

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 12, 2021

P. Faltstrom
Netnod
March 11, 2021

IDNA2008 and Unicode 12.0.0
draft-faltstrom-unicode12-02

Abstract

This document describes the changes between Unicode 6.2.0 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 September 12, 2021.

Internet-Draft

IDNA2008 and Unicode 12

March 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 6.2.0 and 12.0.0	7
3.1.	Changes between Unicode 6.2.0 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	8
3.4.	Changes between Unicode 11.0.0 and 12.0.0	10
4.	U+111C9 SHARADA SANDHI MARK	10
5.	Conclusion	10
6.	IANA Considerations	11
7.	Security Considerations	11
8.	Acknowledgements	11
9.	References	11
9.1.	Normative References	12
9.2.	Non-normative references	12
Appendix A.	Changes from Unicode 6.3.0 to Unicode 7.0.0	15
Appendix B.	Changes from Unicode 7.0.0 to Unicode 8.0.0	18
Appendix C.	Changes from Unicode 8.0.0 to Unicode 9.0.0	19
Appendix D.	Changes from Unicode 9.0.0 to Unicode 10.0.0	20
Appendix E.	Changes from Unicode 10.0.0 to Unicode 11.0.0	21
Appendix F.	Changes from Unicode 11.0.0 to Unicode 12.0.0	23
	Author's Address	24

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

Faltstrom

Expires September 12, 2021

[Page 2]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

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.
- 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 after Unicode 5.2.0 [[Unicode-5.2.0](#)] up to including Unicode 6.0.0 [[Unicode-6.0.0](#)], as described in [RFC 6452](#) [[RFC6452](#)]. The code points U+0CF1 and U+0CF2 had a derived property value change from DISALLOWED to PVALID while U+19DA had a change in derived property value from PVALID to DISALLOWED. They were examined in great detail and IETF concluded that the consensus is that no update was needed to [RFC 5892](#) [[RFC5892](#)] based on the changes made to the Unicode standard.

As described in [Section 3](#), more changes have been made to code points between Unicode version 6.0.0 and Unicode version 12.0.0 [[Unicode-12.0.0](#)] so that the derived property values have been changed in an incompatible way. This document concludes that no exceptions are to be added to [RFC 5892](#) [[RFC5892](#)] even though there are changes in the derived property value as a result of the changes made in Unicode between version 6.2.0 and 12.0.0.

Further, in 2015, the Internet Architecture Board (IAB) issued a statement [[IAB](#)] which requested 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](#)]. This document concludes that this code point is not to be added to the exception list either. It should be noted that the review on U+08A1 indicated that it 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 [RFC 8753](#) [[RFC8753](#)]. This clarification of the review process will impact review of Unicode versions after version 12.0.0.

[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).

Faltstrom

Expires September 12, 2021

[Page 4]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

- 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 4892](#) 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 version 3.2 of The Unicode Standard [[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, CONTEXTJ, and CONTEXTO 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.

3. Notable Changes Between Unicode 6.2.0 and 12.0.0

3.1. Changes between Unicode 6.2.0 and 7.0.0

Change in number of characters in each category:

Code points that changed derived property value: 0

PVALID changed from 97946 to 99867 (+1921)

UNASSIGNED changed from 864348 to 861509 (-2839)

CONTEXTJ did not change, at 2

CONTEXT0 did not change, at 25

DISALLOWED changed from 151791 to 152709 (+918)

TOTAL did not change, at 1114112

There are no changes made to Unicode between version 6.2.0 and 7.0.0 that impact IDNA2008 calculation of the derived property values.

The character ARABIC LETTER BEH WITH HAMZA ABOVE (U+08A1) was introduced in Unicode 7.0.0. This was discussed extensively in the IETF, and by the IAB in their statement [[IAB](#)] requesting the IETF to 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 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 ZANABZAR 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 ZANABZAR 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

CONTEXTO 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 6.2.0 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](#). (BackwardCompatible (G)). 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 editor proposes that such a BackwardCompatible rule NOT to be added to IDNA2008. This also ensures all sandhi marks being treated in an equal way.

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.

