

.WS Emoji IDN

VERSION 11.0.0.1

Last updated: October 10, 2018

Contents

Introduction	4
Abstract	4
Background	4
Registrar Requirements	4
Version	4
Versioning	4
This Version of the Specification	5
Version History	5
Release Cycle for Version 11	5
Emoji Character Support	6
Definition	6
Emoji Translation	6
U+FE0F: "VARIATION SELECTOR-16"	6
U+200D: "ZERO WIDTH JOINER"	6
Disallowed Emoji	6
Homography	7
Blacklisted Emoji	7
Glyph Facing Direction	7
Combining Characters	8
Example Domain Reference	9
Valid Emoji Domains	9
Invalid Emoji Domains	9
EPP Usage	9
Standard EPP Considerations	9
EPP Fee Extension Considerations	9
Changes to EPP Responses	10
Client Validation	10
EPP Examples	10
Valid Emoji CHECK Response	10
Invalid Emoji CHECK Response	10
Server Validation and Hints	11

Appendix.....	12
.WS Emoji IDN Chart	12
EPP Processing Flowchart	13
References.....	14

Introduction

Abstract

This document aims to explain how a registrar can adopt the usage of Emoji IDN with the .WS Registry. The implementation of Emoji domains is based on the UTS #46 algorithm, namely the conversion of Emoji Unicode code points into ASCII using the Punycode algorithm.

Background

With the growth of Emoji becoming a means of communication, the Unicode Consortium has implemented a standardized method of representing such symbols. These symbols are represented by a single Unicode code point to establish an Emoji (such as U+1F609 🙄 winking face), or multiple code points in combination to produce a more elaborate Emoji (such as U+1F466 U+1F3FB 🧑 boy: light skin tone). More information regarding Emoji is available at Unicode's website:

http://www.unicode.org/faq/emoji_dingbats.html .

Registrar Requirements

Implementation of .WS Emoji IDN is not required of .WS registrars. However, we do recommend supporting such domains. For many registrars, no changes may be necessary. If your system is performing minimal validation against domain labels prior to performing an EPP transaction, then .WS Emoji IDN may immediately work. However, if you are performing validation for labels as defined by IDNA2003, IDNA2008, or UTS #46, then modifications may be necessary. *Be advised that the .WS Registry checks conformance to the specifications outlined in this document, so unsupported domains will be blocked (e.g. return as "unavailable" via EPP).*

Version

At the time of writing, this document refers to the following versioned references:

Unicode version 11.0.0 (as released on June 5, 2018):

The Unicode Consortium. The Unicode Standard, Version 11.0.0, (Mountain View, CA: The Unicode Consortium, 2018. ISBN 978-1-936213-19-1)

<https://www.unicode.org/versions/Unicode11.0.0/>

Unicode Emoji, Version 11.0 (as released on June 5, 2018):

Unicode Technical Report #51, "Unicode Emoji," by Mark Davis and Peter Edberg. 2018-05-21

<https://www.unicode.org/reports/tr51/tr51-14.html>

Versioning

After version 1.0, the version number of this document will correlate to the version number used by the implemented standard as defined by Unicode. Additionally, a fourth versioning number will be added for internal revisions for the WS Emoji IDN Specification.

Versioning are then separated into four numbers:

1. Major version of the Unicode Standard

2. Minor version of the Unicode Standard
3. Update version of the Unicode Standard
4. Supplementary version of the WS Emoji IDN Specification

For example, WS Emoji IDN Specification version 11.0.0.0 refers to our first implementation of the Unicode version 11.0.0 . WS Emoji IDN Specification 12.3.9.1 would refer to the theoretical 1st update to the WS Emoji IDN Specification (previously 12.3.9.0) which implements theoretical Unicode version 12.3.9 .

This Version of the Specification

This specification has been bumped to version 11.0.0.0 (rather than 2.0). The reasoning of this is to pin this document's version against the version number used by Unicode for their technical standard.

Version History

.WS Emoji IDN version 11.0.0.1 (This document, October 10, 2018)

- Document updated to reference now-released Unicode 11.0.0 , as opposed to beta.
- Unicode Standard references updated to 11.0.0, and UTS#51 references updated to 11.0 .
- Minor grammatical fixes.
- Language updated to be more readable to those not familiar with registry/registrar operations, or EPP in general.

