610
A.6	Unicode collections	1613
Annex B (normative) List of combining characters		1622
Annex C (normative) Transformation format for planes 01 to 10 of the UCS (UTF-16)		1623
Annex D (normative) UCS Transformation Format 8 (UTF-8)		1624
Annex E (normative) Mirrored characters in bidirectional context		1625
Annex F (informative) Format characters		1626
F.1	General format characters	1626
F.2	Script-specific format characters	1628
F.3	Interlinear annotation characters	1629
F.4	Subtending format characters	1629
F.5	Contiguity operators	1629
F.6	Western musical symbols	1630
F.7	Language tagging using Tag characters	1631
Annex G (informative) Alphabetically sorted list of character names		1633
Annex H (informative) The use of signatures to identify UCS		1634
Annex I (informative) Ideographic description characters		1635
Annex J (informative) Recommendation for combined receiving/originating devices with internal storage		1638
Annex K (informative) Notations of octet value representations		1639
Annex L (informative) Character naming guidelines		1640
Annex M (informative) Sources of characters		1643
Annex N (informative) External references to character repertoires		1660

N.1	Methods of reference to character repertoires and their coding	1660
N.2	Identification of ASN.1 character abstract syntaxes	1660
N.3	Identification of ASN.1 character transfer syntaxes	1661
Annex P (informative) Additional information on CJK Unified Ideographs		1662
Annex Q (informative) Code mapping table for Hangul syllables		1663
Annex R (informative) Names of Hangul syllables		1664
Annex S (informative) Procedure for the unification and arrangement of CJK Ideographs		1665
S.1	Unification procedure	1665
S.2	Arrangement procedure	1669
S.3	Source code separation examples	1669
S.4	Non-unification examples	1674
Annex T (informative) Language tagging using Tag Characters		1676
Annex U (informative) Characters in identifiers		1677