[5.](#) Conclusion

As described in [Section 3](#) and [Section 4](#), changes have been made to Unicode between version 6.2.0 and 12.0.0. Some changes to specific characters changed their derived property value, whereas other

Faltstrom

Expires September 12, 2021

[Page 10]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

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 of this document is to not add any exception rules to IDNA2008.

This document addresses only changes to Unicode between version 6.2.0 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 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 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 do 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](#).

[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

Faltstrom	Expires September 12, 2021	[Page 11]
-----------	----------------------------	-----------

Internet-Draft	IDNA2008 and Unicode 12	March 2021
----------------	-------------------------	------------

[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

[Changes-11.0.0]

The Unicode Consortium, "Unicode Standard Annex #44", Unicode Standard Annex #44, UNICODE CHARACTER DATABASE, Change History https://www.unicode.org/reports/tr44/tr44-21d4.html#Change_History, May 2018.

- [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.

Faltstrom	Expires September 12, 2021	[Page 12]
-----------	----------------------------	-----------

Internet-Draft	IDNA2008 and Unicode 12	March 2021
----------------	-------------------------	------------

- [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. Faeltsstroem, "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.

[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.

Faltstrom

Expires September 12, 2021

[Page 13]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

[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-6.3.0]

The Unicode Consortium, "The Unicode Standard, Version 6.3.0", The Unicode Standard, Version 6.3.0 ISBN 978-1-936213-08-5, September 2013.

[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.

[Unicode-9.0.0]

The Unicode Consortium, "The Unicode Standard, Version 9.0.0", The Unicode Standard, Version 9.0.0 ISBN 978-1-936213-13-9, June 2016.

[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 6.3.0 to Unicode 7.0.0

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

037F	; DISALLOWED	# GREEK CAPITAL LETTER YOT
0528..052F	; DISALLOWED	# CYRILLIC CAPITAL LETTER EN WITH LEFT HOOK..C
058D..058E	; DISALLOWED	# RIGHT-FACING ARMENIAN ETERNITY SIGN..LEFT-FA
0605	; DISALLOWED	# ARABIC NUMBER MARK ABOVE
08A1	; PVALID	# ARABIC LETTER BEH WITH HAMZA ABOVE
08AD..08B2	; PVALID	# ARABIC LETTER LOW ALEF..ARABIC LETTER ZAIN W
08FF	; PVALID	# ARABIC MARK SIDEWAYS NOON GHUNNA
0978	; PVALID	# DEVANAGARI LETTER MARWARI DDA
0980	; PVALID	# BENGALI ANJI
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
16F1..16F8	; PVALID	# RUNIC LETTER K..RUNIC LETTER FRANKS CASKET A
191D..191E	; PVALID	# LIMBU LETTER GYAN..LIMBU LETTER TRA
1AB0..1ABE	; PVALID	# COMBINING DOUBLED CIRCUMFLEX ACCENT..COMBINI
1CF8..1CF9	; PVALID	# VEDIC TONE RING ABOVE..VEDIC TONE DOUBLE RIN
1DE7..1DF5	; PVALID	# COMBINING LATIN SMALL LETTER ALPHA..COMBININ
20BB..20BD	; DISALLOWED	# NORDIC MARK SIGN..RUBLE SIGN
23F4..23FA	; DISALLOWED	# BLACK MEDIUM LEFT-POINTING TRIANGLE..BLACK C
2700	; DISALLOWED	# BLACK SAFETY SCISSORS
2B4D..2B4F	; DISALLOWED	# DOWNWARDS TRIANGLE-HEADED ZIGZAG ARROW..SHOR
2B5A..2B73	; DISALLOWED	# SLANTED NORTH ARROW WITH HOOKED HEAD..DOWNWA
2B76..2B95	; DISALLOWED	# NORTH WEST TRIANGLE-HEADED ARROW TO BAR..RIG
2B98..2BB9	; DISALLOWED	# THREE-D TOP-LIGHTED LEFTWARDS EQUILATERAL AR
2BBD..2BC8	; DISALLOWED	# BALLOT BOX WITH LIGHT X..BLACK MEDIUM RIGHT-
2BCA..2BD1	; DISALLOWED	# TOP HALF BLACK CIRCLE..UNCERTAINTY SIGN
2E3C..2E42	; DISALLOWED	# STENOGRAPHIC FULL STOP..DOUBLE LOW-REVERSED-
A698..A69D	; DISALLOWED	# CYRILLIC CAPITAL LETTER DOUBLE O..MODIFIER L
A794..A79F	; PVALID	# LATIN SMALL LETTER C WITH PALATAL HOOK..LATI
A7AB..A7AD	; DISALLOWED	# LATIN CAPITAL LETTER REVERSED OPEN E..LATIN
A7B0..A7B1	; DISALLOWED	# LATIN CAPITAL LETTER TURNED K..LATIN CAPITAL
A7F7	; PVALID	# LATIN EPIGRAPHIC LETTER SIDEWAYS I
A9E0..A9FE	; PVALID	# MYANMAR LETTER SHAN GHA..MYANMAR LETTER TAI
AA7C..AA7F	; PVALID	# MYANMAR SIGN TAI LAING TONE-2..MYANMAR LETTE
AB30..AB5F	; PVALID	# LATIN SMALL LETTER BARRED ALPHA..MODIFIER LE
AB64..AB65	; PVALID	# LATIN SMALL LETTER INVERTED ALPHA..GREEK LET
FE27..FE2D	; PVALID	# COMBINING LIGATURE LEFT HALF BELOW..COMBININ
1018B..1018C	; DISALLOWED	# GREEK ONE QUARTER SIGN..GREEK SINUSOID SIGN
101A0	; DISALLOWED	# GREEK SYMBOL TAU RHO
102E0..102FB	; PVALID	# COPTIC EPACT THOUSANDS MARK..COPTIC EPACT NU
1031F	; PVALID	# OLD ITALIC LETTER ESS