.WS Emoji IDN version 11.0.0.0 (January 16, 2018)

- Adds support for new Emoji of Unicode 11.0
 - o All newly-introduced Emoji are supported except for 6 Emoji: superhero and supervillain variants (genderless, male, and female) due to potential homography
- Disallows support of Emoji Glyph Facing Direction
- Disallows support of skin tone modifiers for some Emoji that have "unrealistic" skin tones

.WS Emoji IDN version 1.0 – Initial

Release Cycle for Version 11

The following chart shows the intended release cycle for new versions of the .WS Emoji Domains specification. The release cycle is closely mirrored to that of the Unicode Standard and Technical Standards. For Version 11, the following releases are scheduled.

2017 Q4

- Unicode adds Draft Emoji Candidates from Emoji 5.0 into Emoji 11.0 Beta
- WS Registry releases an update through EPP OT&E channels with updates that conform to Unicode Emoji 11.0 beta (this document).

2018 Q1

- Unicode announces finalized list of Emoji to be introduced in Emoji 11.0
- WS Registry introduces Emoji 11.0 to production EPP channels, as well as through website.ws .

Exact release dates are decided by Unicode. To follow news, announcements, and other updates from Unicode, we recommend checking the Unicode Blog: <http://blog.unicode.org> .

Emoji Character Support

Definition

The allowed Emoji are defined by the latest released version of Unicode. The latest version is available at <http://unicode.org/emoji/charts/emoji-ordering.txt> . However, not all Emoji on this list are supported. Because of how third-party vendors implement Emoji, many Emoji pose a problem as they appear the same visually as other characters, Emoji or otherwise. These visually-matching characters are treated as potential homographic attacks, and are **unavailable for registration in any combination**. Also, some Emoji domains are represented differently than specified in Unicode's definitions in certain situations outlined in the following sections.

Emoji Translation

Akin to how IDN domain names are converted into Punycode, .WS Emoji Domains follow the same encoding rules as set forth in RFC3492. Vendors that support the usage of Emoji in URLs have a common set of rules that remove any non-printing character from a string prior to translating from Unicode to Punycode. The following code points appear in many Emoji, but **should be stripped from any query before translating into Punycode and sending via EPP**.

U+FE0F: "VARIATION SELECTOR-16"

<http://unicode.org/cldr/utility/character.jsp?a=FE0F>

The Variation Selector is often used as the final codepoint in a complex multi-codepoint Emoji, signifying the end of a sequence in most cases. Such Variation Selectors are not permitted in domain name labels, and must be stripped from a label prior to the Punycode encoding process.

U+200D: "ZERO WIDTH JOINER"

<http://unicode.org/cldr/utility/character.jsp?a=200D>

The Zero Width Joiner (ZWJ) is used when the surrounding codepoints should be joined together as one visual glyph when presented to the user. Such ZWJ are not permitted in domain name labels, and must be stripped from a label prior to the Punycode encoding process.

Disallowed Emoji

The Emoji listed in our mapping tables may have STATUS "disallowed" or "mapped" to another codepoint that is disallowed. Please consult this list for up-to-date information about how the .WS Registry handles Emoji IDN. Reasons for disallowing certain Emoji is as follows.

- Production (currently Unicode 11.0.0): <https://www.website.ws/WSEmojiMappingTable.txt>
- OT&E (currently Unicode 11.0.0): <https://www.website.ws/WSEmojiMappingTableOTE.txt>

Homography

When determining homographic collisions within Emoji, we take into consideration the multiple popular platforms in which Emoji are rendered. Such platforms include Apple (iOS), Google (Android), Twitter, Facebook, and others. The .WS Registry evaluates the rendering of each Emoji on each platform. If a visual match is found, the Emoji is disallowed from registration.

For example, the representation of the following Emoji are visually the same on iOS, Google, and other platforms:

PERSON GETTING HAIRCUT (U+1F487)

WOMAN GETTING HAIRCUT (U+1F487 U+200D U+2640 U+FE0F)

No	Code	Browser	Appl	Goog	Twtr.	One	FB	FBM	Sams.	Wind.	GMail	SB	DCM	KDDI	CLDR Short Name
816	U+1F487												—		person getting haircut
828	U+1F487 U+200D U+2640 U+FE0F				—		—		—	—	—	—	—	—	woman getting haircut

As a result, both Emoji are currently disallowed.

