

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: June 6, 2022

P. Faltstrom
Netnod
December 03, 2021

IDNA2008 and Unicode 12.0.0
draft-faltstrom-unicode12-05

Abstract

This document describes the changes between Unicode 6.0.0 and Unicode 12.0.0 in the context of IDNA2008. Some additions and changes have been made in the Unicode Standard that affect the values produced by the algorithm IDNA2008 specifies. IDNA2008 allows adding exceptions to the algorithm for backward compatibility; however, this document does not add any such exceptions. This document provides the necessary tables to IANA to make its database consistent with Unicode 12.0.0.

To improve understanding, this document describes systems that are being used as alternatives to those that conform to IDNA2008.

TO BE REMOVED AT TIME OF PUBLICATION AS AN RFC:

This document is discussed on the i18n-discuss@ietf.org mailing list of the IETF.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on June 6, 2022.

Internet-Draft

IDNA2008 and Unicode 12

December 2021

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Background	4
2.1.	IDNA2008 Documents	4
2.2.	Additional important IDNA2008-related documents	5
2.3.	Deployment	6
3.	Notable Changes Between Unicode 6.0.0 and 12.0.0	7
3.1.	Changes between Unicode 6.0.0 and 7.0.0	7
3.2.	Changes between Unicode 7.0.0 and 10.0.0	8
3.3.	Changes between Unicode 10.0.0 and 11.0.0	9
3.4.	Changes between Unicode 11.0.0 and 12.0.0	10
4.	U+111C9 SHARADA SANDHI MARK	10
5.	Conclusion	11
6.	IANA Considerations	11
7.	Security Considerations	11
8.	Acknowledgements	12
9.	References	12
9.1.	Normative References	12
9.2.	Non-normative references	12
Appendix A.	Changes from Unicode 6.0.0 to Unicode 7.0.0	15
Appendix B.	Changes from Unicode 7.0.0 to Unicode 8.0.0	20
Appendix C.	Changes from Unicode 8.0.0 to Unicode 9.0.0	22
Appendix D.	Changes from Unicode 9.0.0 to Unicode 10.0.0	24
Appendix E.	Changes from Unicode 10.0.0 to Unicode 11.0.0	26
Appendix F.	Changes from Unicode 11.0.0 to Unicode 12.0.0	27
	Author's Address	29

1. Introduction

The current version of Internationalized Domain Names for Applications (IDNA) was initiated in 2008, and despite not being completed until 2010, is widely known as "IDNA2008". It is specified

in the series of documents listed in [Section 2.1](#). The IDNA2008 standard includes an algorithm by which a derived property value is calculated based on the properties defined from the Unicode Standard.

The derived property values that can be calculated are defined in [RFC 5892](#) [[RFC5892](#)]. The summary below is a summary to make the reading of this document easier. For definition of the terms, please see [RFC 5892](#) [[RFC5892](#)].

- o **PROTOCOL VALID:** Those that are allowed to be used in IDNs. Code points with this property value are permitted for general use in IDNs. However, that a label consists only of code points that have this property value does not imply that the label can be used in DNS. The abbreviated term PVALID is used to refer to this value.
- o **CONTEXTUAL RULE REQUIRED:** Some characteristics of the character, such as it being invisible in certain contexts or problematic in others, require that it not be used in labels unless specific other characters or properties are present. The abbreviated term **CONTEXT** is used to refer to this value. As explained in [RFC 5892](#) [[RFC5892](#)] **CONTEXT** is in turn divided into **CONTEXTJ** and **CONTEXT0**.
- o **DISALLOWED:** Those that should clearly not be included in IDNs. Code points with this property value are not permitted in IDNs.
- o **UNASSIGNED:** Those code points that are not designated (i.e., are unassigned) in the Unicode Standard.

When the Unicode Standard is updated, new code points are assigned and already-assigned code points can have their property values changed.

- o Assigning code points can create problems if the newly-assigned code points are compositions of existing code points and because of that the normalization relationships associated with those code

points should have been changed.

- o Changing properties for already-assigned code points can create problems if the property change results in changes to the derived property value. This might make an earlier allowed code point whose derived property value is PVALID to then not be allowed anymore if its derived property value changes to DISALLOWED. The problem can also happen the other way around: a code point that was not allowed (and thus is prohibited) can suddenly end up being allowed.

- o Problems can also be created if the properties assigned to those code points are inconsistent with IDNA2008 assumptions about how properties are assigned and/or about how code points with those properties are used or behave.

There were three incompatible changes in the Unicode standard between Unicode 5.2.0 [[Unicode-5.2.0](#)] and Unicode 6.0.0 [[Unicode-6.0.0](#)]; they are described in [RFC 6452](#) [[RFC6452](#)]. The code points U+0CF1 and U+0CF2 had a derived property value change from DISALLOWED to PVALID, and the code point U+19DA had a change in derived property value from PVALID to DISALLOWED. These changes were examined in great detail, but the IETF concluded that these changes to the Unicode standard did not warrant an update to [RFC 5892](#) [[RFC5892](#)].

As described in [Section 3](#), more incompatible changes have been made to code points between Unicode 6.0.0 and Unicode 12.0.0 [[Unicode-12.0.0](#)]; however, the changes in the derived property values do not result in exceptions (as defined in [section 2.6 of RFC 5892](#) [[RFC5892](#)]) being added to [RFC 5892](#) [[RFC5892](#)].

Further, in 2015, the Internet Architecture Board (IAB) issued a statement [[IAB](#)] that asked the IETF to resolve the issues related to the code point ARABIC LETTER BEH WITH HAMZA ABOVE (U+08A1) that was introduced in Unicode 7.0.0 [[Unicode-7.0.0](#)]. Again, no exception is added to [RFC 5892](#) [[RFC5892](#)]; however, it should be noted that the review of the issues around U+08A1 indicated that this code point is not an isolated case and that a number of PVALID code points of long standing may have similar issues. The problem resulted in a clarification of the review process of new Unicode versions, which

are published in [RFC 8753](#) [[RFC8753](#)]. Any review of Unicode versions after 12.0.0 should be made according to [RFC 8753](#) [[RFC8753](#)]; an objective of this document is to ensure that a proper review of such versions after version 12.0.0 can be made.

[2.](#) Background

[2.1.](#) IDNA2008 Documents

IDNA2008 consists of the following documents. The documents in the set have informal names.

- o Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework [[RFC5890](#)], informally called "Defs" or "Definitions", contains definitions and other material that are needed for understanding other documents in the set.
- o Internationalized Domain Names in Applications (IDNA): Protocol [[RFC5891](#)], informally called "Protocol", describes the core

IDNA2008 protocol and its operations. It needs to be interpreted in combination with the Bidi document (described below).

- o The Unicode Code Points and Internationalized Domain Names for Applications (IDNA) [[RFC5892](#)], informally called "Tables", lists the categories and rules that identify the code points allowed in a label written in native character form (called a "U-label"), and is based on Unicode 5.2.0 [[Unicode-5.2.0](#)] code point assignments and additional rules unique to IDNA2008. The Unicode-based rules in [RFC 5892](#) are expected to be stable across Unicode updates and hence independent of Unicode versions. [RFC 5892](#) [[RFC5892](#)] obsoletes [RFC 3491](#) [[RFC3491](#)], and in particular the use of the tables to which [RFC 3491](#) [[RFC3491](#)] refers.
- o Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA) [[RFC5893](#)], informally called "Bidi", specifies special rules for labels that contain characters that are written from right to left.
- o Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale [[RFC5894](#)], informally called "Rationale", provides an overview of the protocol and

associated tables, and gives explanatory material and some rationale for the decisions that led to IDNA2008. It also contains advice for DNS registry operators and others who use Internationalized Domain Names (IDNs).

- o Mapping Characters for Internationalized Domain Names in Applications (IDNA) 2008 [[RFC5895](#)], informally called "Mapping", discusses the issue of mapping characters into other characters and provides guidance for doing so when that is appropriate. [RFC 5895](#) provides advice only and is not a required part of IDNA.

[2.2.](#) Additional important IDNA2008-related documents

There are other documents important for the understanding and functioning of IDNA2008, for example this.

- o The Unicode Code Points and Internationalized Domain Names for Applications (IDNA) - Unicode 6.0 [[RFC6452](#)] describes some changes made to Unicode 6.0.0 [[Unicode-6.0.0](#)] that resulted in derived property value change for the code points U+0CF1, U+0CF2 and U+19DA. U+0CF1 and U+0CF2 changed from DISALLOWED to PVALID, while U+19DA changed from PVALID to DISALLOWED. The IETF concluded that no update to [RFC 5892](#) [[RFC5892](#)] was needed based on the changes made in Unicode 6.0.0 [[Unicode-6.0.0](#)]. As a result, the derived property value remained aligned with the Unicode Standard. Specifically, no exception was added.

[2.3.](#) Deployment

There are many variations on the general IDNA model in use in the various parts of the community. The following lists some of the strategies that implementations that claim to be IDNA compliant are known to use, but it should be noted the list is not complete:

- o IDNA2003 as specified in [RFC 3490](#) [[RFC3490](#)] and [RFC 3491](#) [[RFC3491](#)]. Those specifications are dependent on case folding and NFKC normalization and on tables that specify for each code point whether it is allowed to be used or not, with a distinction made between use for "stored strings" and "query strings". The tables themselves are dependent on Unicode 3.2 [[Unicode-3.2.0](#)].
- o A number of variations on IDNA2003, sometimes presented as

"updated IDNA2003" or the like, which follow the principles of IDNA2003 as understood by the implementers but that use tables that represent how the implementers believe Stringprep [[RFC3454](#)] and Nameprep [[RFC3491](#)] would have evolved had the IETF not moved in the direction of IDNA2008 instead.

- o A mix between IDNA2003 and IDNA2008 where code points assigned to Unicode after Unicode 3.2.0 [[Unicode-3.2.0](#)] have derived property value calculated according to the algorithm specified in IDNA2008.
- o A mix between IDNA2003 and IDNA2008 according to the Unicode Technical Standard #46 [[UTS-46](#)]. Because that document specifies different profiles, there are several different variations that leave users with no guarantee that two applications claiming conformance to UTS#46 will interoperate well with each other much less with conforming IDNA2008 implementations. UTS#46 is ultimately based on a normative table very much like the one used by Stringprep [[RFC3454](#)] but updated for each new version of Unicode.
- o The (normative) IDNA2008 algorithm applied to whatever version of Unicode Standard exists in the operating system and/or libraries used, independent of whatever version of tables appears in the (non-normative) IANA database.

In practice, the Unicode Consortium creates a maximum set of code points by assigning code points in the Unicode Standard. The IDNA2008 rules use the Unicode Standard to create a further subset of code points and context that are permitted in DNS labels associated with its PVALID, and CONTEXT (CONTEXTJ or CONTEXT0) derived property values. DNS registries and other organizations that deal with IDNs are supposed to create their own subsets from IDNA2008 for use by those registries and organizations.