10350..1037A; PVALID	# OLD PERMIC LETTER AN..COMBINING OLD PERMIC L
10500..10527; PVALID	# ELBASAN LETTER A..ELBASAN LETTER KHE
10530..10563; PVALID	# CAUCASIAN ALBANIAN LETTER ALT..CAUCASIAN ALB
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..1089E; PVALID	# PALMYRENE LETTER ALEPH..NABATAEAN LETTER TAW
108A7..108AF; DISALLOWED	# NABATAEAN NUMBER ONE..NABATAEAN NUMBER ONE H
10A80..10A9F; PVALID	# OLD NORTH ARABIAN LETTER HEH..OLD NORTH ARAB
10AC0..10AE6; PVALID	# MANICHAEAN LETTER ALEPH..MANICHAEAN ABBREVIA
10AEB..10AF6; DISALLOWED	# MANICHAEAN NUMBER ONE..MANICHAEAN PUNCTUATIO
10B80..10B91; PVALID	# PSALTER PAHLAVI LETTER ALEPH..PSALTER PAHLAV
10B99..10B9C; DISALLOWED	# PSALTER PAHLAVI SECTION MARK..PSALTER PAHLAV
10BA9..10BAF; DISALLOWED	# PSALTER PAHLAVI NUMBER ONE..PSALTER PAHLAVI
1107F ; PVALID	# BRAHMI NUMBER JOINER
11150..11176; PVALID	# MAHAJANI LETTER A..MAHAJANI LIGATURE SHRI
111CD ; DISALLOWED	# SHARADA SUTRA MARK
111DA ; PVALID	# SHARADA EKAM
111E1..111F4; DISALLOWED	# SINHALA ARCHAIC DIGIT ONE..SINHALA ARCHAIC N
11200..11211; PVALID	# KHOJKI LETTER A..KHOJKI LETTER JJA
11213..1123D; PVALID	# KHOJKI LETTER NYA..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 VISAR
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
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 VOCAL
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 VOCAL
11366..1136C; PVALID	# COMBINING GRANTHA DIGIT ZERO..COMBINING GRAN
11370..11374; PVALID	# COMBINING GRANTHA LETTER A..COMBINING GRANTH
11480..114C7; PVALID	# TIRHUTA ANJI..TIRHUTA OM
114D0..114D9; PVALID	# TIRHUTA DIGIT ZERO..TIRHUTA DIGIT NINE
11580..115B5; PVALID	# SIDDHAM LETTER A..SIDDHAM VOWEL SIGN VOCALIC
115B8..115C9; PVALID	# SIDDHAM VOWEL SIGN E..SIDDHAM END OF TEXT MA
11600..11644; PVALID	# MODI LETTER A..MODI SIGN HUVA
11650..11659; PVALID	# MODI DIGIT ZERO..MODI DIGIT NINE
118A0..118F2; DISALLOWED	# WARANG CITI CAPITAL LETTER NGAA..WARANG CITI
118FF ; PVALID	# WARANG CITI OM
11AC0..11AF8; PVALID	# PAU CIN HAU LETTER PA..PAU CIN HAU GLOTTAL S