Blacklisted Emoji

Some Emoji are unsupported due to their code point representation using ASCII characters unsupported by IDNA2008. Such an example is U+0023 U+FE0F U+20E3 KEYCAP: # .

Glyph Facing Direction



This latest proposed version of TR51 introduces the ability to change the facing direction of an Emoji, limited to left or right. The .WS Registry has deemed such sequences as **invalid** for the following reasons:

1. As the Emoji Specification does not define the "default" direction a specific Emoji must display, it is up to the implementing vendor to decide how they wish to visually represent the codepoint. As a result, the potential for visually confusing characters is high.
2. The protocol for introducing new Emoji that may be sequenced with a facing direction is the same as submitting a new Emoji proposal. Because there are no currently-defined RGI sequences for supported Emoji Glyph Facing Direction, the .WS Registry will prevent such combinations from registration.
3. Emoji are compatible for use with bi-directional (Bidi), but how vendors choose to interpret Emoji directional modifiers within Bidi input is not known. To allow for better interoperability with future considerations of .WS Emoji domains (such as allowing for the combination of RTL languages [Hebrew, etc.] with Emoji), such Glyph Facing Direction combinations are disallowed.

Rules for Detection

The following rules are implemented at the .WS Registry to check for domain name validation, particularly relating to the usage of Emoji Glyph Facing Direction:

The following codepoints are defined by TR-51 as the directionally-modifying characters for use in Emoji sequences:

Codepoint	Text Representation	Example Emoji	CLDR Short Name
U+27A1	→		right arrow
U+2B05	←		left arrow

Any domain names that contain the above codepoints **directly after** any other Emoji character will be returned as "Unavailable".

Should a document be published for Unicode defining which Emoji may be combined with the above codepoints, the restriction of domain registrations containing these Emoji may be reduced.

Combining Characters

Emoji may be combined only with other Emoji or standard ASCII characters. Please note: this diverges from typical IDNA processing, where any Latin characters may be combined with other Latin characters. .WS IDN does not consider all "Latin" to be the same. The following examples explain differences between IDNA2008 Latin Character specification versus .WS IDN:

Codepoint Range	Example Character	Character Description	Unicode Classification	.WS IDN Classification
0030..0039 002D 0041..005A 0061..007A	0 - A a	DIGIT ZERO HYPHEN-MINUS LATIN CAPITAL LETTER A LATIN SMALL LETTER A	C0 Controls and Basic Latin	Basic Latin
00C0 00E1	á	LATIN SMALL LETTER A WITH ACUTE	C1 Controls and Latin-1 Supplement	Latin Extended
0100..017F	Ā	LATIN CAPITAL LETTER A WITH MACRON	Latin Extended-A	Latin Extended
0180..024F	ḃ	LATIN SMALL LETTER B WITH STROKE	Latin Extended-B	Latin Extended
2C60..2C7F	Ł	LATIN CAPITAL LETTER L WITH DOUBLE BAR	Latin Extended-C	Latin Extended
A720..17FF	◌́	MODIFIER LETTER STRESS AND HIGH TONE	Latin Extended-D	Latin Extended
AB30..AB6F	Ɑ	LATIN SMALL LETTER BARRED ALPHA	Latin Extended-E	Latin Extended
1E00..1EFF	Å	LATIN CAPITAL LETTER A WITH RING BELOW	Latin Extended	Latin Extended

Combining Emoji with other IDN characters will return as invalid. The following domain name is invalid as a result:

Unicode	Image Substitution	Punycode	Reason
---------	--------------------	----------	--------

café☕.ws	café☕.ws	xn--caf-dma4968b.ws	Script Mixing. "é" is Latin Extended, and may not be mixed with Emoji.
----------	----------	---------------------	---

Bear in mind, that the .WS Registry does not discern language usage when mixing Emoji with ASCII characters, just as the Registry does not inquire about the source language of an IDN. The only distinction made is that the script(s) used are compatible. For example, **café☕.ws** (Spanish for "coffee") is not allowed because of the "é" character. However, **kaffee☕.ws** (German for "coffee") is valid, since it does not use any *Basic Latin* characters besides the single Emoji.