This progressive subsetting and narrowing of the repertoire of code points that can be used in labels is an implementation of the principles of being conservative when deciding what code points to include in such a subset. SAC-084 [[SAC-084](#)] and [RFC 6912](#) [[RFC6912](#)] recommend to DNS registries and other organizations to be conservative when creating their subsets, and to use the principle of creating subsets by inclusion.

See also the Security Considerations section in this document.

3. Notable Changes Between Unicode 6.0.0 and 12.0.0

Among the changes between the Unicode versions, most code points that change derived property value change from UNASSIGNED to PVALID or from UNASSIGNED to DISALLOWED. The interesting changes in derived property values include other changes. All changes between the major versions of Unicode can be found in [Appendix A](#) (6.0.0–7.0.0), [Appendix B](#) (7.0.0–8.0.0), [Appendix C](#) (8.0.0–9.0.0), [Appendix D](#) (9.0.0–10.0.0), [Appendix E](#) (10.0.0–11.0.0) and [Appendix F](#) (11.0.0–12.0.0).

3.1. Changes between Unicode 6.0.0 and 7.0.0

Change in number of characters in each category:

PVALID changed from 97418 to 99867 (+2449)

UNASSIGNED changed from 865081 to 861509 (–3572)

CONTEXTJ did not change, at 2

CONTEXT0 did not change, at 25

DISALLOWED changed from 151586 to 152709 (+1123)

TOTAL did not change, at 1114112

There are no changes made to Unicode between version 6.0.0 and 7.0.0 that impact IDNA2008 calculation of the derived property values.

The code points U+17B4 KHMER VOWEL INHERENT AQ and U+17B5 KHMER VOWEL INHERENT AA both changed the general category from Cf (Format) to Mn (Nonspacing_Mark), but that did not impact the calculation of the derived property value which stayed at DISALLOWED.

The character ARABIC LETTER BEH WITH HAMZA ABOVE (U+08A1) was introduced in Unicode 7.0.0. This was discussed extensively in the

investigate the issue. Specifically, the IAB stated:

On the same precautionary principle, the IAB recommends that the Internationalized Domain Names for Applications (IDNA) Parameters registry (<http://www.iana.org/assignments/idna-tables/>) not be updated to Unicode 7.0.0 until the IETF has consensus on a solution to this problem.

The discussion in the IETF concluded that although it is possible to create "the same" character in multiple ways, the issue with U+08A1 is not unique. The character U+08A1 (ARABIC LETTER BEH WITH HAMZA ABOVE) can be represented with the sequence ARABIC LETTER BEH (U+0628) and ARABIC HAMZA ABOVE (U+0654). This is identical to LATIN SMALL LETTER O WITH STROKE (U+00F8), which can be represented with the sequence LATIN SMALL LETTER O (U+006F) followed by COMBINING SHORT SOLIDUS OVERLAY (U+0337).

Although the discussion about this specific code point resulted in acceptance of the derived property value of PVALID, the underlying problem with combining sequences is not understood fully. Therefore it cannot be claimed that this case can be extrapolated to other situations and other code points.

3.2. Changes between Unicode 7.0.0 and 10.0.0

Change in number of characters in each category:

Code points that changed derived property value: 0

PVALID changed from 99867 to 122411 (+22544)

UNASSIGNED changed from 861509 to 837775 (-23734)

CONTEXTJ did not change, at 2

CONTEXT0 did not change, at 25

DISALLOWED changed from 152709 to 153899 (+1190)

TOTAL did not change, at 1114112

There are no changes made to Unicode between version 7.0.0 and 10.0.0 that impact IDNA2008 calculation of the derived property values.

[3.3.](#) Changes between Unicode 10.0.0 and 11.0.0

Change in number of characters in each category:

Code points that changed derived property value: 1

PVALID changed from 122411 to 122734 (+323)

UNASSIGNED changed from 837775 to 837091 (-684)

CONTEXTJ did not change, at 2

CONTEXT0 did not change, at 25

DISALLOWED changed from 153899 to 154260 (+361)

TOTAL did not change, at 1114112

Georgian letters in the ranges U+10D0..U+10FA and U+10FD..U+10FF had their General Properties changed from Lo to Ll, to reflect their status as the lowercase of new Georgian case pairs. Case mappings were also added.

SHARADA SANDHI MARK (U+111C9) was changed from Po to Mn, and from bc=L to bc=NSM.

The properties for ZANABAZAR SQUARE VOWEL SIGN AI (U+11A07) and ZANABAZAR SQUARE VOWEL SIGN AU (U+11A08) were corrected from Mc to Mn.

SPHERICAL ANGLE OPENING UP (U+29A1) was changed to Bidi_M=N.

These changes to the Unicode Standard have the following implications for these code points:

- o The newly assigned 684 characters are assigned a derived property value as of a result of applying the IDNA2008 algorithm.
- o The Georgian letters in the ranges U+10D0..U+10FA and U+10FD..U+10FF existed before IDNA2008 was created. Applying the IDNA2008 algorithm to the code points assigned the derived property value PVALID, and that value is unchanged even if the underlying Unicode properties have changed. The newly encoded Mtavruli letters have general category "Lu" and are therefore DISALLOWED.

- o The U+111C9 SHARADA SANDHI MARK was added to Unicode 8.0.0 [[Unicode-8.0.0](#)]. Applying the IDNA2008 algorithm to the code

point assigned the derived property value DISALLOWED. The changes in the underlying properties in the Unicode Standard Version 11.0.0 [[Unicode-11.0.0](#)] caused the derived property value to change to PVALID.

- o The characters ZANABAZAR SQUARE VOWEL SIGN AI (U+11A07) and ZANABAZAR SQUARE VOWEL SIGN AU (U+11A08) were added to Unicode 10.0.0 [[Unicode-10.0.0](#)]. Applying the IDNA2008 algorithm to the code points assigned the derived property value PVALID, and that value is unchanged even if the underlying Unicode properties have changed.
- o SPHERICAL ANGLE OPENING UP (U+29A1) existed before IDNA2008 was created. Applying the IDNA2008 algorithm to the code point assigned the derived property value DISALLOWED, and that value is unchanged even if the underlying Unicode properties have changed.

[3.4](#). Changes between Unicode 11.0.0 and 12.0.0

Change in number of characters in each category:

Code points that changed derived property value: 0

PVALID changed from 122734 to 123006 (+272)

UNASSIGNED changed from 837091 to 836537 (-554)

CONTEXTJ did not change, at 2

CONTEXT0 did not change, at 25

DISALLOWED changed from 154260 to 154542 (+282)

TOTAL did not change, at 1114112

[4](#). U+111C9 SHARADA SANDHI MARK

As one can see in [Section 3](#), an incompatible property change was made between Unicode 6.0.0 and 12.0.0, affecting the code point U+111C9.

Its derived property value thus changed from DISALLOWED to PVALID. In situations like these, IDNA2008 allow for addition of rules to [RFC 5892 \[RFC5892\] section 2.7](#). If the code point is accepted, it might still be rejected if validated by software based on older versions of Unicode than 11.0.0. As the character is rarely used outside of the group of Sharada specialists, and used in some records for indicating sandhi breaks, the conclusion is that it could either be added as an exception or allowed to change its property value, as the use of the code point is limited outside a special community. As including an

exception would require implementation changes in deployed implementations of IDNA2008, the IETF has decided to not add a BackwardCompatible rule to IDNA2008 (i.e. [Section 2.7 of RFC 5892 \[RFC5892\]](#) for this code point. This also ensures all sandhi marks being treated in an equal way.

[5](#). Conclusion

As described in [Section 3](#) and [Section 4](#), changes have been made to Unicode between version 6.0.0 and 12.0.0. Some changes to specific characters changed their derived property value, whereas other changes did not. Given the deployment considerations described in [Section 2.3](#) and changes in the Unicode Standard described in [Section 3](#) and [Section 4](#), including implications to normalization, the conclusion is to not add any exception rules to IDNA2008.

This document addresses only changes to Unicode between version 6.0.0 and version 12.0.0. Changes in future Unicode versions might result in the conclusion that exception rules need to be added to IDNA2008 after the review process explained in [RFC 8753 \[RFC8753\]](#). Separately from any changes in Unicode, the IETF might conclude that updates to [RFC 5892 \[RFC5892\]](#) or other IDNA2008 documents might become necessary; such updates might include changes to the algorithm specified in IDNA2008 as well as additional rules, categories, or other forms of tuning, like the clarifications in [RFC 8753 \[RFC8753\]](#).

[6](#). IANA Considerations

IANA is requested to update the IDNA Parameters registry [[IANA-IDNA](#)] of derived property values, after the expert reviewer validates that the derived property values are calculated correctly.

7. Security Considerations

This document makes recommendations regarding the use of the IDNA2008 algorithm for calculation of derived property values, based on Unicode version 12.0.0. This recommendation does not say anything about what recommendations to make for future versions of the Unicode Standard.

Not following these recommendations can lead to various security issues. Specifically, allowing confusable characters may lead to various phishing attacks, as described in the Security Consideration Sections in the documents listed in [Section 2.1](#).

Faltstrom

Expires June 6, 2022

[Page 11]

Internet-Draft

IDNA2008 and Unicode 12

December 2021

8. Acknowledgements

Thanks to Harald Alvestrand, Marc Blanchet, Martin Duerst, Asmus Freytag, Ted Hardie, John Klensin, Erik Nordmark, Pete Resnick, Peter Saint-Andre, Michel Suignard, Andrew Sullivan and Suzanne Woolf for input to this document.

9. References

9.1. Normative References

- [RFC3491] Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", [RFC 3491](#), DOI 10.17487/RFC3491, March 2003, <<https://www.rfc-editor.org/info/rfc3491>>.
- [RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", [RFC 5890](#), DOI 10.17487/RFC5890, August 2010, <<https://www.rfc-editor.org/info/rfc5890>>.
- [RFC5891] Klensin, J., "Internationalized Domain Names in Applications (IDNA): Protocol", [RFC 5891](#), DOI 10.17487/RFC5891, August 2010, <<https://www.rfc-editor.org/info/rfc5891>>.

- [RFC5892] Faltstrom, P., Ed., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)", [RFC 5892](#), DOI 10.17487/RFC5892, August 2010, <<https://www.rfc-editor.org/info/rfc5892>>.
- [RFC5893] Alvestrand, H., Ed. and C. Karp, "Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA)", [RFC 5893](#), DOI 10.17487/RFC5893, August 2010, <<https://www.rfc-editor.org/info/rfc5893>>.
- [RFC6452] Faltstrom, P., Ed. and P. Hoffman, Ed., "The Unicode Code Points and Internationalized Domain Names for Applications (IDNA) - Unicode 6.0", [RFC 6452](#), DOI 10.17487/RFC6452, November 2011, <<https://www.rfc-editor.org/info/rfc6452>>.