1236F..12398; PVALID # CUNEIFORM SIGN KAP ELAMITE..CUNEIFORM SIGN U

Faltstrom

Expires September 12, 2021

[Page 16]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

12463..1246E; DISALLOWED # CUNEIFORM NUMERIC SIGN ONE QUARTER GUR..CUNE
[12474](#) ; DISALLOWED # CUNEIFORM PUNCTUATION SIGN DIAGONAL QUADCOLO
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..16AF5; PVALID # BASSA VAH COMBINING HIGH TONE..BASSA VAH FUL
16B00..16B45; PVALID # PAHAHW HMONG VOWEL KEEB..PAHAHW HMONG SIGN C
16B50..16B59; PVALID # PAHAHW HMONG DIGIT ZERO..PAHAHW HMONG DIGIT
16B5B..16B61; DISALLOWED # PAHAHW HMONG NUMBER TENS..PAHAHW HMONG NUMBE
16B63..16B77; PVALID # PAHAHW HMONG SIGN VOS LUB..PAHAHW HMONG SIGN
16B7D..16B8F; PVALID # PAHAHW HMONG CLAN SIGN TSHEEJ..PAHAHW HMONG
1BC00..1BC6A; PVALID # DUPLOYAN LETTER H..DUPLOYAN LETTER VOCALIC M
1BC70..1BC7C; PVALID # DUPLOYAN AFFIX LEFT HORIZONTAL SECANT..DUPLO
1BC80..1BC88; PVALID # DUPLOYAN AFFIX HIGH ACUTE..DUPLOYAN AFFIX HI
1BC90..1BC99; PVALID # DUPLOYAN AFFIX LOW ACUTE..DUPLOYAN AFFIX LOW
1BC9C..1BCA3; DISALLOWED # DUPLOYAN SIGN O WITH CROSS..SHORTHAND FORMAT
1E800..1E8C4; PVALID # MENDE KIKAKUI SYLLABLE M001 KI..MENDE KIKAKU
1E8C7..1E8D6; DISALLOWED # MENDE KIKAKUI DIGIT ONE..MENDE KIKAKUI COMBI
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..DINGB
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 TICKET
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..1F53F; DISALLOWED # LOWER RIGHT SHADOWED WHITE CIRCLE..UPPER RIG
1F544..1F54A; DISALLOWED # NOTCHED RIGHT SEMICIRCLE WITH THREE DOTS..DO
1F568..1F579; DISALLOWED # RIGHT SPEAKER..JOYSTICK
1F57B..1F5A3; DISALLOWED # LEFT HAND TELEPHONE RECEIVER..BLACK DOWN POI
1F5A5..1F5FA; DISALLOWED # DESKTOP COMPUTER..WORLD MAP
1F641..1F642; DISALLOWED # SLIGHTLY FROWNING FACE..SLIGHTLY SMILING FAC

1F650..1F67F;	DISALLOWED	# NORTH WEST POINTING LEAF..REVERSE CHECKER BO
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 ARROWHEA
1F810..1F847;	DISALLOWED	# LEFTWARDS ARROW WITH SMALL EQUILATERAL ARROW
1F850..1F859;	DISALLOWED	# LEFTWARDS SANS-SERIF ARROW..UP DOWN SANS-SER

