FINAL
DRAFT

INTERNATIONAL
STANDARD

**ISO/IEC
FDIS
10646**

ISO/IEC JTC **1**/SC **2**

Secretariat: JISC

Voting begins on:
2020-10-08

Voting terminates on:
2020-12-03

Information technology — Universal coded character set (UCS)

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

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857c-46d7-af40-a8b5890f2d11/iso-iec-fdis-10646

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

Reference number
ISO/IEC FDIS 10646:2020(E)

© ISO/IEC 2020

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857c-46d7-af40-a8b5890f2d11/iso-iec-fdis-10646

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2020

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

 © ISO/IEC 2020 – All rights reserved

CONTENTS

1 Scope ... 1
2 Normative references ... 1
3 Terms and definitions ... 2
4 Conformance ... 8
4.1 General ... 8
4.2 Conformance of information interchange ... 8
4.3 Conformance of devices ... 8
5 Electronic data attachments ... 9
6 General structure of the UCS ... 10
7 Basic structure and nomenclature ... 11
7.1 Structure ... 11
7.2 Coding of characters ... 12
7.3 Types of code points ... 12
7.4 Naming of characters ... 13
7.5 Short identifiers for code points (UIDs) ... 14
7.6 UCS Sequence Identifiers ... 14
7.7 Octet sequence identifiers ... 15
8 Revision and updating of the UCS ... 15
9 Subsets ... 15
9.1 General ... 15
9.2 Limited subset ... 15
9.3 Selected subset ... 15
10 UCS encoding forms ... 15
10.1 General ... 15
10.2 UTF-8 ... 15
10.3 UTF-16 ... 16
10.4 UTF-32 ... 17
11 UCS encoding schemes ... 17
11.1 General ... 17
11.2 UTF-8 ... 17
11.3 UTF-16BE ... 17
11.4 UTF-16LE ... 18
11.5 UTF-16 ... 18
11.6 UTF-32BE ... 18
11.7 UTF-32LE ... 18
11.8 UTF-32 ... 19
12 Use of control functions with the UCS ... 19
13 Declaration of identification of features ... 20
13.1 Purpose and context of identification ... 20
13.2 Identification of a UCS encoding scheme ... 20
13.3 Identification of subsets of graphic characters ... 21

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857c-46d7-af40-a8b5e990f2d11/iso-iec-fdis-10646

© ISO/IEC 2020 – All rights reserved

13.4 Identification of control function set ... 21
13.5 Identification of the coding system of ISO/IEC 2022 ... 21
14 Structure of the code charts and lists ... 22
15 Block and collection names ... 22
15.1 Block names ... 22
15.2 Collection names ... 23
16 Mirrored characters in bidirectional context ... 23
16.1 Mirrored characters ... 23
16.2 Directionality of bidirectional text ... 23
17 Special characters ... 23
17.1 General ... 23
17.2 Space characters ... 23
17.3 Currency symbols ... 24
17.4 Format characters ... 24
17.5 Ideographic description characters ... 24
17.6 Variation selectors and variation sequences ... 25
18 Presentation forms of characters ... 27
19 Compatibility characters ... 27
20 Order of characters ... 27
21 Combining characters ... 28
21.1 Order of combining characters ... 28
21.2 Combining class and canonical ordering ... 28
21.3 Appearance in code charts ... 28
21.4 Alternate coded representations ... 28
21.5 Multiple combining characters ... 28
21.6 Collections containing combining characters ... 29
21.7 Combining Grapheme Joiner ... 29
22 Normalization forms ... 29
23 Special features of individual scripts and symbol repertoires ... 30
23.1 Hangul syllable composition method ... 30
23.2 Features of scripts used in India and some other South Asian countries ... 30
23.3 Byzantine musical symbols ... 31
23.4 Source references for pictographic symbols ... 31
24 Source references for CJK ideographs ... 32
24.1 List of source references ... 32
24.2 Source references file for CJK ideographs ... 35
24.3 Source reference presentation for CJK Unified ideographs ... 37
24.4 Source references presentation for CJK Compatibility ideographs ... 40
25 Source references for Tangut ideographs ... 40
25.1 List of source references ... 40
25.2 Source reference file for Tangut ideographs ... 41
25.3 Source reference presentation for Tanguts ideographs ... 42
26 Source references for Nüshu characters ... 42