[9.2](#). Non-normative references

Faltstrom Expires June 6, 2022 [Page 12]

Internet-Draft IDNA2008 and Unicode 12 December 2021

- [IAB] Internet Architecture Board, "IAB Statement on Identifiers and Unicode 7.0.0", IAB Statement on Identifiers and Unicode 7.0.0
<https://www.iab.org/documents/correspondence-reports-documents/2015-2/iab-statement-on-identifiers-and-unicode-7-0-0/>, January 2015.
- [IANA-IDNA] IANA, "IDNA Rules and Derived Property Values", IDNA Rules and Derived Property Values
<https://www.iana.org/assignments/idna-tables-6.0.0/idna-tables-6.0.0.xhtml>, April 2020.
- [RFC3454] Hoffman, P. and M. Blanchet, "Preparation of Internationalized Strings ("stringprep")", [RFC 3454](#), DOI 10.17487/RFC3454, December 2002, <<https://www.rfc-editor.org/info/rfc3454>>.

- [RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", [RFC 3490](#), DOI 10.17487/RFC3490, March 2003, <<https://www.rfc-editor.org/info/rfc3490>>.
- [RFC5894] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale", [RFC 5894](#), DOI 10.17487/RFC5894, August 2010, <<https://www.rfc-editor.org/info/rfc5894>>.
- [RFC5895] Resnick, P. and P. Hoffman, "Mapping Characters for Internationalized Domain Names in Applications (IDNA) 2008", [RFC 5895](#), DOI 10.17487/RFC5895, September 2010, <<https://www.rfc-editor.org/info/rfc5895>>.
- [RFC6912] Sullivan, A., Thaler, D., Klensin, J., and O. Kolkman, "Principles for Unicode Code Point Inclusion in Labels in the DNS", [RFC 6912](#), DOI 10.17487/RFC6912, April 2013, <<https://www.rfc-editor.org/info/rfc6912>>.
- [RFC8753] Klensin, J. and P. Faelstroem, "Internationalized Domain Names for Applications (IDNA) Review for New Unicode Versions", [RFC 8753](#), DOI 10.17487/RFC8753, April 2020, <<https://www.rfc-editor.org/info/rfc8753>>.
- [SAC-084] The Security and Stability Advisory Committee, "SAC084", SSAC Comments on Guidelines for the Extended Process Similarity Review Panel for the IDN ccTLD Fast Track Process <https://www.icann.org/en/system/files/files/sac-084-en.pdf>, August 2016.

Faltstrom

Expires June 6, 2022

[Page 13]

Internet-Draft

IDNA2008 and Unicode 12

December 2021

[Unicode-10.0.0]

The Unicode Consortium, "The Unicode Standard, Version 10.0.0", The Unicode Standard, Version 10.0.0 ISBN 978-1-936213-16-0, June 2017.

[Unicode-11.0.0]

The Unicode Consortium, "The Unicode Standard, Version 11.0.0", The Unicode Standard, Version 11.0.0 ISBN 978-1-936213-19-1, June 2018.

[Unicode-12.0.0]

The Unicode Consortium, "The Unicode Standard, Version 12.0.0", The Unicode Standard, Version 12.0.0 ISBN 978-1-936213-22-1, March 2019.

[Unicode-3.2.0]

The Unicode Consortium, "The Unicode Standard, Version 3.2.0", The Unicode Standard, Version 3.2.0 ISBN 0-201-61633-5, March 2002.

[Unicode-5.2.0]

The Unicode Consortium, "The Unicode Standard, Version 5.2.0", The Unicode Standard, Version 5.2.0 ISBN 978-1-936213-00-9, October 2009.

[Unicode-6.0.0]

The Unicode Consortium, "The Unicode Standard, Version 6.0.0", The Unicode Standard, Version 6.0.0 ISBN 978-1-936213-01-6, October 2011.

[Unicode-7.0.0]

The Unicode Consortium, "The Unicode Standard, Version 7.0.0", The Unicode Standard, Version 7.0.0 ISBN 978-1-936213-09-2, June 2014.

[Unicode-8.0.0]

The Unicode Consortium, "The Unicode Standard, Version 8.0.0", The Unicode Standard, Version 8.0.0 ISBN 978-1-936213-10-8, June 2015.

[UTS-46] The Unicode Consortium, "Unicode Technical Standard #46, Version 12.0.0", UNICODE IDNA COMPATIBILITY PROCESSING <http://www.unicode.org/reports/tr46/>, March 2019.

[Appendix A](#). Changes from Unicode 6.0.0 to Unicode 7.0.0

Changes from derived property value UNASSIGNED to either PVALID or DISALLOWED.

037F	; DISALLOWED # GREEK CAPITAL LETTER YOT
0528	; DISALLOWED # CYRILLIC CAPITAL LETTER EN WITH LEFT HOOK
0529	; PVALID # CYRILLIC SMALL LETTER EN WITH LEFT HOOK
052A	; DISALLOWED # CYRILLIC CAPITAL LETTER DZZHE
052B	; PVALID # CYRILLIC SMALL LETTER DZZHE
052C	; DISALLOWED # CYRILLIC CAPITAL LETTER DCHE
052D	; PVALID # CYRILLIC SMALL LETTER DCHE
052E	; DISALLOWED # CYRILLIC CAPITAL LETTER EL WITH DESCENDER
052F	; PVALID # CYRILLIC SMALL LETTER EL WITH DESCENDER
058D..058F	; DISALLOWED # RIGHT-FACING ARMENIAN ETERNITY SIGN..ARMENIAN
0604..0605	; DISALLOWED # ARABIC SIGN SAMVAT..ARABIC NUMBER MARK ABOVE
061C	; DISALLOWED # ARABIC LETTER MARK
08A0..08B2	; PVALID # ARABIC LETTER BEH WITH SMALL V BELOW..ARABIC
08E4..08FF	; PVALID # ARABIC CURLY FATHA..ARABIC MARK SIDEWAYS NOON
0978	; PVALID # DEVANAGARI LETTER MARWARI DDA
0980	; PVALID # BENGALI ANJI
0AF0	; DISALLOWED # GUJARATI ABBREVIATION SIGN
0C00	; PVALID # TELUGU SIGN COMBINING CANDRABINDU ABOVE
0C34	; PVALID # TELUGU LETTER LLLA
0C81	; PVALID # KANNADA SIGN CANDRABINDU
0D01	; PVALID # MALAYALAM SIGN CANDRABINDU
0DE6..0DEF	; PVALID # SINHALA LITH DIGIT ZERO..SINHALA LITH DIGIT N
0EDE..0EDF	; PVALID # LAO LETTER KHMU GO..LAO LETTER KHMU NYO
10C7	; DISALLOWED # GEORGIAN CAPITAL LETTER YN
10CD	; DISALLOWED # GEORGIAN CAPITAL LETTER AEN
10FD..10FF	; PVALID # GEORGIAN LETTER AEN..GEORGIAN LETTER LABIAL S
16F1..16F8	; PVALID # RUNIC LETTER K..RUNIC LETTER FRANKS CASKET AE
17B4..17B5	; DISALLOWED # KHMER VOWEL INHERENT AQ..KHMER VOWEL INHERENT
191D..191E	; PVALID # LIMBU LETTER GYAN..LIMBU LETTER TRA
1AB0..1ABD	; PVALID # COMBINING DOUBLED CIRCUMFLEX ACCENT..COMBININ
1ABE	; DISALLOWED # COMBINING PARENTHESES OVERLAY
1BAB..1BAD	; PVALID # SUNDANESE SIGN VIRAMA..SUNDANESE CONSONANT SI
1BBA..1BBF	; PVALID # SUNDANESE AVAGRAHA..SUNDANESE LETTER FINAL M
1CC0..1CC7	; DISALLOWED # SUNDANESE PUNCTUATION BINDU SURYA..SUNDANESE
1CF3..1CF6	; PVALID # VEDIC SIGN ROTATED ARDHAVISARGA..VEDIC SIGN U
1CF8..1CF9	; PVALID # VEDIC TONE RING ABOVE..VEDIC TONE DOUBLE RING
1DE7..1DF5	; PVALID # COMBINING LATIN SMALL LETTER ALPHA..COMBINING
2066..2069	; DISALLOWED # LEFT-TO-RIGHT ISOLATE..POP DIRECTIONAL ISOLAT
20BA..20BD	; DISALLOWED # TURKISH LIRA SIGN..RUBLE SIGN
23F4..23FA	; DISALLOWED # BLACK MEDIUM LEFT-POINTING TRIANGLE..BLACK CI
2700	; DISALLOWED # BLACK SAFETY SCISSORS
27CB	; DISALLOWED # MATHEMATICAL RISING DIAGONAL
27CD	; DISALLOWED # MATHEMATICAL FALLING DIAGONAL

2B4D..2B4F ; DISALLOWED # DOWNWARDS TRIANGLE-HEADED ZIGZAG ARROW..SHORT
2B5A..2B73 ; DISALLOWED # SLANTED NORTH ARROW WITH HOOKED HEAD..DOWNWAR
2B76..2B95 ; DISALLOWED # NORTH WEST TRIANGLE-HEADED ARROW TO BAR..RIGH
2B98..2BB9 ; DISALLOWED # THREE-D TOP-LIGHTED LEFTWARDS EQUILATERAL ARR
2BBD..2BC8 ; DISALLOWED # BALLOT BOX WITH LIGHT X..BLACK MEDIUM RIGHT-P
2BCA..2BD1 ; DISALLOWED # TOP HALF BLACK CIRCLE..UNCERTAINTY SIGN
2CF2 ; DISALLOWED # COPTIC CAPITAL LETTER BOHAIRIC KHEI
2CF3 ; PVALID # COPTIC SMALL LETTER BOHAIRIC KHEI
2D27 ; PVALID # GEORGIAN SMALL LETTER YN
2D2D ; PVALID # GEORGIAN SMALL LETTER AEN
2D66..2D67 ; PVALID # TIFINAGH LETTER YE..TIFINAGH LETTER YO
2E32..2E42 ; DISALLOWED # TURNED COMMA..DOUBLE LOW-REVERSED-9 QUOTATION
9FCC ; PVALID # <CJK Ideograph>
A674..A67B ; PVALID # COMBINING CYRILLIC LETTER UKRAINIAN IE..COMBI
A698 ; DISALLOWED # CYRILLIC CAPITAL LETTER DOUBLE O
A699 ; PVALID # CYRILLIC SMALL LETTER DOUBLE O
A69A ; DISALLOWED # CYRILLIC CAPITAL LETTER CROSSED O
A69B ; PVALID # CYRILLIC SMALL LETTER CROSSED O
A69C..A69D ; DISALLOWED # MODIFIER LETTER CYRILLIC HARD SIGN..MODIFIER
A69F ; PVALID # COMBINING CYRILLIC LETTER IOTIFIED E
A792 ; DISALLOWED # LATIN CAPITAL LETTER C WITH BAR
A793..A795 ; PVALID # LATIN SMALL LETTER C WITH BAR..LATIN SMALL LE
A796 ; DISALLOWED # LATIN CAPITAL LETTER B WITH FLOURISH
A797 ; PVALID # LATIN SMALL LETTER B WITH FLOURISH
A798 ; DISALLOWED # LATIN CAPITAL LETTER F WITH STROKE
A799 ; PVALID # LATIN SMALL LETTER F WITH STROKE
A79A ; DISALLOWED # LATIN CAPITAL LETTER VOLAPUK AE
A79B ; PVALID # LATIN SMALL LETTER VOLAPUK AE
A79C ; DISALLOWED # LATIN CAPITAL LETTER VOLAPUK OE
A79D ; PVALID # LATIN SMALL LETTER VOLAPUK OE
A79E ; DISALLOWED # LATIN CAPITAL LETTER VOLAPUK UE
A79F ; PVALID # LATIN SMALL LETTER VOLAPUK UE
A7AA..A7AD ; DISALLOWED # LATIN CAPITAL LETTER H WITH HOOK..LATIN CAPIT
A7B0..A7B1 ; DISALLOWED # LATIN CAPITAL LETTER TURNED K..LATIN CAPITAL
A7F7 ; PVALID # LATIN EPIGRAPHIC LETTER SIDEWAYS I
A7F8..A7F9 ; DISALLOWED # MODIFIER LETTER CAPITAL H WITH STROKE..MODIFI
A9E0..A9FE ; PVALID # MYANMAR LETTER SHAN GHA..MYANMAR LETTER TAI L
AA7C..AA7F ; PVALID # MYANMAR SIGN TAI LAING TONE-2..MYANMAR LETTER
AAE0..AAEF ; PVALID # MEETEI MAYEK LETTER E..MEETEI MAYEK VOWEL SIG
AAF0..AAF1 ; DISALLOWED # MEETEI MAYEK CHEIKHAN..MEETEI MAYEK AHANG KHU
AAF2..AAF6 ; PVALID # MEETEI MAYEK ANJI..MEETEI MAYEK VIRAMA
AB30..AB5A ; PVALID # LATIN SMALL LETTER BARRED ALPHA..LATIN SMALL
AB5B..AB5F ; DISALLOWED # MODIFIER BREVE WITH INVERTED BREVE..MODIFIER
AB64..AB65 ; PVALID # LATIN SMALL LETTER INVERTED ALPHA..GREEK LETT
FA2E..FA2F ; DISALLOWED # CJK COMPATIBILITY IDEOGRAPH-FA2E..CJK COMPATI
FE27..FE2D ; PVALID # COMBINING LIGATURE LEFT HALF BELOW..COMBINING
1018B..1018C; DISALLOWED # GREEK ONE QUARTER SIGN..GREEK SINUSOID SIGN

