



SLOVENSKI STANDARD
oSIST ISO/IEC 10646:2013
01-marec-2013

Informacijska tehnologija - Univerzalni večoktetni nabor znakov (UCS)

Information technology -- Universal Coded Character Set (UCS)

Technologies de l'information -- Jeu universel de caractères codés (JUC)

Ta slovenski standard je istoveten z: ISO/IEC 10646:2012

ICS:

35.040	Nabori znakov in kodiranje informacij	Character sets and information coding
--------	---------------------------------------	---------------------------------------

oSIST ISO/IEC 10646:2013 **en,fr,de**

INTERNATIONAL
STANDARD

ISO/IEC
10646

Third edition
2012-06-01

**Information technology — Universal
Coded Character Set (UCS)**

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

Reference number
ISO/IEC 10646:2012(E)





COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2012

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.org
Web www.iso.org

Published in Switzerland

CONTENTS

Foreword.....	vii
Introduction.....	viii
1 Scope	1
2 Conformance.....	1
2.1 General	1
2.2 Conformance of information interchange.....	2
2.3 Conformance of devices	2
3 Normative references	2
4 Terms and definitions	3
5 General structure of the UCS.....	9
6 Basic structure and nomenclature.....	9
6.1 Structure.....	9
6.2 Coding of characters.....	11
6.3 Types of code points.....	11
6.4 Naming of characters.....	12
6.5 Short identifiers for code points (UIDs).....	12
6.6 UCS Sequence Identifiers.....	13
6.7 Octet sequence identifiers	13
7 Revision and updating of the UCS	14
8 Subsets.....	14
8.1 Limited subset.....	14
8.2 Selected subset.....	14
9 UCS encoding forms	14
9.1 UTF-8.....	14
9.2 UTF-16.....	15
9.3 UTF-32 (UCS-4).....	16
10 UCS Encoding schemes	16
10.1 UTF-8.....	16
10.2 UTF-16BE	16
10.3 UTF-16LE.....	16
10.4 UTF-16.....	16
10.5 UTF-32BE	17
10.6 UTF-32LE.....	17
10.7 UTF-32.....	17
11 Use of control functions with the UCS.....	17
12 Declaration of identification of features	18
12.1 Purpose and context of identification	18
12.2 Identification of a UCS encoding form	19
12.3 Identification of subsets of graphic characters.....	19
12.4 Identification of control function set.....	19
12.5 Identification of the coding system of ISO/IEC 2022	20
13 Structure of the code charts and lists.....	20

ISO/IEC 10646:2012 (E)

14	Block and collection names.....	21
14.1	Block names.....	21
14.2	Collection names.....	21
15	Mirrored characters in bidirectional context	21
15.1	Mirrored characters	21
15.2	Directionality of bidirectional text	21
16	Special characters.....	22
16.1	Space characters	22
16.2	Currency symbols	22
16.3	Format Characters	22
16.4	Ideographic description characters	23
16.5	Variation selectors and variation sequences	23
17	Presentation forms of characters	26
18	Compatibility characters	26
19	Order of characters	26
20	Combining characters	27
20.1	Order of combining characters.....	27
20.2	Combining class and canonical ordering	27
20.3	Appearance in code charts	27
20.4	Alternate coded representations	27
20.5	Multiple combining characters	27
20.6	Collections containing combining characters.....	28
20.7	Combining Grapheme Joiner	28
21	Normalization forms	29
22	Special features of individual scripts and symbol repertoires	29
22.1	Hangul syllable composition method	29
22.2	Features of scripts used in India and some other South Asian countries.....	29
22.3	Byzantine musical symbols.....	30
22.4	Source references for pictographic symbols.....	30
23	Source references for CJK Ideographs.....	31
23.1	List of source references.....	31
23.2	Source references for CJK Unified Ideographs	33
23.3	Source reference presentation for CJK Unified Ideographs	34
23.4	Source references for CJK Compatibility Ideographs.....	36
23.5	Source references presentation for CJK Compatibility Ideographs.....	37
24	Character names and annotations	38
24.1	Entity names	38
24.2	Name formation.....	38
24.3	Single name	38
24.4	Name immutability.....	39
24.5	Name uniqueness	39
24.6	Character names for CJK Ideographs	40
24.7	Character names for Hangul syllables	40
25	Named UCS Sequence Identifiers	41

26	Structure of the Basic Multilingual Plane	43
27	Structure of the Supplementary Multilingual Plane for scripts and symbols (SMP).....	45
28	Structure of the Supplementary Ideographic Plane (SIP)	47
29	Structure of the Tertiary Ideographic Plane (TIP)	47
30	Structure of the Supplementary Special-purpose Plane (SSP)	47
31	Code charts and lists of character names	48
31.1	Code chart.....	48
31.2	Character names list	48
31.3	Pointers to code charts and lists of character names	49
Annex A	(normative) Collections of graphic characters for subsets	2181
A.1	Collections of coded graphic characters	2181
A.2	Blocks lists	2186
A.3	Fixed collections of the whole UCS (except Unicode collections).....	2188
A.4	CJK collections.....	2192
A.5	Other collections	2193
A.6	Unicode collections	2196
Annex B	(normative) List of combining characters.....	2207
Annex C	(normative) Transformation format for planes 01 to 10 of the UCS (UTF-16).....	2208
Annex D	(normative) UCS Transformation Format 8 (UTF-8)	2209
Annex E	(normative) Mirrored characters in bidirectional context.....	2210
Annex F	(informative) Format characters	2211
F.1	General format characters	2211
F.2	Script-specific format characters.....	2213
F.3	Interlinear annotation characters	2214
F.4	Subtending format characters	2214
F.5	Contiguity operators	2214
F.6	Western musical symbols	2215
F.7	Language tagging using Tag characters	2216
Annex G	(informative) Alphabetically sorted list of character names.....	2218
Annex H	(informative) The use of “signatures” to identify UCS	2219
Annex I	(informative) Ideographic description characters	2220
Annex J	(informative) Recommendation for combined receiving/originating devices with internal storage.....	2223
Annex K	(informative) Notations of octet value representations	2224
Annex L	(informative) Character naming guidelines	2225
Annex M	(informative) Sources of characters	2228
Annex N	(informative) External references to character repertoires	2245
N.1	Methods of reference to character repertoires and their coding	2245
N.2	Identification of ASN.1 character abstract syntaxes	2245
N.3	Identification of ASN.1 character transfer syntaxes	2246
Annex P	(informative) Additional information on CJK Unified Ideographs	2247
Annex Q	(informative) Code mapping table for Hangul syllables	2248
Annex R	(informative) Names of Hangul syllables	2249

ISO/IEC 10646:2012 (E)

Annex S (informative) Procedure for the unification and arrangement of CJK Ideographs	2250
S.1 Unification procedure	2250
S.2 Arrangement procedure	2254
S.3 Source code separation examples.....	2254
S.4 Non-unification examples.....	2259
Annex T (informative) Language tagging using Tag Characters.....	2261
Annex U (informative) Characters in identifiers.....	2262