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857e-46d7-af40-a8b589f02d11/iso-iec-fdis-10646

 © ISO/IEC 2020 – All rights reserved

26.1 List of source references ... 42
26.2 Source reference file for Nüshu characters ... 42
27 Character names and annotations ... 43
27.1 Entity names ... 43
27.2 Name formation ... 43
27.3 Single name ... 44
27.4 Name immutability ... 44
27.5 Name uniqueness ... 44
27.6 Character names for CJK ideographs ... 45
27.7 Character names for Tangut ideographs ... 45
27.8 Character names for Nüshu characters ... 45
27.9 Character names for Khitan Small Script characters ... 46
27.10 Character names for Hangul syllables ... 46
28 Named UCS Sequence Identifiers ... 47
29 Structure of the Basic Multilingual Plane ... 49
30 Structure of the Supplementary Multilingual Plane for scripts and symbols (SMP) ... 51
31 Structure of the Supplementary Ideographic Plane (SIP) ... 54
32 Structure of the Tertiary Ideographic Plane (TIP) ... 55
33 Structure of the Supplementary Special-purpose Plane (SSP) ... 56
34 Code charts and lists of character names ... 57
34.1 General ... 57
34.2 Code chart ... 57
34.3 Character names list ... 57
34.4 Summary of standardized variation sequences ... 58
34.5 Code charts and lists of character names ... 58
Annex A (normative) Collections of graphic characters for subsets ... 2743
A.1 Collections of coded graphic characters ... 2743
A.2 Blocks lists ... 2750
A.3 Fixed collections of the whole UCS (except Unicode collections) ... 2753
A.4 CJK collections ... 2756
A.5 Other collections ... 2757
A.6 Unicode collections ... 2761
Annex B (normative) List of combining characters ... 2763
Annex C (normative) Transformation format for planes 01 to 10 of the UCS (UTF-16) ... 2764
Annex D (normative) UCS Transformation Format 8 (UTF-8) ... 2765
Annex E (normative) Mirrored characters in bidirectional context ... 2766
Annex F (informative) Format characters ... 2767
F.1 General format characters ... 2767
F.2 Script-specific format characters ... 2769
F.3 Interlinear annotation characters ... 2770
F.4 Subtending format characters ... 2770
F.5 Shorthand format characters ... 2771
F.6 Invisible mathematical operators ... 2771

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaaa0a7b-857e-46d6ea4f4ea8b5890d2d1/iso-iec-fdis-10646

© ISO/IEC 2020 – All rights reserved

F.7 Western musical symbols ... 2771
F.8 Language tagging using Tag characters ... 2772
Annex G (informative) Alphabetically sorted list of character names ... 2774
Annex H (informative) The use of "signatures" to identify UCS ... 2775
Annex I (informative) Ideographic description characters ... 2776
I.1 General ... 2776
I.2 Syntax of an ideographic description sequence ... 2776
I.3 Individual definitions of the ideographic description characters ... 2777
Annex J (informative) Recommendation for combined receiving/originating devices with internal storage ... 2779
Annex K (informative) Notations of octet value representations ... 2780
Annex L (informative) Character naming guidelines ... 2781
Annex M (informative) Sources of characters ... 2784
Annex N (informative) External references to character repertoires ... 2785
N.1 Methods of reference to character repertoires and their coding ... 2785
N.2 Identification of ASN.1 character abstract syntaxes ... 2785
N.3 Identification of ASN.1 character transfer syntaxes ... 2786
Annex P (informative) Additional information on CJK Unified ideographs ... 2787
Annex Q (informative) Code mapping table for Hangul syllables ... 2790
Annex R (informative) Names of Hangul syllables ... 2791
Annex S (informative) Procedure for the unification and arrangement of CJK ideographs ... 2792
S.1 Unification procedure ... 2792
S.2 Arrangement procedure ... 2796
S.3 Source separation examples ... 2796
S.4 Non-unification examples ... 2801
Annex T (informative) Language tagging using Tag Characters ... 2803
Annex U (informative) Characters in identifiers ... 2804

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857c-4667-a4a0-a8b5890f2d11/iso-iec-tr-10646

 © ISO/IEC 2020 – All rights reserved

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organiza-tion (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

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

This sixth edition of ISO/IEC 10646 cancels and replaces the fifth edition (ISO/IEC 10646:2017), which has been technically revised. It also incorporates ISO/IEC 10646:2017/Amd 1:2019 and ISO/IEC 10646:2017/Amd 2:2019.