101A0 ; DISALLOWED # GREEK SYMBOL TAU RHO

Faltstrom

Expires June 6, 2022

[Page 16]

Internet-Draft

IDNA2008 and Unicode 12

December 2021

102E0 ; PVALID # COPTIC EPACT THOUSANDS MARK
102E1..102FB; DISALLOWED # COPTIC EPACT DIGIT ONE..COPTIC EPACT NUMBER N
1031F ; PVALID # OLD ITALIC LETTER ESS
10350..1037A; PVALID # OLD PERMIC LETTER AN..COMBINING OLD PERMIC LE
10500..10527; PVALID # ELBASAN LETTER A..ELBASAN LETTER KHE
10530..10563; PVALID # CAUCASIAN ALBANIAN LETTER ALT..CAUCASIAN ALBA
1056F ; DISALLOWED # CAUCASIAN ALBANIAN CITATION MARK
10600..10736; PVALID # LINEAR A SIGN AB001..LINEAR A SIGN A664
10740..10755; PVALID # LINEAR A SIGN A701 A..LINEAR A SIGN A732 JE
10760..10767; PVALID # LINEAR A SIGN A800..LINEAR A SIGN A807
10860..10876; PVALID # PALMYRENE LETTER ALEPH..PALMYRENE LETTER TAW
10877..1087F; DISALLOWED # PALMYRENE LEFT-POINTING FLEURON..PALMYRENE NU
10880..1089E; PVALID # NABATAEAN LETTER FINAL ALEPH..NABATAEAN LETTE
108A7..108AF; DISALLOWED # NABATAEAN NUMBER ONE..NABATAEAN NUMBER ONE HU
10980..109B7; PVALID # MEROITIC HIEROGLYPHIC LETTER A..MEROITIC CURS
109BE..109BF; PVALID # MEROITIC CURSIVE LOGOGRAM RMT..MEROITIC CURSI
10A80..10A9C; PVALID # OLD NORTH ARABIAN LETTER HEH..OLD NORTH ARABI
10A9D..10A9F; DISALLOWED # OLD NORTH ARABIAN NUMBER ONE..OLD NORTH ARABI
10AC0..10AC7; PVALID # MANICHAEAN LETTER ALEPH..MANICHAEAN LETTER WA
10AC8 ; DISALLOWED # MANICHAEAN SIGN UD
10AC9..10AE6; PVALID # MANICHAEAN LETTER ZAYIN..MANICHAEAN ABBREVIAT
10AEB..10AF6; DISALLOWED # MANICHAEAN NUMBER ONE..MANICHAEAN PUNCTUATION
10B80..10B91; PVALID # PSALTER PAHLAVI LETTER ALEPH..PSALTER PAHLAVI
10B99..10B9C; DISALLOWED # PSALTER PAHLAVI SECTION MARK..PSALTER PAHLAVI
10BA9..10BAF; DISALLOWED # PSALTER PAHLAVI NUMBER ONE..PSALTER PAHLAVI N
1107F ; PVALID # BRAHMI NUMBER JOINER
110D0..110E8; PVALID # SORA SOMPENG LETTER SAH..SORA SOMPENG LETTER
110F0..110F9; PVALID # SORA SOMPENG DIGIT ZERO..SORA SOMPENG DIGIT N
11100..11134; PVALID # CHAKMA SIGN CANDRABINDU..CHAKMA MAAYYAA
11136..1113F; PVALID # CHAKMA DIGIT ZERO..CHAKMA DIGIT NINE
11140..11143; DISALLOWED # CHAKMA SECTION MARK..CHAKMA QUESTION MARK
11150..11173; PVALID # MAHAJANI LETTER A..MAHAJANI SIGN NUKTA
11174..11175; DISALLOWED # MAHAJANI ABBREVIATION SIGN..MAHAJANI SECTION
[11176](#) ; PVALID # MAHAJANI LIGATURE SHRI
11180..111C4; PVALID # SHARADA SIGN CANDRABINDU..SHARADA OM
111C5..111C8; DISALLOWED # SHARADA DANDA..SHARADA SEPARATOR
111CD ; DISALLOWED # SHARADA SUTRA MARK
111D0..111DA; PVALID # SHARADA DIGIT ZERO..SHARADA EKAM
111E1..111F4; DISALLOWED # SINHALA ARCHAIC DIGIT ONE..SINHALA ARCHAIC NU
11200..11211; PVALID # KHOJKI LETTER A..KHOJKI LETTER JJA

11213..11237;	PVALID	# KHOJKI LETTER NYA..KHOJKI SIGN SHADDA
11238..1123D;	DISALLOWED	# KHOJKI DANDA..KHOJKI ABBREVIATION SIGN
112B0..112EA;	PVALID	# KHUDAWADI LETTER A..KHUDAWADI SIGN VIRAMA
112F0..112F9;	PVALID	# KHUDAWADI DIGIT ZERO..KHUDAWADI DIGIT NINE
11301..11303;	PVALID	# GRANTHA SIGN CANDRABINDU..GRANTHA SIGN VISARG
11305..1130C;	PVALID	# GRANTHA LETTER A..GRANTHA LETTER VOCALIC L
1130F..11310;	PVALID	# GRANTHA LETTER EE..GRANTHA LETTER AI
11313..11328;	PVALID	# GRANTHA LETTER OO..GRANTHA LETTER NA

Faltstrom

Expires June 6, 2022

[Page 17]

Internet-Draft

IDNA2008 and Unicode 12

December 2021

1132A..11330;	PVALID	# GRANTHA LETTER PA..GRANTHA LETTER RA
11332..11333;	PVALID	# GRANTHA LETTER LA..GRANTHA LETTER LLA
11335..11339;	PVALID	# GRANTHA LETTER VA..GRANTHA LETTER HA
1133C..11344;	PVALID	# GRANTHA SIGN NUKTA..GRANTHA VOWEL SIGN VOCALI
11347..11348;	PVALID	# GRANTHA VOWEL SIGN EE..GRANTHA VOWEL SIGN AI
1134B..1134D;	PVALID	# GRANTHA VOWEL SIGN OO..GRANTHA SIGN VIRAMA
11357	; PVALID	# GRANTHA AU LENGTH MARK
1135D..11363;	PVALID	# GRANTHA SIGN PLUTA..GRANTHA VOWEL SIGN VOCALI
11366..1136C;	PVALID	# COMBINING GRANTHA DIGIT ZERO..COMBINING GRANT
11370..11374;	PVALID	# COMBINING GRANTHA LETTER A..COMBINING GRANTHA
11480..114C5;	PVALID	# TIRHUTA ANJI..TIRHUTA GVANG
114C6	; DISALLOWED	# TIRHUTA ABBREVIATION SIGN
114C7	; PVALID	# TIRHUTA OM
114D0..114D9;	PVALID	# TIRHUTA DIGIT ZERO..TIRHUTA DIGIT NINE
11580..115B5;	PVALID	# SIDDHAM LETTER A..SIDDHAM VOWEL SIGN VOCALIC
115B8..115C0;	PVALID	# SIDDHAM VOWEL SIGN E..SIDDHAM SIGN NUKTA
115C1..115C9;	DISALLOWED	# SIDDHAM SIGN SIDDHAM..SIDDHAM END OF TEXT MAR
11600..11640;	PVALID	# MODI LETTER A..MODI SIGN ARDHACANDRA
11641..11643;	DISALLOWED	# MODI DANDA..MODI ABBREVIATION SIGN
11644	; PVALID	# MODI SIGN HUVA
11650..11659;	PVALID	# MODI DIGIT ZERO..MODI DIGIT NINE
11680..116B7;	PVALID	# TAKRI LETTER A..TAKRI SIGN NUKTA
116C0..116C9;	PVALID	# TAKRI DIGIT ZERO..TAKRI DIGIT NINE
118A0..118BF;	DISALLOWED	# WARANG CITI CAPITAL LETTER NGAA..WARANG CITI
118C0..118E9;	PVALID	# WARANG CITI SMALL LETTER NGAA..WARANG CITI DI
118EA..118F2;	DISALLOWED	# WARANG CITI NUMBER TEN..WARANG CITI NUMBER NI
118FF	; PVALID	# WARANG CITI OM
11AC0..11AF8;	PVALID	# PAU CIN HAU LETTER PA..PAU CIN HAU GLOTTAL ST
1236F..12398;	PVALID	# CUNEIFORM SIGN KAP ELAMITE..CUNEIFORM SIGN UM
12463..1246E;	DISALLOWED	# CUNEIFORM NUMERIC SIGN ONE QUARTER GUR..CUNEI
12474	; DISALLOWED	# CUNEIFORM PUNCTUATION SIGN DIAGONAL QUADCOLON
16A40..16A5E;	PVALID	# MRO LETTER TA..MRO LETTER TEK
16A60..16A69;	PVALID	# MRO DIGIT ZERO..MRO DIGIT NINE