Faltstrom

Expires September 12, 2021

[Page 17]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

1F860..1F887;	DISALLOWED	# WIDE-HEADED LEFTWARDS LIGHT BARB ARROW..WIDE
---------------	------------	--

[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..ARA
08E3	; PVALID	# ARABIC TURNED DAMMA BELOW
0AF9	; PVALID	# GUJARATI LETTER ZHA
0C5A	; PVALID	# TELUGU LETTER RRRA
0D5F	; PVALID	# MALAYALAM LETTER ARCHAIC II
13F5	; PVALID	# CHEROKEE LETTER MV
13F8..13FD	; DISALLOWED	# CHEROKEE SMALL LETTER YE..CHEROKEE SMALL LET
20BE	; DISALLOWED	# LARI SIGN
218A..218B	; DISALLOWED	# TURNED DIGIT TWO..TURNED DIGIT THREE
2BEC..2BEF	; DISALLOWED	# LEFTWARDS TWO-HEADED ARROW WITH TRIANGLE ARR
9FCD..9FD5	; PVALID	# <CJK Ideograph>..<CJK Ideograph>
A69E	; PVALID	# COMBINING CYRILLIC LETTER EF
A78F	; PVALID	# LATIN LETTER SINOLOGICAL DOT
A7B2..A7B7	; DISALLOWED	# LATIN CAPITAL LETTER J WITH CROSSED-TAIL..LA
A8FC..A8FD	; DISALLOWED	# DEVANAGARI SIGN SIDDHAM..DEVANAGARI JAIN OM
AB60..AB63	; PVALID	# LATIN SMALL LETTER SAKHA YAT..LATIN SMALL LE
AB70..ABBF	; DISALLOWED	# CHEROKEE SMALL LETTER A..CHEROKEE SMALL LETT
FE2E..FE2F	; PVALID	# COMBINING CYRILLIC TITLO LEFT HALF..COMBININ
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..M
109C0..109CF;	DISALLOWED	# MEROITIC CURSIVE NUMBER ONE..MEROITIC CURSIV
109D2..109FF;	DISALLOWED	# MEROITIC CURSIVE NUMBER ONE HUNDRED..MEROITI
10C80..10CB2;	DISALLOWED	# OLD HUNGARIAN CAPITAL LETTER A..OLD HUNGARIA
10CC0..10CF2;	PVALID	# OLD HUNGARIAN SMALL LETTER A..OLD HUNGARIAN

10CFA..10CFF;	DISALLOWED	# OLD HUNGARIAN NUMBER ONE..OLD HUNGARIAN NUMB
111C9..111CC;	DISALLOWED	# SHARADA SANDHI MARK..SHARADA EXTRA SHORT VOW
111DB..111DF;	DISALLOWED	# SHARADA SIGN SIDDHAM..SHARADA SECTION MARK-2
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..112A9;	PVALID	# MULTANI LETTER BHA..MULTANI SECTION MARK
11300	; PVALID	# GRANTHA SIGN COMBINING ANUSVARA ABOVE
11350	; PVALID	# GRANTHA OM
115CA..115DD;	DISALLOWED	# SIDDHAM SECTION MARK WITH TRIDENT AND U-SHAP
11700..11719;	PVALID	# AHOM LETTER KA..AHOM LETTER JHA
1171D..1172B;	PVALID	# AHOM CONSONANT SIGN MEDIAL LA..AHOM SIGN KIL
11730..1173F;	PVALID	# AHOM DIGIT ZERO..AHOM SYMBOL VI
12399	; PVALID	# CUNEIFORM SIGN U U

Faltstrom

Expires September 12, 2021

[Page 18]

Internet-Draft

IDNA2008 and Unicode 12

March 2021

12480..12543;	PVALID	# CUNEIFORM SIGN AB TIMES NUN TENU..CUNEIFORM
14400..14646;	PVALID	# ANATOLIAN HIEROGLYPH A001..ANATOLIAN HIEROGL
1D1DE..1D1E8;	DISALLOWED	# MUSICAL SYMBOL KIEVAN C CLEF..MUSICAL SYMBOL
1D800..1DA8B;	DISALLOWED	# SIGNWRITING HAND-FIST INDEX..SIGNWRITING PAR
1DA9B..1DA9F;	PVALID	# SIGNWRITING FILL MODIFIER-2..SIGNWRITING FIL
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 AN
1F3F8..1F3FF;	DISALLOWED	# BADMINTON RACQUET AND SHUTTLECOCK..EMOJI MOD
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