Example Domain Reference

Valid Emoji Domains

☺.ws => xn--h28h.ws (Valid: Contains an approved Emoji)

I♥NY.ws => xn--iny-4r6a.ws (Valid: Contains an approved combination of Emoji and ASCII)

Invalid Emoji Domains

Unicode	Image Substitution	Punycode	Reason
☹.ws	 .ws	xn--l48h.ws	Homography. Matches with ☹♂.ws [xn--g5hx380o.ws].
★.ws	<n/a>	xn--p3h.ws	Not an Emoji. U+2605 is not defined as an Emoji by Unicode.
#☐ws	 .ws	xn--#-2sn.ws	"#" is not a valid character per RFC5890

EPP Usage

Manipulating Emoji domains uses the same processes as any IDN with regards to our EPP system.

Standard EPP Considerations

As declared in the **dotWS EPP Implementation Guide**, when registering an IDN over EPP, no language tag is necessary in the client request. Likewise with Emoji IDN, no language tag is required in any request.

EPP Fee Extension Considerations

All .WS Emoji domains that are defined in the current version of Unicode (excluding Disallowed Emoji) are available for registration on our base pricing tier. This means that if your registrar has not adopted the .WS EPP Fee Extension, you may still register .WS Emoji Domains from the current version of Unicode.

However, as defined in section "Emoji Tiers," the .WS Registry may make available upcoming Emoji that are finalized to be included in the next version of Unicode. Such Emoji are considered "prospective" and

are assessed a higher price as determined by the registry. Pricing can be found by utilizing the EPP Fee Extension to perform a CHECK on a domain that includes a confirmed, but not-yet-released Emoji.

Changes to EPP Responses

When searching for an Emoji IDN that does not conform to our specification (e.g. includes Punycode-encoded characters such as U+FE0F or U+200D), the EPP server will respond with a standard `<extData>` element signifying that the requested Emoji IDN is invalid. Within the `<extData>` element are child elements providing the corrected Punycode for the requested invalid domain(s).

Client Validation

Though this document explains how to detect and block Invalid Emoji, we do not require registrars to perform any validation on client input. For example, if an Invalid Emoji is sent to the EPP server via a CHECK command, our system will respond with an invalid domain response.

EPP Examples

Valid Emoji CHECK Response

In this example, a valid Emoji IDN is requested in a CHECK, and the response is generated by the server. The response is not any different than requesting a non-IDN domain, but is included in this document for completeness.

```
C: <?xml version="1.0" encoding="UTF-8"?>
C: <epp xmlns="urn:ietf:params:xml:ns:epp-1.0"
xmlns:domain="urn:ietf:params:xml:ns:domain-1.0"
C:   <command>
C:     <check>
C:       <domain:check>
C:         <domain:name>xn--ohmy-fe63ccc873c.ws</domain:name>
C:       </domain:check>
C:     </check>
C:     <clTRID>58ddb3ace1efe</clTRID>
C:   </command>
C: </epp>

S: <?xml version="1.0" encoding="UTF-8" standalone="no"?>
S: <epp xmlns="urn:ietf:params:xml:ns:epp-1.0">
S:   <response>
S:     <result code="1000">
S:       <msg>Command completed successfully</msg>
S:     </result>
S:     <resData>
S:       <domain:chkData xmlns:domain="urn:ietf:params:xml:ns:domain-1.0">
S:         <domain:cd>
S:           <domain:name avail="1">xn--ohmy-fe63ccc873c.ws</domain:name>
S:         </domain:cd>
S:       </domain:chkData>
S:     </resData>
S:     <trID>
S:       <clTRID>58ddb3ace1efe</clTRID>
S:       <svTRID>24494-1490924461</svTRID>
S:     </trID>
S:   </response>
S: </epp>
```