16A6E..16A6F; DISALLOWED # MRO DANDA..MRO DOUBLE DANDA
 16AD0..16AED; PVALID # BASSA VAH LETTER ENNI..BASSA VAH LETTER I
 16AF0..16AF4; PVALID # BASSA VAH COMBINING HIGH TONE..BASSA VAH COMB
 16AF5 ; DISALLOWED # BASSA VAH FULL STOP
 16B00..16B36; PVALID # PAHAHW HMONG VOWEL KEEB..PAHAHW HMONG MARK CI
 16B37..16B3F; DISALLOWED # PAHAHW HMONG SIGN VOS THOM..PAHAHW HMONG SIGN
 16B40..16B43; PVALID # PAHAHW HMONG SIGN VOS SEEV..PAHAHW HMONG SIGN
 16B44..16B45; DISALLOWED # PAHAHW HMONG SIGN XAUS..PAHAHW HMONG SIGN CIM
 16B50..16B59; PVALID # PAHAHW HMONG DIGIT ZERO..PAHAHW HMONG DIGIT N
 16B5B..16B61; DISALLOWED # PAHAHW HMONG NUMBER TENS..PAHAHW HMONG NUMBER
 16B63..16B77; PVALID # PAHAHW HMONG SIGN VOS LUB..PAHAHW HMONG SIGN
 16B7D..16B8F; PVALID # PAHAHW HMONG CLAN SIGN TSHEEJ..PAHAHW HMONG C
 16F00..16F44; PVALID # MIAO LETTER PA..MIAO LETTER HHA
 16F50..16F7E; PVALID # MIAO LETTER NASALIZATION..MIAO VOWEL SIGN NG
 16F8F..16F9F; PVALID # MIAO TONE RIGHT..MIAO LETTER REFORMED TONE-8

Faltstrom

Expires June 6, 2022

[Page 18]

Internet-Draft

IDNA2008 and Unicode 12

December 2021

1BC00..1BC6A; PVALID # DUPLOYAN LETTER H..DUPLOYAN LETTER VOCALIC M
 1BC70..1BC7C; PVALID # DUPLOYAN AFFIX LEFT HORIZONTAL SECANT..DUPLOY
 1BC80..1BC88; PVALID # DUPLOYAN AFFIX HIGH ACUTE..DUPLOYAN AFFIX HIG
 1BC90..1BC99; PVALID # DUPLOYAN AFFIX LOW ACUTE..DUPLOYAN AFFIX LOW
 1BC9C ; DISALLOWED # DUPLOYAN SIGN O WITH CROSS
 1BC9D..1BC9E; PVALID # DUPLOYAN THICK LETTER SELECTOR..DUPLOYAN DOUB
 1BC9F..1BCA3; DISALLOWED # DUPLOYAN PUNCTUATION CHINOOK FULL STOP..SHORT
 1E800..1E8C4; PVALID # MENDE KIKAKUI SYLLABLE M001 KI..MENDE KIKAKUI
 1E8C7..1E8CF; DISALLOWED # MENDE KIKAKUI DIGIT ONE..MENDE KIKAKUI DIGIT
 1E8D0..1E8D6; PVALID # MENDE KIKAKUI COMBINING NUMBER TEENS..MENDE K
 1EE00..1EE03; DISALLOWED # ARABIC MATHEMATICAL ALEF..ARABIC MATHEMATICAL
 1EE05..1EE1F; DISALLOWED # ARABIC MATHEMATICAL WAW..ARABIC MATHEMATICAL
 1EE21..1EE22; DISALLOWED # ARABIC MATHEMATICAL INITIAL BEH..ARABIC MATHE
 1EE24 ; DISALLOWED # ARABIC MATHEMATICAL INITIAL HEH
 1EE27 ; DISALLOWED # ARABIC MATHEMATICAL INITIAL HAH
 1EE29..1EE32; DISALLOWED # ARABIC MATHEMATICAL INITIAL YEH..ARABIC MATHE
 1EE34..1EE37; DISALLOWED # ARABIC MATHEMATICAL INITIAL SHEEN..ARABIC MAT
 1EE39 ; DISALLOWED # ARABIC MATHEMATICAL INITIAL DAD
 1EE3B ; DISALLOWED # ARABIC MATHEMATICAL INITIAL GHAIN
 1EE42 ; DISALLOWED # ARABIC MATHEMATICAL TAILED JEEM
 1EE47 ; DISALLOWED # ARABIC MATHEMATICAL TAILED HAH
 1EE49 ; DISALLOWED # ARABIC MATHEMATICAL TAILED YEH
 1EE4B ; DISALLOWED # ARABIC MATHEMATICAL TAILED LAM
 1EE4D..1EE4F; DISALLOWED # ARABIC MATHEMATICAL TAILED NOON..ARABIC MATHE
 1EE51..1EE52; DISALLOWED # ARABIC MATHEMATICAL TAILED SAD..ARABIC MATHEM
 1EE54 ; DISALLOWED # ARABIC MATHEMATICAL TAILED SHEEN

1EE57 ; DISALLOWED # ARABIC MATHEMATICAL TAILED KHAH
 1EE59 ; DISALLOWED # ARABIC MATHEMATICAL TAILED DAD
 1EE5B ; DISALLOWED # ARABIC MATHEMATICAL TAILED GHAIN
 1EE5D ; DISALLOWED # ARABIC MATHEMATICAL TAILED DOTLESS NOON
 1EE5F ; DISALLOWED # ARABIC MATHEMATICAL TAILED DOTLESS QAF
 1EE61..1EE62; DISALLOWED # ARABIC MATHEMATICAL STRETCHED BEH..ARABIC MAT
 1EE64 ; DISALLOWED # ARABIC MATHEMATICAL STRETCHED HEH
 1EE67..1EE6A; DISALLOWED # ARABIC MATHEMATICAL STRETCHED HAH..ARABIC MAT
 1EE6C..1EE72; DISALLOWED # ARABIC MATHEMATICAL STRETCHED MEEM..ARABIC MA
 1EE74..1EE77; DISALLOWED # ARABIC MATHEMATICAL STRETCHED SHEEN..ARABIC M
 1EE79..1EE7C; DISALLOWED # ARABIC MATHEMATICAL STRETCHED DAD..ARABIC MAT
 1EE7E ; DISALLOWED # ARABIC MATHEMATICAL STRETCHED DOTLESS FEH
 1EE80..1EE89; DISALLOWED # ARABIC MATHEMATICAL LOOPED ALEF..ARABIC MATHE
 1EE8B..1EE9B; DISALLOWED # ARABIC MATHEMATICAL LOOPED LAM..ARABIC MATHEM
 1EEA1..1EEA3; DISALLOWED # ARABIC MATHEMATICAL DOUBLE-STRUCK BEH..ARABIC
 1EEA5..1EEA9; DISALLOWED # ARABIC MATHEMATICAL DOUBLE-STRUCK WAW..ARABIC
 1EEAB..1EEBB; DISALLOWED # ARABIC MATHEMATICAL DOUBLE-STRUCK LAM..ARABIC
 1EEF0..1EEF1; DISALLOWED # ARABIC MATHEMATICAL OPERATOR MEEM WITH HAH WI
 1F0BF ; DISALLOWED # PLAYING CARD RED JOKER
 1F0E0..1F0F5; DISALLOWED # PLAYING CARD FOOL..PLAYING CARD TRUMP-21
 1F10B..1F10C; DISALLOWED # DINGBAT CIRCLED SANS-SERIF DIGIT ZERO..DINGBA
 1F16A..1F16B; DISALLOWED # RAISED MC SIGN..RAISED MD SIGN

1F321..1F32C; DISALLOWED # THERMOMETER..WIND BLOWING FACE
 1F336 ; DISALLOWED # HOT PEPPER
 1F37D ; DISALLOWED # FORK AND KNIFE WITH PLATE
 1F394..1F39F; DISALLOWED # HEART WITH TIP ON THE LEFT..ADMISSION TICKETS
 1F3C5 ; DISALLOWED # SPORTS MEDAL
 1F3CB..1F3CE; DISALLOWED # WEIGHT LIFTER..RACING CAR
 1F3D4..1F3DF; DISALLOWED # SNOW CAPPED MOUNTAIN..STADIUM
 1F3F1..1F3F7; DISALLOWED # WHITE PENNANT..LABEL
 1F43F ; DISALLOWED # CHIPMUNK
 1F441 ; DISALLOWED # EYE
 1F4F8 ; DISALLOWED # CAMERA WITH FLASH
 1F4FD..1F4FE; DISALLOWED # FILM PROJECTOR..PORTABLE STEREO
 1F53E..1F54A; DISALLOWED # LOWER RIGHT SHADOWED WHITE CIRCLE..DOVE OF PE
 1F568..1F579; DISALLOWED # RIGHT SPEAKER..JOYSTICK
 1F57B..1F5A3; DISALLOWED # LEFT HAND TELEPHONE RECEIVER..BLACK DOWN POIN
 1F5A5..1F5FA; DISALLOWED # DESKTOP COMPUTER..WORLD MAP
 1F600 ; DISALLOWED # GRINNING FACE
 1F611 ; DISALLOWED # EXPRESSIONLESS FACE
 1F615 ; DISALLOWED # CONFUSED FACE

1F617 ; DISALLOWED # KISSING FACE
 1F619 ; DISALLOWED # KISSING FACE WITH SMILING EYES
 1F61B ; DISALLOWED # FACE WITH STUCK-OUT TONGUE
 1F61F ; DISALLOWED # WORRIED FACE
 1F626..1F627; DISALLOWED # FROWNING FACE WITH OPEN MOUTH..ANGUISHED FACE
 1F62C ; DISALLOWED # GRIMACING FACE
 1F62E..1F62F; DISALLOWED # FACE WITH OPEN MOUTH..HUSHED FACE
 1F634 ; DISALLOWED # SLEEPING FACE
 1F641..1F642; DISALLOWED # SLIGHTLY FROWNING FACE..SLIGHTLY SMILING FACE
 1F650..1F67F; DISALLOWED # NORTH WEST POINTING LEAF..REVERSE CHECKER BOA
 1F6C6..1F6CF; DISALLOWED # TRIANGLE WITH ROUNDED CORNERS..BED
 1F6E0..1F6EC; DISALLOWED # HAMMER AND WRENCH..AIRPLANE ARRIVING
 1F6F0..1F6F3; DISALLOWED # SATELLITE..PASSENGER SHIP
 1F780..1F7D4; DISALLOWED # BLACK LEFT-POINTING ISOSCELES RIGHT TRIANGLE.
 1F800..1F80B; DISALLOWED # LEFTWARDS ARROW WITH SMALL TRIANGLE ARROWHEAD
 1F810..1F847; DISALLOWED # LEFTWARDS ARROW WITH SMALL EQUILATERAL ARROWH
 1F850..1F859; DISALLOWED # LEFTWARDS SANS-SERIF ARROW..UP DOWN SANS-SERI
 1F860..1F887; DISALLOWED # WIDE-HEADED LEFTWARDS LIGHT BARB ARROW..WIDE-
 1F890..1F8AD; DISALLOWED # LEFTWARDS TRIANGLE ARROWHEAD..WHITE ARROW SHA

[Appendix B.](#) Changes from Unicode 7.0.0 to Unicode 8.0.0

Changes from derived property value UNASSIGNED to either PVALID or DISALLOWED.