[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..ARA
08D4..08E2	; PVALID	# ARABIC SMALL HIGH WORD AR-RUB..ARABIC DISPUT
0C80	; PVALID	# KANNADA SIGN SPACING CANDRABINDU

0D4F	; DISALLOWED	# MALAYALAM SIGN PARA
0D54..0D56	; PVALID	# MALAYALAM LETTER CHILLU M..MALAYALAM LETTER
0D58..0D5E	; DISALLOWED	# MALAYALAM FRACTION ONE ONE-HUNDRED-AND-SIXTI
0D76..0D78	; DISALLOWED	# MALAYALAM FRACTION ONE SIXTEENTH..MALAYALAM
1C80..1C88	; DISALLOWED	# CYRILLIC SMALL LETTER ROUNDED VE..CYRILLIC S
1DFB	; PVALID	# COMBINING DELETION MARK
23FB..23FE	; DISALLOWED	# POWER SYMBOL..POWER SLEEP SYMBOL
2E43..2E44	; DISALLOWED	# DASH WITH LEFT UPTURN..DOUBLE SUSPENSION MAR
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..11459	; PVALID	# NEWA LETTER A..NEWA DIGIT NINE
1145B	; DISALLOWED	# NEWA PLACEHOLDER MARK
1145D	; DISALLOWED	# NEWA INSERTION SIGN
11660..1166C	; DISALLOWED	# MONGOLIAN BIRGA WITH ORNAMENT..MONGOLIAN TUR
11C00..11C08	; PVALID	# BHAIKSUKI LETTER A..BHAIKSUKI LETTER VOCALIC
11C0A..11C36	; PVALID	# BHAIKSUKI LETTER E..BHAIKSUKI VOWEL SIGN VOC
11C38..11C45	; PVALID	# BHAIKSUKI VOWEL SIGN E..BHAIKSUKI GAP FILLER
11C50..11C6C	; PVALID	# BHAIKSUKI DIGIT ZERO..BHAIKSUKI HUNDREDS UNI

11C70..11C8F	; DISALLOWED	# MARCHEN HEAD MARK..MARCHEN LETTER A
11C92..11CA7	; PVALID	# MARCHEN SUBJOINED LETTER KA..MARCHEN SUBJOIN
11CA9..11CB6	; PVALID	# MARCHEN SUBJOINED LETTER YA..MARCHEN SIGN CA
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 G
1E008..1E018	; PVALID	# COMBINING GLAGOLITIC LETTER ZEMLJA..COMBININ
1E01B..1E021	; PVALID	# COMBINING GLAGOLITIC LETTER SHTA..COMBINING
1E023..1E024	; PVALID	# COMBINING GLAGOLITIC LETTER YU..COMBINING GL
1E026..1E02A	; PVALID	# COMBINING GLAGOLITIC LETTER YO..COMBINING GL
1E900..1E94A	; DISALLOWED	# ADLAM CAPITAL LETTER ALIF..ADLAM NUKTA
1E950..1E959	; PVALID	# ADLAM DIGIT ZERO..ADLAM DIGIT NINE
1E95E..1E95F	; DISALLOWED	# ADLAM INITIAL EXCLAMATION MARK..ADLAM INITIA
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 FIN
 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