Invalid Emoji CHECK Response

In this example, the client attempts to CHECK domain xn--iq9h.ws, which is invalid as it contains an Emoji that is not implemented in the current version of Emoji as defined by Unicode. A `<domain:reason>`

element is used to explain that the provided domain name is invalid, via the error message "Parameter value syntax error."

```
C: <?xml version="1.0" encoding="UTF-8"?>
C: <epp xmlns="urn:ietf:params:xml:ns:epp-1.0"
xmlns:domain="urn:ietf:params:xml:ns:domain-1.0">
C:   <command>
C:     <check>
C:       <domain:check>
C:         <domain:name>xn--iq9h.ws</domain:name>
C:       </domain:check>
C:     </check>
C:     <clTRID>58ddb1a74688b</clTRID>
C:   </command>
C: </epp>

S: <?xml version="1.0" encoding="UTF-8" standalone="no"?>
S: <epp xmlns="urn:ietf:params:xml:ns:epp-1.0">
S:   <response>
S:     <result code="1000">
S:       <msg>Command completed successfully</msg>
S:     </result>
S:     <resData>
S:       <domain:chkData xmlns:domain="urn:ietf:params:xml:ns:domain-1.0">
S:         <domain:cd>
S:           <domain:name avail="0">xn--iq9h.ws</domain:name>
S:           <domain:reason>Parameter value syntax error</domain:reason>
S:         </domain:cd>
S:       </domain:chkData>
S:     </resData>
S:     <trID>
S:       <clTRID>58ddb1a74688b</clTRID>
S:       <svTRID>24494-1490923943</svTRID>
S:     </trID>
S:   </response>
S: </epp>
```

Server Validation and Hints

In the event that the server receives an invalid Emoji domain request, but the request can be resolved by proper translation of the requested label, information returned via XML will be extended to include the suggested domain. Such a suggestion would be returned in cases where the requested label contains encoded Zero Width Joiners or Variation Selectors.

In the following example, the Emoji domain "Man Running" is represented as the Punycode-encoded string "xn--1ug66vku9rrn4h.ws". This encoded domain contains two invalid characters: U+200D Zero Width Joiner and U+FE0F Variation Selector-16. If the WS EPP Server receives such a domain with disallowed codepoints, the returned EPP response will be modeled as follows:

```
C: <?xml version="1.0" encoding="UTF-8" standalone="no"?>
C: <epp xmlns="urn:ietf:params:xml:ns:epp-1.0">
C:   <command>
C:     <info>
C:       <obj:info xmlns:obj="urn:ietf:params:xml:ns:obj">
C:         <obj:name>xn--1ug66vku9rrn4h.ws</obj:name>
C:       </obj:info>
C:     </info>
C:     <clTRID>ABC-12345</clTRID>
C:   </command>
C: </epp>

S: <?xml version="1.0" encoding="UTF-8" standalone="no"?>
S: <epp xmlns="urn:ietf:params:xml:ns:epp-1.0">
S:   <response>
```

```

S: <result code="1000">
S: <msg>Command completed successfully</msg>
S: <extValue>
S: <value>
S: <domain:name xmlns:domain="urn:ietf:params:xml:ns:domain-1.0">xn--
1ug66vku9rrn4h.ws</domain:name>
S: </value>
S: <reason>Not valid punycode</reason>
S: </extValue>
S: </result>
S: <resData>
S: <domain:chkData xmlns:domain="urn:ietf:params:xml:ns:domain-1.0">
S: <domain:cd>
S: <domain:name avail="0">xn--1ug66vku9rrn4h.ws</domain:name>
S: <domain:reason>Parameter value syntax error</domain:reason>
S: </domain:cd>
S: </domain:chkData>
S: </resData>
S: <trID>
S: <c1TRID>ABC-12345</c1TRID>
S: <svTRID>54321-XYZ</svTRID>
S: </trID>
S: </response>
S:</epp>