08B3..08B4 ; PVALID # ARABIC LETTER AIN WITH THREE DOTS BELOW..ARAB
 08E3 ; PVALID # ARABIC TURNED DAMMA BELOW
 0AF9 ; PVALID # GUJARATI LETTER ZHA

Faltstrom

Expires June 6, 2022

[Page 20]

Internet-Draft

IDNA2008 and Unicode 12

December 2021

0C5A ; PVALID # TELUGU LETTER RRRRA
 0D5F ; PVALID # MALAYALAM LETTER ARCHAIC II
 13F5 ; PVALID # CHEROKEE LETTER MV
 13F8..13FD ; DISALLOWED # CHEROKEE SMALL LETTER YE..CHEROKEE SMALL LETT
 20BE ; DISALLOWED # LARI SIGN
 218A..218B ; DISALLOWED # TURNED DIGIT TWO..TURNED DIGIT THREE
 2BEC..2BEF ; DISALLOWED # LEFTWARDS TWO-HEADED ARROW WITH TRIANGLE ARRO
 9FCD..9FD5 ; PVALID # <CJK Ideograph>..<CJK Ideograph>
 A69E ; PVALID # COMBINING CYRILLIC LETTER EF
 A78F ; PVALID # LATIN LETTER SINOLOGICAL DOT
 A7B2..A7B4 ; DISALLOWED # LATIN CAPITAL LETTER J WITH CROSSED-TAIL..LAT
 A7B5 ; PVALID # LATIN SMALL LETTER BETA

A7B6 ; DISALLOWED # LATIN CAPITAL LETTER OMEGA
 A7B7 ; PVALID # LATIN SMALL LETTER OMEGA
 A8FC ; DISALLOWED # DEVANAGARI SIGN SIDDHAM
 A8FD ; PVALID # DEVANAGARI JAIN OM
 AB60..AB63 ; PVALID # LATIN SMALL LETTER SAKHA YAT..LATIN SMALL LET
 AB70..ABBF ; DISALLOWED # CHEROKEE SMALL LETTER A..CHEROKEE SMALL LETTE
 FE2E..FE2F ; PVALID # COMBINING CYRILLIC TITLO LEFT HALF..COMBINING
 108E0..108F2; PVALID # HATRAN LETTER ALEPH..HATRAN LETTER QOPH
 108F4..108F5; PVALID # HATRAN LETTER SHIN..HATRAN LETTER TAW
 108FB..108FF; DISALLOWED # HATRAN NUMBER ONE..HATRAN NUMBER ONE HUNDRED
 109BC..109BD; DISALLOWED # MEROITIC CURSIVE FRACTION ELEVEN TWELFTHS..ME
 109C0..109CF; DISALLOWED # MEROITIC CURSIVE NUMBER ONE..MEROITIC CURSIVE
 109D2..109FF; DISALLOWED # MEROITIC CURSIVE NUMBER ONE HUNDRED..MEROITIC
 10C80..10CB2; DISALLOWED # OLD HUNGARIAN CAPITAL LETTER A..OLD HUNGARIAN
 10CC0..10CF2; PVALID # OLD HUNGARIAN SMALL LETTER A..OLD HUNGARIAN S
 10CFA..10CFF; DISALLOWED # OLD HUNGARIAN NUMBER ONE..OLD HUNGARIAN NUMBE
 111C9 ; DISALLOWED # SHARADA SANDHI MARK
 111CA..111CC; PVALID # SHARADA SIGN NUKTA..SHARADA EXTRA SHORT VOWEL
 111DB ; DISALLOWED # SHARADA SIGN SIDDHAM
 111DC ; PVALID # SHARADA HEADSTROKE
 111DD..111DF; DISALLOWED # SHARADA CONTINUATION SIGN..SHARADA SECTION MA
 11280..11286; PVALID # MULTANI LETTER A..MULTANI LETTER GA
[11288](#) ; PVALID # MULTANI LETTER GHA
 1128A..1128D; PVALID # MULTANI LETTER CA..MULTANI LETTER JJA
 1128F..1129D; PVALID # MULTANI LETTER NYA..MULTANI LETTER BA
 1129F..112A8; PVALID # MULTANI LETTER BHA..MULTANI LETTER RHA
 112A9 ; DISALLOWED # MULTANI SECTION MARK
[11300](#) ; PVALID # GRANTHA SIGN COMBINING ANUSVARA ABOVE
[11350](#) ; PVALID # GRANTHA OM
 115CA..115D7; DISALLOWED # SIDDHAM SECTION MARK WITH TRIDENT AND U-SHAPE
 115D8..115DD; PVALID # SIDDHAM LETTER THREE-CIRCLE ALTERNATE I..SIDD
 11700..11719; PVALID # AHOM LETTER KA..AHOM LETTER JHA
 1171D..1172B; PVALID # AHOM CONSONANT SIGN MEDIAL LA..AHOM SIGN KILL
 11730..11739; PVALID # AHOM DIGIT ZERO..AHOM DIGIT NINE
 1173A..1173F; DISALLOWED # AHOM NUMBER TEN..AHOM SYMBOL VI
[12399](#) ; PVALID # CUNEIFORM SIGN U U

12480..12543; PVALID # CUNEIFORM SIGN AB TIMES NUN TENU..CUNEIFORM S
 14400..14646; PVALID # ANATOLIAN HIEROGLYPH A001..ANATOLIAN HIEROGLY
 1D1DE..1D1E8; DISALLOWED # MUSICAL SYMBOL KIEVAN C CLEF..MUSICAL SYMBOL
 1D800..1D9FF; DISALLOWED # SIGNWRITING HAND-FIST INDEX..SIGNWRITING HEAD
 1DA00..1DA36; PVALID # SIGNWRITING HEAD RIM..SIGNWRITING AIR SUCKING

1DA37..1DA3A; DISALLOWED # SIGNWRITING AIR BLOW SMALL ROTATIONS..SIGNWRI
 1DA3B..1DA6C; PVALID # SIGNWRITING MOUTH CLOSED NEUTRAL..SIGNWRITING
 1DA6D..1DA74; DISALLOWED # SIGNWRITING SHOULDER HIP SPINE..SIGNWRITING T
 1DA75 ; PVALID # SIGNWRITING UPPER BODY TILTING FROM HIP JOINT
 1DA76..1DA83; DISALLOWED # SIGNWRITING LIMB COMBINATION..SIGNWRITING LOC
 1DA84 ; PVALID # SIGNWRITING LOCATION HEAD NECK
 1DA85..1DA8B; DISALLOWED # SIGNWRITING LOCATION TORSO..SIGNWRITING PAREN
 1DA9B..1DA9F; PVALID # SIGNWRITING FILL MODIFIER-2..SIGNWRITING FILL
 1DAA1..1DAAF; PVALID # SIGNWRITING ROTATION MODIFIER-2..SIGNWRITING
 1F32D..1F32F; DISALLOWED # HOT DOG..BURRITO
 1F37E..1F37F; DISALLOWED # BOTTLE WITH POPPING CORK..POPCORN
 1F3CF..1F3D3; DISALLOWED # CRICKET BAT AND BALL..TABLE TENNIS PADDLE AND
 1F3F8..1F3FF; DISALLOWED # BADMINTON RACQUET AND SHUTTLECOCK..EMOJI MODI
 1F4FF ; DISALLOWED # PRAYER BEADS
 1F54B..1F54F; DISALLOWED # KAABA..BOWL OF HYGIEIA
 1F643..1F644; DISALLOWED # UPSIDE-DOWN FACE..FACE WITH ROLLING EYES
 1F6D0 ; DISALLOWED # PLACE OF WORSHIP
 1F910..1F918; DISALLOWED # ZIPPER-MOUTH FACE..SIGN OF THE HORNS
 1F980..1F984; DISALLOWED # CRAB..UNICORN FACE
 1F9C0 ; DISALLOWED # CHEESE WEDGE
 2B820..2CEA1; PVALID # <CJK Ideograph Extension E>..<CJK Ideograph E

[Appendix C](#). Changes from Unicode 8.0.0 to Unicode 9.0.0

Changes from derived property value UNASSIGNED to either PVALID or DISALLOWED.

08B6..08BD ; PVALID # ARABIC LETTER BEH WITH SMALL MEEM ABOVE..ARAB
 08D4..08E1 ; PVALID # ARABIC SMALL HIGH WORD AR-RUB..ARABIC SMALL H
 08E2 ; DISALLOWED # ARABIC DISPUTED END OF AYAH
 0C80 ; PVALID # KANNADA SIGN SPACING CANDRABINDU
 0D4F ; DISALLOWED # MALAYALAM SIGN PARA
 0D54..0D56 ; PVALID # MALAYALAM LETTER CHILLU M..MALAYALAM LETTER C
 0D58..0D5E ; DISALLOWED # MALAYALAM FRACTION ONE ONE-HUNDRED-AND-SIXTIE
 0D76..0D78 ; DISALLOWED # MALAYALAM FRACTION ONE SIXTEENTH..MALAYALAM F
 1C80..1C88 ; DISALLOWED # CYRILLIC SMALL LETTER ROUNDED VE..CYRILLIC SM
 1DFB ; PVALID # COMBINING DELETION MARK
 23FB..23FE ; DISALLOWED # POWER SYMBOL..POWER SLEEP SYMBOL
 2E43..2E44 ; DISALLOWED # DASH WITH LEFT UPTURN..DOUBLE SUSPENSION MARK
 A7AE ; DISALLOWED # LATIN CAPITAL LETTER SMALL CAPITAL I
 A8C5 ; PVALID # SAURASHTRA SIGN CANDRABINDU
 1018D..1018E; DISALLOWED # GREEK INDICTION SIGN..NOMISMA SIGN
 104B0..104D3; DISALLOWED # OSAGE CAPITAL LETTER A..OSAGE CAPITAL LETTER