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

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

0860..086A ; PVALID # SYRIAC LETTER MALAYALAM NGA..SYRIAC LETTER M
 09FC..09FD ; PVALID # BENGALI LETTER VEDIC ANUSVARA..BENGALI ABBRE
 0AFA..0AFF ; PVALID # GUJARATI SIGN SUKUN..GUJARATI SIGN TWO-CIRCL
 0D00 ; PVALID # MALAYALAM SIGN COMBINING ANUSVARA ABOVE
 0D3B..0D3C ; PVALID # MALAYALAM SIGN VERTICAL BAR VIRAMA..MALAYALA
 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 SOUT
11A00..11A47	; PVALID	# ZANABAZAR SQUARE LETTER A..ZANABAZAR SQUARE
11A50..11A83	; PVALID	# SOYOMBO LETTER A..SOYOMBO LETTER KSSA
11A86..11A9C	; PVALID	# SOYOMBO CLUSTER-INITIAL LETTER RA..SOYOMBO M
11A9E..11AA2	; DISALLOWED	# SOYOMBO HEAD MARK WITH MOON AND SUN AND TRIP
11D00..11D06	; PVALID	# MASARAM GONDI LETTER A..MASARAM GONDI LETTER
11D08..11D09	; PVALID	# MASARAM GONDI LETTER AI..MASARAM GONDI LETTE
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 V
11D3F..11D47	; PVALID	# MASARAM GONDI VOWEL SIGN AU..MASARAM GONDI R
11D50..11D59	; PVALID	# MASARAM GONDI DIGIT ZERO..MASARAM GONDI DIGI
16FE1	; PVALID	# NUSHU ITERATION MARK
1B002..1B11E	; PVALID	# HENTAIGANA LETTER A-1..HENTAIGANA LETTER N-M
1B170..1B2FB	; PVALID	# NUSHU CHARACTER-1B170..NUSHU CHARACTER-1B2FB
1F260..1F265	; DISALLOWED	# ROUNDED SYMBOL FOR FU..ROUNDED SYMBOL FOR CA
1F6D3..1F6D4	; DISALLOWED	# STUPA..PAGODA
1F6F7..1F6F8	; DISALLOWED	# SLED..FLYING SAUCER
1F900..1F90B	; DISALLOWED	# CIRCLED CROSS FORMEE WITH FOUR DOTS..DOWNWAR
1F91F	; DISALLOWED	# I LOVE YOU HAND SIGN
1F928..1F92F	; DISALLOWED	# FACE WITH ONE EYEBROW RAISED..SHOCKED FACE W
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

[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..07FF	; PVALID	# NKO DANTAYALAN..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..GEORGIA
1CBD..1CBF	; DISALLOWED	# GEORGIAN MTAVRULI CAPITAL LETTER AEN..GEORGI
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..A7B9	; DISALLOWED	# LATIN CAPITAL LETTER U WITH STROKE..LATIN SM
A8FE..A8FF	; PVALID	# DEVANAGARI LETTER AY..DEVANAGARI VOWEL SIGN
10A34..10A35	; PVALID	# KHAROSHTHI LETTER TTTA..KHAROSHTHI LETTER VH
10A48	; DISALLOWED	# KHAROSHTHI FRACTION ONE HALF
10D00..10D27	; PVALID	# HANIFI ROHINGYA LETTER A..HANIFI ROHINGYA SI
10D30..10D39	; PVALID	# HANIFI ROHINGYA DIGIT ZERO..HANIFI ROHINGYA
10F00..10F27	; PVALID	# OLD SOGDIAN LETTER ALEPH..OLD SOGDIAN LIGATU
10F30..10F59	; PVALID	# SOGDIAN LETTER ALEPH..SOGDIAN PUNCTUATION HA
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..1183B	; PVALID	# DOGRA LETTER A..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 LETTE
11D6A..11D8E	; PVALID	# GUNJALA GONDI LETTER OO..GUNJALA GONDI VOWEL
11D90..11D91	; PVALID	# GUNJALA GONDI VOWEL SIGN EE..GUNJALA GONDI V
11D93..11D98	; PVALID	# GUNJALA GONDI VOWEL SIGN OO..GUNJALA GONDI O
11DA0..11DA9	; PVALID	# GUNJALA GONDI DIGIT ZERO..GUNJALA GONDI DIGI
11EE0..11EF8	; PVALID	# MAKASAR LETTER KA..MAKASAR END OF SECTION
16E40 ..16E9A	; DISALLOWED	# MEDEFAIDRIN CAPITAL LETTER M..MEDEFAIDRIN EX
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 ALTERNAT
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
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