```

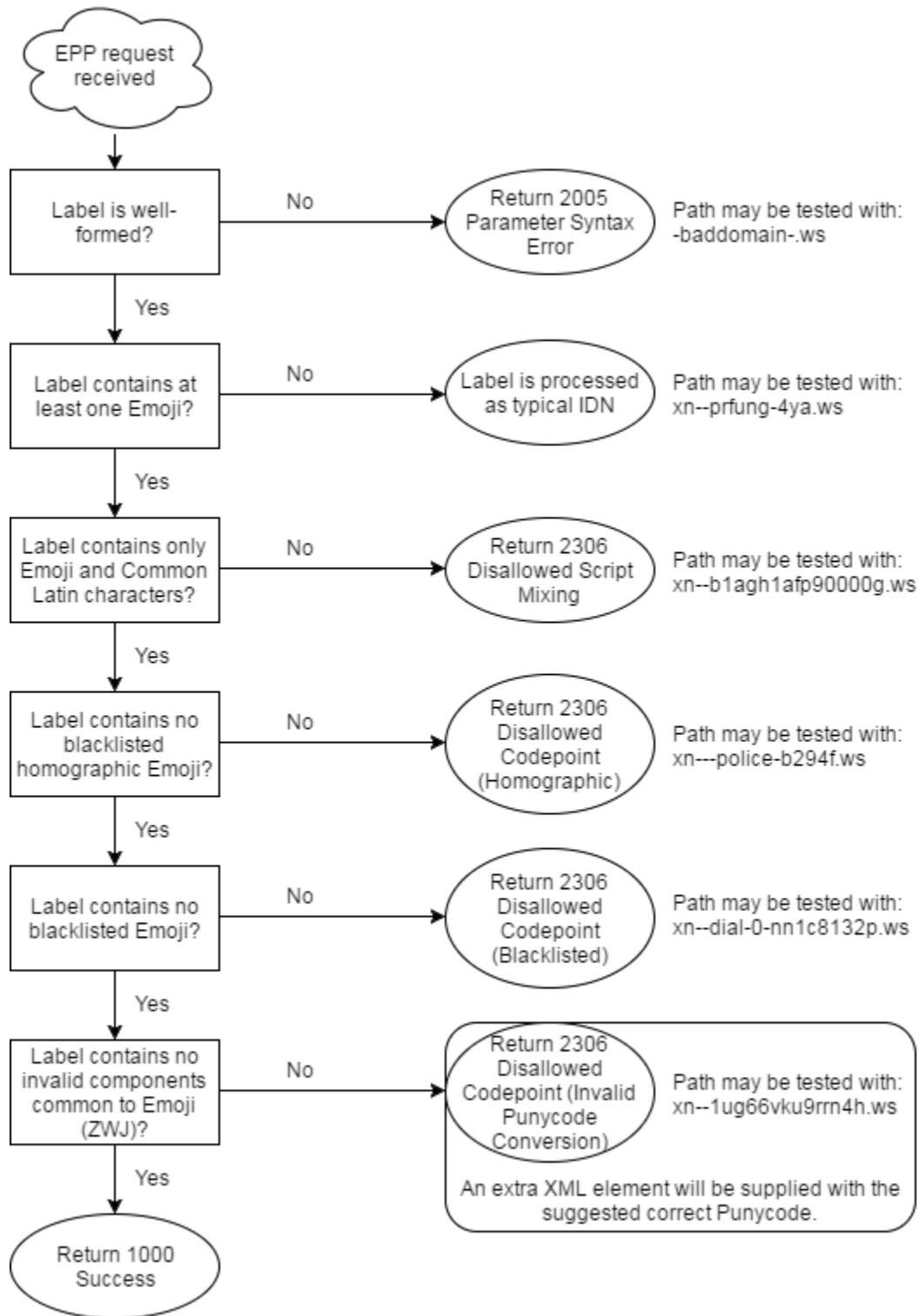
It is at the discretion of each registrar if they would like to implement any sort of mechanism to detect such responses and present the "corrected" label to end users.

Appendix

.WS Emoji IDN Chart

The current list of Emoji supported by .WS is available at <https://www.website.ws/WSEmojiMappingTable.txt> for production usage, and <https://www.website.ws/WSEmojiMappingTableOTE.txt> for OT&E. This document contains all current Emoji codepoints, and information on how each codepoint is handled. The format of this document is very similar to <http://www.unicode.org/Public/idna/9.0.0/IdnaMappingTable.txt>.

EPP Processing Flowchart



References

<http://www.rfc-editor.org/rfc/rfc3490.txt> - Internationalizing Domain Names in Applications (IDNA)

<http://www.rfc-editor.org/rfc/rfc3454.txt> - Stringprep

<http://www.rfc-editor.org/rfc/rfc3491.txt> - Nameprep

<http://www.rfc-editor.org/rfc/rfc3492.txt> - Punycode

<http://www.unicode.org/Public/emoji/latest/> - Latest Emoji Data (machine readable)