This edition includes the following significant changes with respect to the previous edition:

- New scripts covered: Chorasmian, Dives Akuru, Dogra, Elymaic, Gunjala Gondi, Hanifi Rohingya, Khitan Small Script, Makasar, Medefaidrin, Nandinagari, Nyiakeng Puachue Hmong, Old Sogdian, Sogdian, Yezidi, Wancho;
- Existing scripts significantly extended: Georgian, CJK Unified Ideographs (Extension G);
- New symbol sets: Chess Symbols, Symbols for Legacy Computing;
- New set of Emoji symbols.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/...-iso-iec-fdis-10646

© ISO/IEC 2020 – All rights reserved

Introduction

This document specifies the Universal 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 additional symbols.

By defining a consistent way of encoding multilingual text it enables the exchange of data internationally. The information technology industry gains data stability, greater global interoperability and data interchange. This International Standard has been widely adopted in new Internet protocols and implemented in modern operating systems and computer languages. This edition covers over 130 000 characters from the world's scripts.

The UCS is an encoding system different from that specified in ISO/IEC 2022. The method to designate UCS from ISO/IEC 2022 is specified in 13.2.

A graphic character will be assigned only one code point in the standard, located either in the BMP or in one of the supplementary planes.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857c-46d7-af40-a8b5890f2d11/iso-iec-fdis-10646

 © ISO/IEC 2020 – All rights reserved

Information technology — Universal Coded Character Set (UCS)

1 Scope

This document

- specifies the architecture of the UCS;
- defines terms used for the UCS;
- describes the general structure of the UCS codespace;
- specifies the assigned planes of the UCS: the Basic Multilingual Plane (BMP) of the UCS, the Supplementary Multilingual Plane (SMP), the Supplementary Ideographic Plane (SIP), the Tertiary Ideographic Plane (TIP), and the Supplementary Special-purpose Plane (SSP);
- 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 and format characters of the BMP, SMP, SIP, TIP, SSP and their coded representations within the UCS codespace;
- specifies the coded representations for control characters and private use characters;
- specifies three encoding forms of the UCS: UTF-8, UTF-16, and UTF-32;
- specifies seven encoding schemes of the UCS: UTF-8, UTF-16, UTF-16BE, UTF-16LE, UTF-32, UTF-32BE, and UTF-32LE;
- specifies the management of future additions to this coded character set.

NOTE – The determination of suitability of these characters for use as identifiers in programming languages is not specified by this document but can be found in an external reference. See Annex U.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaa0a7b-857c-4d7-af40-a8b3890f2d11/iso-iec-fdis-10646

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 2022 *Information technology — Character code structure and extension techniques.*

ISO/IEC 6429 *Information technology — Control functions for coded character sets.*

Unicode Standard Annex, UAX #9, *The Unicode Bidirectional Algorithm:*
http://www.unicode.org/reports/tr9/tr9-42.html

Unicode Standard Annex, UAX #15, *Unicode Normalization Forms:*
http://www.unicode.org/reports/tr15/tr15-50.html

Unicode Technical Standard, *UTS #37, Ideographic Variation Database:*
http://www.unicode.org/reports/tr37/tr37-12.html

Unicode Standard Version 13.0, *Chapter 4, Character Properties*
http://www.unicode.org/versions/Unicode13.0.0/ch04.pdf
Section 4.3, Combining Classes – Normative
Section 4.5, General Category – Normative
Section 4.7, Bidi Mirrored – Normative

Unicode Standard Version 12.1, *Age Property:*
https://www.unicode.org/Public/13.0.0/ucd/DerivedAge.txt

© ISO/IEC 2020 – All rights reserved

Note – Parts of this document which use machine-readable format are available as electronic data attachments. See Clause 5.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
base character
graphic character which is not a combining character

Note 1 to entry – Most graphic characters are base characters. This sense of graphic combination does not preclude the presentation of base characters from adopting different contextual forms or from participating in ligatures.