[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..0C7F	; DISALLOWED	# TELUGU SIGN SIDDHAM..TELUGU SIGN TUUMU
0E86..0E8A	; PVALID	# LAO LETTER PALI GHA..LAO LETTER SO TAM
0E8C..0EA3	; PVALID	# LAO LETTER PALI JHA..LAO LETTER LO LING
0EA7..0EB2	; PVALID	# LAO LETTER WO..LAO VOWEL SIGN AA
0EB4..0EBD	; PVALID	# LAO VOWEL SIGN I..LAO SEMIVOWEL SIGN NYO
1CD4..1CFA	; PVALID	# VEDIC SIGN YAJURVEDIC MIDLINE SVARITA..VEDIC
2B98..2C2E	; DISALLOWED	# THREE-D TOP-LIGHTED LEFTWARDS EQUILATERAL AR
2E30..2E4F	; DISALLOWED	# RING POINT..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..LA
AB60..AB67	; PVALID	# LATIN SMALL LETTER SAKHA YAT..LATIN SMALL LE
10FE0..10FF6;	PVALID	# ELYMAIC LETTER ALEPH..ELYMAIC LIGATURE ZAYIN
1145E..1145F;	PVALID	# NEWA SANDHI MARK..NEWA LETTER VEDIC ANUSVARA
11680..116B8;	PVALID	# TAKRI LETTER A..TAKRI LETTER ARCHAIC KHA
119A0..119A7;	PVALID	# NANDINAGARI LETTER A..NANDINAGARI LETTER VOC
119AA..119D7;	PVALID	# NANDINAGARI LETTER E..NANDINAGARI VOWEL SIGN
119DA..119E1;	PVALID	# NANDINAGARI VOWEL SIGN E..NANDINAGARI SIGN A
119E2	; DISALLOWED	# NANDINAGARI SIGN SIDDHAM
119E3..119E4;	PVALID	# NANDINAGARI HEADSTROKE..NANDINAGARI VOWEL SI
11A50..11A99;	PVALID	# SOYOMBO LETTER A..SOYOMBO SUBJOINER
11FC0..11FF1;	DISALLOWED	# TAMIL FRACTION ONE THREE-HUNDRED-AND-TWENTIE
11FFF	; DISALLOWED	# TAMIL PUNCTUATION END OF TEXT
13430..13438;	DISALLOWED	# EGYPTIAN HIEROGLYPH VERTICAL JOINER..EGYPTIA
16F00..16F4A;	PVALID	# MIAO LETTER PA..MIAO LETTER RTE
16F4F..16F87;	PVALID	# MIAO SIGN CONSONANT MODIFIER BAR..MIAO VOWEL

Internet-Draft

IDNA2008 and Unicode 12

March 2021

```
16FE2      ; DISALLOWED # OLD CHINESE HOOK MARK
16FE3      ; PVALID     # OLD CHINESE ITERATION MARK
17000..187F7; PVALID     # <Tangut Ideograph>..<Tangut Ideograph>
1B150..1B152; PVALID     # HIRAGANA LETTER SMALL WI..HIRAGANA LETTER SM
1B164..1B167; PVALID     # KATAKANA LETTER SMALL WI..KATAKANA LETTER SM
1E100..1E12C; PVALID     # NYIAKENG PUACHUE HMONG LETTER MA..NYIAKENG P
1E130..1E13D; PVALID     # NYIAKENG PUACHUE HMONG TONE-B..NYIAKENG PUAC
1E140..1E149; PVALID     # NYIAKENG PUACHUE HMONG DIGIT ZERO..NYIAKENG
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
1E922..1E94B; PVALID     # ADLAM SMALL LETTER ALIF..ADLAM NASALIZATION
1ED01..1ED3D; DISALLOWED # OTTOMAN SIYAQ NUMBER ONE..OTTOMAN SIYAQ FRAC
1F110..1F16C; DISALLOWED # PARENTHESES L A T I N CAPITAL LETTER A..RAISED
1F300..1F6D5; DISALLOWED # CYCLONE..HINDU TEMPLE
1F6F0..1F6FA; DISALLOWED # SATELLITE..AUTO RICKSHAW
1F7E0..1F7EB; DISALLOWED # LARGE ORANGE CIRCLE..LARGE BROWN SQUARE
1F90D..1F971; DISALLOWED # WHITE HEART..YAWNING FACE
1F97A..1F9A2; DISALLOWED # FACE WITH PLEADING EYES..SWAN
1F9A5..1F9AA; DISALLOWED # SLOTH..OYSTER
1F9AE..1F9CA; DISALLOWED # GUIDE DOG..ICE CUBE
1F9CD..1FA53; DISALLOWED # STANDING PERSON..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

Faltstrom

Expires September 12, 2021

[Page 24]