104D8..104FB;	PVALID	# OSAGE SMALL LETTER A..OSAGE SMALL LETTER ZHA
1123E	; PVALID	# KHOJKI SIGN SUKUN
11400..1144A;	PVALID	# NEWA LETTER A..NEWA SIDDHI
1144B..1144F;	DISALLOWED	# NEWA DANDA..NEWA ABBREVIATION SIGN
11450..11459;	PVALID	# NEWA DIGIT ZERO..NEWA DIGIT NINE
1145B	; DISALLOWED	# NEWA PLACEHOLDER MARK
1145D	; DISALLOWED	# NEWA INSERTION SIGN
11660..1166C;	DISALLOWED	# MONGOLIAN BIRGA WITH ORNAMENT..MONGOLIAN TURN
11C00..11C08;	PVALID	# BHAIKSUKI LETTER A..BHAIKSUKI LETTER VOCALIC
11C0A..11C36;	PVALID	# BHAIKSUKI LETTER E..BHAIKSUKI VOWEL SIGN VOCA
11C38..11C40;	PVALID	# BHAIKSUKI VOWEL SIGN E..BHAIKSUKI SIGN AVAGRA
11C41..11C45;	DISALLOWED	# BHAIKSUKI DANDA..BHAIKSUKI GAP FILLER-2
11C50..11C59;	PVALID	# BHAIKSUKI DIGIT ZERO..BHAIKSUKI DIGIT NINE
11C5A..11C6C;	DISALLOWED	# BHAIKSUKI NUMBER ONE..BHAIKSUKI HUNDREDS UNIT
11C70..11C71;	DISALLOWED	# MARCHEN HEAD MARK..MARCHEN MARK SHAD
11C72..11C8F;	PVALID	# MARCHEN LETTER KA..MARCHEN LETTER A
11C92..11CA7;	PVALID	# MARCHEN SUBJOINED LETTER KA..MARCHEN SUBJOINE
11CA9..11CB6;	PVALID	# MARCHEN SUBJOINED LETTER YA..MARCHEN SIGN CAN
16FE0	; PVALID	# TANGUT ITERATION MARK
17000..187EC;	PVALID	# <Tangut Ideograph>..<Tangut Ideograph>
18800..18AF2;	PVALID	# TANGUT COMPONENT-001..TANGUT COMPONENT-755
1E000..1E006;	PVALID	# COMBINING GLAGOLITIC LETTER AZU..COMBINING GL
1E008..1E018;	PVALID	# COMBINING GLAGOLITIC LETTER ZEMLJA..COMBINING
1E01B..1E021;	PVALID	# COMBINING GLAGOLITIC LETTER SHTA..COMBINING G
1E023..1E024;	PVALID	# COMBINING GLAGOLITIC LETTER YU..COMBINING GLA
1E026..1E02A;	PVALID	# COMBINING GLAGOLITIC LETTER YO..COMBINING GLA
1E900..1E921;	DISALLOWED	# ADLAM CAPITAL LETTER ALIF..ADLAM CAPITAL LETT
1E922..1E94A;	PVALID	# ADLAM SMALL LETTER ALIF..ADLAM NUKTA
1E950..1E959;	PVALID	# ADLAM DIGIT ZERO..ADLAM DIGIT NINE
1E95E..1E95F;	DISALLOWED	# ADLAM INITIAL EXCLAMATION MARK..ADLAM INITIAL
1F19B..1F1AC;	DISALLOWED	# SQUARED THREE D..SQUARED VOD
1F23B	; DISALLOWED	# SQUARED CJK UNIFIED IDEOGRAPH-914D
1F57A	; DISALLOWED	# MAN DANCING
1F5A4	; DISALLOWED	# BLACK HEART
1F6D1..1F6D2;	DISALLOWED	# OCTAGONAL SIGN..SHOPPING TROLLEY
1F6F4..1F6F6;	DISALLOWED	# SCOOTER..CANOE
1F919..1F91E;	DISALLOWED	# CALL ME HAND..HAND WITH INDEX AND MIDDLE FING
1F920..1F927;	DISALLOWED	# FACE WITH COWBOY HAT..SNEEZING FACE
1F930	; DISALLOWED	# PREGNANT WOMAN
1F933..1F93E;	DISALLOWED	# SELFIE..HANDBALL
1F940..1F94B;	DISALLOWED	# WILTED FLOWER..MARTIAL ARTS UNIFORM
1F950..1F95E;	DISALLOWED	# CROISSANT..PANCAKES
1F985..1F991;	DISALLOWED	# EAGLE..SQUID

Internet-Draft

IDNA2008 and Unicode 12

December 2021

[Appendix D](#). Changes from Unicode 9.0.0 to Unicode 10.0.0

Changes from derived property value UNASSIGNED to either PVALID or DISALLOWED.

Internet-Draft

IDNA2008 and Unicode 12

December 2021

0860..086A	; PVALID	# SYRIAC LETTER MALAYALAM NGA..SYRIAC LETTER MA
09FC	; PVALID	# BENGALI LETTER VEDIC ANUSVARA
09FD	; DISALLOWED	# BENGALI ABBREVIATION SIGN
0AFA..0AFF	; PVALID	# GUJARATI SIGN SUKUN..GUJARATI SIGN TWO-CIRCLE
0D00	; PVALID	# MALAYALAM SIGN COMBINING ANUSVARA ABOVE
0D3B..0D3C	; PVALID	# MALAYALAM SIGN VERTICAL BAR VIRAMA..MALAYALAM
1CF7	; PVALID	# VEDIC SIGN ATIKRAMA
1DF6..1DF9	; PVALID	# COMBINING KAVYKA ABOVE RIGHT..COMBINING WIDE
20BF	; DISALLOWED	# BITCOIN SIGN
23FF	; DISALLOWED	# OBSERVER EYE SYMBOL
2BD2	; DISALLOWED	# GROUP MARK
2E45..2E49	; DISALLOWED	# INVERTED LOW KAVYKA..DOUBLE STACKED COMMA
312E	; PVALID	# BOPOMOFO LETTER O WITH DOT ABOVE
9FD6..9FEA	; PVALID	# <CJK Ideograph>..<CJK Ideograph>
1032D..1032F	; PVALID	# OLD ITALIC LETTER YE..OLD ITALIC LETTER SOUTH
11A00..11A3E	; PVALID	# ZANABAZAR SQUARE LETTER A..ZANABAZAR SQUARE C
11A3F..11A46	; DISALLOWED	# ZANABAZAR SQUARE INITIAL HEAD MARK..ZANABAZAR
11A47	; PVALID	# ZANABAZAR SQUARE SUBJOINER
11A50..11A83	; PVALID	# SOYOMBO LETTER A..SOYOMBO LETTER KSSA
11A86..11A99	; PVALID	# SOYOMBO CLUSTER-INITIAL LETTER RA..SOYOMBO SU
11A9A..11A9C	; DISALLOWED	# SOYOMBO MARK TSHEG..SOYOMBO MARK DOUBLE SHAD
11A9E..11AA2	; DISALLOWED	# SOYOMBO HEAD MARK WITH MOON AND SUN AND TRIPL
11D00..11D06	; PVALID	# MASARAM GONDI LETTER A..MASARAM GONDI LETTER
11D08..11D09	; PVALID	# MASARAM GONDI LETTER AI..MASARAM GONDI LETTER
11D0B..11D36	; PVALID	# MASARAM GONDI LETTER AU..MASARAM GONDI VOWEL
11D3A	; PVALID	# MASARAM GONDI VOWEL SIGN E
11D3C..11D3D	; PVALID	# MASARAM GONDI VOWEL SIGN AI..MASARAM GONDI VO
11D3F..11D47	; PVALID	# MASARAM GONDI VOWEL SIGN AU..MASARAM GONDI RA
11D50..11D59	; PVALID	# MASARAM GONDI DIGIT ZERO..MASARAM GONDI DIGIT
16FE1	; PVALID	# NUSHU ITERATION MARK
1B002..1B11E	; PVALID	# HENTAIGANA LETTER A-1..HENTAIGANA LETTER N-MU
1B170..1B2FB	; PVALID	# NUSHU CHARACTER-1B170..NUSHU CHARACTER-1B2FB
1F260..1F265	; DISALLOWED	# ROUNDED SYMBOL FOR FU..ROUNDED SYMBOL FOR CAI
1F6D3..1F6D4	; DISALLOWED	# STUPA..PAGODA
1F6F7..1F6F8	; DISALLOWED	# SLED..FLYING SAUCER

1F900..1F90B; DISALLOWED # CIRCLED CROSS FORMEE WITH FOUR DOTS..DOWNWARD
 1F91F ; DISALLOWED # I LOVE YOU HAND SIGN
 1F928..1F92F; DISALLOWED # FACE WITH ONE EYEBROW RAISED..SHOCKED FACE WI
 1F931..1F932; DISALLOWED # BREAST-FEEDING..PALMS UP TOGETHER
 1F94C ; DISALLOWED # CURLING STONE
 1F95F..1F96B; DISALLOWED # DUMPLING..CANNED FOOD
 1F992..1F997; DISALLOWED # GIRAFFE FACE..CRICKET
 1F9D0..1F9E6; DISALLOWED # FACE WITH MONOCLE..SOCKS
 2CEB0..2EBE0; PVALID # <CJK Ideograph Extension F>..<CJK Ideograph E

[Appendix E](#). Changes from Unicode 10.0.0 to Unicode 11.0.0

Changes from derived property value DISALLOWED to PVALID.

111C9 ; PVALID # SHARADA SANDHI MARK

Changes from derived property value UNASSIGNED to either PVALID or DISALLOWED.

[0560](#) ; PVALID # ARMENIAN SMALL LETTER TURNED AYB
[0588](#) ; PVALID # ARMENIAN SMALL LETTER YI WITH STROKE
 05EF ; PVALID # HEBREW YOD TRIANGLE
 07FD ; PVALID # NKO DANTAYALAN
 07FE..07FF ; DISALLOWED # NKO DOROME SIGN..NKO TAMAN SIGN
 08D3 ; PVALID # ARABIC SMALL LOW WAW
 09FE ; PVALID # BENGALI SANDHI MARK
 0A76 ; DISALLOWED # GURMUKHI ABBREVIATION SIGN
 0C04 ; PVALID # TELUGU SIGN COMBINING ANUSVARA ABOVE
 0C84 ; DISALLOWED # KANNADA SIGN SIDDHAM
[1878](#) ; PVALID # MONGOLIAN LETTER CHA WITH TWO DOTS
 1C90..1CBA ; DISALLOWED # GEORGIAN MTAVRULI CAPITAL LETTER AN..GEORGIAN
 1CBD..1CBF ; DISALLOWED # GEORGIAN MTAVRULI CAPITAL LETTER AEN..GEORGIA
 2BBA..2BBC ; DISALLOWED # OVERLAPPING WHITE SQUARES..OVERLAPPING BLACK
 2BD3..2BEB ; DISALLOWED # PLUTO FORM TWO..STAR WITH RIGHT HALF BLACK
 2BF0..2BFE ; DISALLOWED # ERIS FORM ONE..REVERSED RIGHT ANGLE
 2E4A..2E4E ; DISALLOWED # DOTTED SOLIDUS..PUNCTUS ELEVATUS MARK
 312F ; PVALID # BOPOMOFO LETTER NN
 9FEB..9FEF ; PVALID # <CJK Ideograph>..<CJK Ideograph>