Note 2 to entry – A base character typically does not graphically combine with preceding characters. There are exceptions for some complex writing systems.

3.2
Basic Multilingual Plane
BMP
plane 00 of the UCS codespace

3.3
block
contiguous range of code points to which a set of characters that share common characteristics, such as a script, are allocated; a block does not overlap another block; one or more of the code points within a block may have no character allocated to them

3.4
canonical form
form with which characters of this coded character set are specified using a single code point within the UCS codespace

Note 1 to entry – The canonical form is not to be confused with an encoding form which describes the relationship between UCS code points and one or several code units (see 3.23).

3.5
character
member of a set of elements used for the organization, control, or representation of textual data

Note 1 to entry – A graphic symbol can be represented by a sequence of one or several coded characters.

3.6
character boundary
(code unit sequence) demarcation between the last code unit of a coded character and the first code unit of the next coded character

3.7
code chart
code table
rectangular array showing the representation of coded characters allocated within a range of the UCS codespace

3.8
coded character
association between a character and a code point

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aae0a7b-857c-46d7-afd0-a8b5899d2d11/iso-iec-fdis-10646

 © ISO/IEC 2020 – All rights reserved

3.9
code point
code position
value in the UCS codespace

Note 1 to entry – Code points in the UCS codespace are integer values. Throughout this document, UCS code points are cited in hexadecimal. UCS code points range from 0 to 10FFFF.

3.10
code unit
minimal bit combination that can represent a unit of encoded text for processing or interchange

Note 1 to entry – Examples of code units are octets (8-bit code units) used in the UTF-8 encoding form, 16-bit code units in the UTF-16 encoding form, and 32-bit code units in the UTF-32 encoding form.

3.11
code unit sequence
CC-data-element
coded-character-data-element
element of interchanged information that is specified to consist of a sequence of code units, in accordance with one or more identified standards for coded character sets

Note 1 to entry – Such sequence can contain code units associated with any type of code point (see 7.3).

Note 2 to entry – Since its second edition: ISO/IEC 10646:2011, this document does not use implementation levels. Its definition of code unit sequence corresponds to the former unrestricted implementation level 3. Other definitions of code unit sequence, previously known as level 1 and 2, are deprecated. To maintain compatibility with these previous editions, in the context of identification of coded representation in International Standards such as ISO/IEC 8824 and ISO/IEC 8825, the concept of implementation level can still be referenced as 'Implementation level 3'. See Annex N.

3.12
collection
numbered and named set of entities made of code points or sequences of code points, the sequences conforming to Normalization C; code points lie within one or more identified ranges

Note 1 to entry – Non extended collections do not contain sequences of code points (see 3.25 for extended collection).

Note 2 to entry – If any of the identified ranges include code points to which no character is allocated, the repertoire of the collection will change if an additional character is assigned to any of those code points at a future amendment of this document. However, it is intended that the collection number and name will remain unchanged in future editions of this document.

3.13
combining character
character which has General Category values of Spacing Combining Mark (Mc), Non Spacing Mark (Mn), and Enclosing Mark (Me)

Note 1 to entry – These characters are intended for combination with the preceding base character, or with a sequence of combining characters preceded by a base character (see also 3.17).

3.14
combining class
value associated with each combining character determining its typographical interaction and its canonical ordering within a sequence of combining character

Note 1 to entry – See 21.2 for details on canonical ordering.

3.15
compatibility character
graphic character included as a coded character of this document primarily for compatibility with existing coded character sets

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0d7b-b72c-46d7-af4b-a1b5e90f5d11/iso-iec-fdis-10646

© ISO/IEC 2020 – All rights reserved

3.16
composite sequence
combining character sequence
sequence of graphic characters consisting of a base character followed by one or more combining characters, ZERO WIDTH JOINER, or ZERO WIDTH NON-JOINER

Note 1 to entry – See also 3.14.

Note 2 to entry – A graphic symbol for a composite sequence generally consists of the combination of the graphic symbols of each character in the sequence.

Note 3 to entry – A composite sequence can be used to represent characters not encoded in the repertoire of this document.

3.17
control character
control function the coded representation of which consists of a single code point

Note 1 to entry – Although control characters are often 'named' using terms such as DELETE, FORM FEED, ESC, these qualifiers do not correspond to formal character names. See Clause 12 for a list of the long names used by ISO/IEC 6429 in association with the control characters.

3.18
control function
action that affects the recording, processing, transmission, or interpretation of data, and that is represented by a code unit sequence

3.19
decomposition mapping
mapping from a character to a sequence of one or more characters

Note 1 to entry – Decomposition mappings are of two types: canonical decompositions, and compatibility decompositions. These are used in the derivation of various normalization forms (see Clause 22). The code charts for various blocks include decomposition mappings and distinguish between the two types of mapping (see 34.3).

3.20
default state
state that is assumed when no state has been explicitly specified

Note 1 to entry – See F.2.1, F.2.2, and F.2.3.

3.21
device
component of information processing equipment which can transmit and/or receive coded information within code unit sequences

Note 1 to entry – It may be an input/output device in the conventional sense, or a process such as an application program or gateway function.

3.22
encoding form
form that determines how each UCS code point for a UCS character is to be expressed as one or more code units used by the encoding form

Note 1 to entry – This document specifies UTF-8, UTF-16, and UTF-32.

3.23
encoding scheme
scheme that specifies the serialization of the code units from the encoding form into octets

Note 1 to entry – Some of the UCS encoding schemes have the same labels as UCS encoding form. However, references to encoding schemes and encoding forms generally occur in different contexts. UCS encoding forms refer to in-memory and application interface representation of textual data. UCS encoding schemes refer to octet-serialized textual data.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/aaae0a7b-857c-46d7-af40-a8b5f583e2d1/iso-iec-10646-2020

 © ISO/IEC 2020 – All rights reserved

3.24
extended collection
collection for which the entities can also consist of sequences of code points that are in Normalization Form C (NFC)

Note 1 to entry – Some collections such as 3 LATIN EXTENDED-A, 4 LATIN EXTENDED-B, 15 ARABIC EXTENDED, and many more, have the term 'extended' in their name. This does not make them extended collections.

Note 2 to entry – See Clause 22 for discussion of Normalization Form C.

Note 3 to entry – The sequences of code points are typically referenced by Named UCS Sequence Identifiers (NUSI) (see Clause 28).

3.25
fixed collection
collection in which every code point within the identified range(s) has a character allocated to it, and which is intended to remain unchanged in future editions of this document

3.26
format character
character whose primary function is to affect the layout or processing of characters around it, or that is presented in a complex, graphic interaction with neighbouring characters

Note 1 to entry – A format character generally does not have a visible representation of its own.

3.27
General Category
GC
value assigned to each UCS code point which determines its major class, such as letter, punctuation, and symbol

Note 1 to entry – Possible values are two-letter abbreviations for the General Category in the Unicode Standard (see reference to the current Unicode Standard General Category in Clause 2).

Note 2 to entry – When referred as a group containing all GC values sharing the same first letter, the group may be described using the first letter only. For example, 'L' stands for all letters: 'Lu', 'Ll', 'Lt', 'Lm', and 'Lo'.

3.28
graphic character
character, other than a control function or a format character, that has a visual representation normally handwritten, printed, or displayed

3.29
graphic symbol
visual representation of a graphic character or of a composite sequence

3.30
high-surrogate code point
code point in the range D800 to DBFF

Note 1 to entry – Reserved for use in UTF-16 (see 10.3).

3.31
high-surrogate code unit
16-bit code unit in the range D800 to DBFF and used in UTF-16

Note 1 to entry – A high-surrogate code unit is used as the leading code unit of a surrogate pair (see also 3.39, 3.54, and 10.3).

3.32
ill-formed code unit sequence
UCS code unit sequence that purports to be in a UCS encoding form that does not conform to the specification of that encoding form

EXAMPLE – An unpaired surrogate code unit is an ill-formed code unit sequence.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/eaaeb7b-857c-46d7-af40-a8b5e90f2d11/iso-iec-fdis-10646