A7AF	; PVALID	# LATIN LETTER SMALL CAPITAL Q
A7B8	; DISALLOWED	# LATIN CAPITAL LETTER U WITH STROKE
A7B9	; PVALID	# LATIN SMALL LETTER U WITH STROKE
A8FE..A8FF	; PVALID	# DEVANAGARI LETTER AY..DEVANAGARI VOWEL SIGN A
10A34..10A35	; PVALID	# KHAROSHTHI LETTER TTTA..KHAROSHTHI LETTER VHA
10A48	; DISALLOWED	# KHAROSHTHI FRACTION ONE HALF
10D00..10D27	; PVALID	# HANIFI ROHINGYA LETTER A..HANIFI ROHINGYA SIG
10D30..10D39	; PVALID	# HANIFI ROHINGYA DIGIT ZERO..HANIFI ROHINGYA D
10F00..10F1C	; PVALID	# OLD SOGDIAN LETTER ALEPH..OLD SOGDIAN LETTER
10F1D..10F26	; DISALLOWED	# OLD SOGDIAN NUMBER ONE..OLD SOGDIAN FRACTION
10F27	; PVALID	# OLD SOGDIAN LIGATURE AYIN-DALETH
10F30..10F50	; PVALID	# SOGDIAN LETTER ALEPH..SOGDIAN COMBINING STROK
10F51..10F59	; DISALLOWED	# SOGDIAN NUMBER ONE..SOGDIAN PUNCTUATION HALF
110CD	; DISALLOWED	# KAITHI NUMBER SIGN ABOVE
11144..11146	; PVALID	# CHAKMA LETTER LHAA..CHAKMA VOWEL SIGN EI
1133B	; PVALID	# COMBINING BINDU BELOW
1145E	; PVALID	# NEWA SANDHI MARK
1171A	; PVALID	# AHOM LETTER ALTERNATE BA
11800..1183A	; PVALID	# DOGRA LETTER A..DOGRA SIGN NUKTA
1183B	; DISALLOWED	# DOGRA ABBREVIATION SIGN

11A9D	; PVALID	# SOYOMBO MARK PLUTA
11D60..11D65	; PVALID	# GUNJALA GONDI LETTER A..GUNJALA GONDI LETTER
11D67..11D68	; PVALID	# GUNJALA GONDI LETTER EE..GUNJALA GONDI LETTER
11D6A..11D8E	; PVALID	# GUNJALA GONDI LETTER OO..GUNJALA GONDI VOWEL
11D90..11D91	; PVALID	# GUNJALA GONDI VOWEL SIGN EE..GUNJALA GONDI VO
11D93..11D98	; PVALID	# GUNJALA GONDI VOWEL SIGN OO..GUNJALA GONDI OM
11DA0..11DA9	; PVALID	# GUNJALA GONDI DIGIT ZERO..GUNJALA GONDI DIGIT
11EE0..11EF6	; PVALID	# MAKASAR LETTER KA..MAKASAR VOWEL SIGN O
11EF7..11EF8	; DISALLOWED	# MAKASAR PASSIMBANG..MAKASAR END OF SECTION
16E40 ..16E5F	; DISALLOWED	# MEDEFAIDRIN CAPITAL LETTER M..MEDEFAIDRIN CAP
16E60..16E7F	; PVALID	# MEDEFAIDRIN SMALL LETTER M..MEDEFAIDRIN SMALL
16E80..16E9A	; DISALLOWED	# MEDEFAIDRIN DIGIT ZERO..MEDEFAIDRIN EXCLAMATI
187ED..187F1	; PVALID	# <Tangut Ideograph>..<Tangut Ideograph>
1D2E0..1D2F3	; DISALLOWED	# MAYAN NUMERAL ZERO..MAYAN NUMERAL NINETEEN
1D372..1D378	; DISALLOWED	# IDEOGRAPHIC TALLY MARK ONE..TALLY MARK FIVE
1EC71..1ECB4	; DISALLOWED	# INDIC SIYAQ NUMBER ONE..INDIC SIYAQ ALTERNATE
1F12F	; DISALLOWED	# COPYLEFT SYMBOL
1F6F9	; DISALLOWED	# SKATEBOARD
1F7D5..1F7D8	; DISALLOWED	# CIRCLED TRIANGLE..NEGATIVE CIRCLED SQUARE
1F94D..1F94F	; DISALLOWED	# LACROSSE STICK AND BALL..FLYING DISC
1F96C..1F970	; DISALLOWED	# LEAFY GREEN..SMILING FACE WITH SMILING EYES A

1F973..1F976; DISALLOWED # FACE WITH PARTY HORN AND PARTY HAT..FREEZING
 1F97A ; DISALLOWED # FACE WITH PLEADING EYES
 1F97C..1F97F; DISALLOWED # LAB COAT..FLAT SHOE
 1F998..1F9A2; DISALLOWED # KANGAROO..SWAN
 1F9B0..1F9B9; DISALLOWED # EMOJI COMPONENT RED HAIR..SUPERVILLAIN
 1F9C1..1F9C2; DISALLOWED # CUPCAKE..SALT SHAKER
 1F9E7..1F9FF; DISALLOWED # RED GIFT ENVELOPE..NAZAR AMULET
 1FA60..1FA6D; DISALLOWED # XIANGQI RED GENERAL..XIANGQI BLACK SOLDIER

[Appendix F.](#) Changes from Unicode 11.0.0 to Unicode 12.0.0

Changes from derived property value UNASSIGNED to either PVALID or DISALLOWED.

0C77 ; DISALLOWED # TELUGU SIGN SIDDHAM
 0E86 ; PVALID # LAO LETTER PALI GHA
 0E89 ; PVALID # LAO LETTER PALI CHA
 0E8C ; PVALID # LAO LETTER PALI JHA
 0E8E..0E93 ; PVALID # LAO LETTER PALI NYA..LAO LETTER PALI NNA
 0E98 ; PVALID # LAO LETTER PALI DHA
 0EA0 ; PVALID # LAO LETTER PALI BHA
 0EA8..0EA9 ; PVALID # LAO LETTER SANSKRIT SHA..LAO LETTER SANSKRIT
 0EAC ; PVALID # LAO LETTER PALI LLA
 0EBA ; PVALID # LAO SIGN PALI VIRAMA
 1CFA ; PVALID # VEDIC SIGN DOUBLE ANUSVARA ANTARGOMUKHA
 2BC9 ; DISALLOWED # NEPTUNE FORM TWO
 2BFF ; DISALLOWED # HELLSCHREIBER PAUSE SYMBOL

2E4F ; DISALLOWED # CORNISH VERSE DIVIDER
 A7BA ; DISALLOWED # LATIN CAPITAL LETTER GLOTTAL A
 A7BB ; PVALID # LATIN SMALL LETTER GLOTTAL A
 A7BC ; DISALLOWED # LATIN CAPITAL LETTER GLOTTAL I
 A7BD ; PVALID # LATIN SMALL LETTER GLOTTAL I
 A7BE ; DISALLOWED # LATIN CAPITAL LETTER GLOTTAL U
 A7BF ; PVALID # LATIN SMALL LETTER GLOTTAL U
 A7C2 ; DISALLOWED # LATIN CAPITAL LETTER ANGLICANA W
 A7C3 ; PVALID # LATIN SMALL LETTER ANGLICANA W
 A7C4..A7C6 ; DISALLOWED # LATIN CAPITAL LETTER C WITH PALATAL HOOK..LAT
 AB66..AB67 ; PVALID # LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX
 10FE0..10FF6; PVALID # ELYMAIC LETTER ALEPH..ELYMAIC LIGATURE ZAYIN-
 1145F ; PVALID # NEWA LETTER VEDIC ANUSVARA
 116B8 ; PVALID # TAKRI LETTER ARCHAIC KHA

119A0..119A7; PVALID # NANDINAGARI LETTER A..NANDINAGARI LETTER VOCA
 119AA..119D7; PVALID # NANDINAGARI LETTER E..NANDINAGARI VOWEL SIGN
 119DA..119E1; PVALID # NANDINAGARI VOWEL SIGN E..NANDINAGARI SIGN AV
 119E2 ; DISALLOWED # NANDINAGARI SIGN SIDDHAM
 119E3..119E4; PVALID # NANDINAGARI HEADSTROKE..NANDINAGARI VOWEL SIG
 11A84..11A85; PVALID # SOYOMBO SIGN JIHVAMULIYA..SOYOMBO SIGN UPADHM
 11FC0..11FF1; DISALLOWED # TAMIL FRACTION ONE THREE-HUNDRED-AND-TWENTIET
 11FFF ; DISALLOWED # TAMIL PUNCTUATION END OF TEXT
 13430..13438; DISALLOWED # EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIAN
 16F45..16F4A; PVALID # MIAO LETTER BRI..MIAO LETTER RTE
 16F4F ; PVALID # MIAO SIGN CONSONANT MODIFIER BAR
 16F7F..16F87; PVALID # MIAO VOWEL SIGN UOG..MIAO VOWEL SIGN UI
 16FE2 ; DISALLOWED # OLD CHINESE HOOK MARK
 16FE3 ; PVALID # OLD CHINESE ITERATION MARK
 187F2..187F7; PVALID # <Tangut Ideograph>..<Tangut Ideograph>
 1B150..1B152; PVALID # HIRAGANA LETTER SMALL WI..HIRAGANA LETTER SMA
 1B164..1B167; PVALID # KATAKANA LETTER SMALL WI..KATAKANA LETTER SMA
 1E100..1E12C; PVALID # NYIAKENG PUACHUE HMONG LETTER MA..NYIAKENG PU
 1E130..1E13D; PVALID # NYIAKENG PUACHUE HMONG TONE-B..NYIAKENG PUACH
 1E140..1E149; PVALID # NYIAKENG PUACHUE HMONG DIGIT ZERO..NYIAKENG P
 1E14E ; PVALID # NYIAKENG PUACHUE HMONG LOGOGRAM NYAJ
 1E14F ; DISALLOWED # NYIAKENG PUACHUE HMONG CIRCLED CA
 1E2C0..1E2F9; PVALID # WANCHO LETTER AA..WANCHO DIGIT NINE
 1E2FF ; DISALLOWED # WANCHO NGUN SIGN
 1E94B ; PVALID # ADLAM NASALIZATION MARK
 1ED01..1ED3D; DISALLOWED # OTTOMAN SIYAQ NUMBER ONE..OTTOMAN SIYAQ FRACT
 1F16C ; DISALLOWED # RAISED MR SIGN
 1F6D5 ; DISALLOWED # HINDU TEMPLE
 1F6FA ; DISALLOWED # AUTO RICKSHAW
 1F7E0..1F7EB; DISALLOWED # LARGE ORANGE CIRCLE..LARGE BROWN SQUARE
 1F90D..1F90F; DISALLOWED # WHITE HEART..PINCHING HAND
 1F93F ; DISALLOWED # DIVING MASK
 1F971 ; DISALLOWED # YAWNING FACE
 1F97B ; DISALLOWED # SARI

1F9A5..1F9AA; DISALLOWED # SLOTH..OYSTER
 1F9AE..1F9AF; DISALLOWED # GUIDE DOG..PROBING CANE
 1F9BA..1F9BF; DISALLOWED # SAFETY VEST..MECHANICAL LEG
 1F9C3..1F9CA; DISALLOWED # BEVERAGE BOX..ICE CUBE
 1F9CD..1F9CF; DISALLOWED # STANDING PERSON..DEAF PERSON
 1FA00..1FA53; DISALLOWED # NEUTRAL CHESS KING..BLACK CHESS KNIGHT-BISHOP
 1FA70..1FA73; DISALLOWED # BALLET SHOES..SHORTS

1FA78..1FA7A; DISALLOWED # DROP OF BLOOD..STETHOSCOPE
1FA80..1FA82; DISALLOWED # YO-YO..PARACHUTE
1FA90..1FA95; DISALLOWED # RINGED PLANET..BANJO

Author's Address

Patrik Faltstrom
Netnod

Email: paf@netnod.se