

# EAI Technical Education and Awareness to the Developers' Community Websites

–

## Proposed FAQs

Submitted by



ThinkTrans

We convey  
your words...

ThinkTrans LLP

India

email: [akshat@thinktrans.in](mailto:akshat@thinktrans.in)

**Date of commencement**

10<sup>th</sup> June 2021

**Date of submission**

28<sup>th</sup> June 2022

## Contents

<b>Introduction:</b> .....	3
<b>Proposed FAQs:</b> .....	3
<b>1. Character/Length specific</b> .....	3
<b>2. General Validation specific</b> .....	3
<b>3. Programming Language Specific</b> .....	3
<b>4. General IDN/EAI protocol specific questions</b> .....	3
<b>FAQs with Answers:</b> .....	4
<b>1. What is the simplest regular expression to validate emails to not accept them blindly?</b> .....	4
<b>2. What is the maximum length of a valid email address?</b> .....	4
<b>3. What is the maximum length of a valid domain name?</b> .....	4
<b>4. Are the IDNs case-sensitive?</b> .....	4
<b>5. Are the Internationalized Email Addresses case-sensitive?</b> .....	5
<b>6. How to detect language/script of a punycode?</b> .....	5
<b>7. Is there a way to avoid showing “xn--” for IDNs?</b> .....	5
<b>8. How to compare punycodes for visible similarity?</b> .....	5
<b>9. Is the whole URL required to be converted to punycode or only the domain name part?</b> .....	6
<b>10. Can I treat all the domain names (IDNs and ASCII based) to be IDNs and process them as such without any ill-effects in my code flow?</b> .....	6
<b>11. While converting IDN to ASCII (or vice-versa) in any programming language, can the whole URL (with protocol parameters such as HTTPS, Colons etc) be submitted as it is?</b> .....	6
<b>12. Are IDN Conversion Functions reversible?</b> .....	6
<b>13. Are ASCII symbols prohibited altogether from IDNs?</b> .....	6
<b>14. Just like there is a list of valid TLDs, where can I find a list of valid “Email TLDs”?</b> .....	7
<b>15. What is the level of support for IDNs in Android in terms of UA related requirements?</b> .....	7
<b>16. What is the level of support for IDNs in iOS in terms of UA related requirements?</b> .....	7
<b>17. I see a special character in an IDN which I want to type out. How should I type it?</b> 7	

## Introduction:

This document is a sub-document of the Final Report on the execution of the project titled “EAI Technical Education and Awareness to the Developers’ Community Websites”. This sub-document gives detailed summary of one of the major deliverables of the project i.e. “Augmenting the Frequently Asked Questions”. The deliverable required that the existing UASG FAQs list to be vetted, and augmented with additional questions which were deemed necessary as per the requirements of the UA cause.

## Proposed FAQs:

The FAQs are categorized in four broad categories viz.

- 1. Character/Length specific**
- 2. General Validation specific**
- 3. Programming Language Specific**
- 4. General IDN/EAI protocol specific questions**

Following are the actual questions as per their respective categories.

1. Character/Length specific
  - a. [What is the maximum length of a valid email address?](#)
  - b. [What is the maximum length of a valid domain name?](#)
  - c. [How to detect language/script of a punycode?](#)
  - d. [Can I treat all the domain names \(IDNs and ASCII based\) to be IDNs and process them as such without any ill-effects in my code flow?](#)
  - e. [Are ASCII symbols prohibited altogether from IDNs?](#)
  - f. [I see a special character in an IDN which I want to type out. How should I type it?](#)
2. General Validation specific
  - a. [What is the maximum length of a valid email address?](#)
  - b. [What is the maximum length of a valid domain name?](#)
  - c. [What is the simplest regular expression to validate emails to not accept them blindly?](#)
  - d. [Are ASCII symbols prohibited altogether from IDNs?](#)
3. Programming Language specific
  - a. [What is the level of support for IDNs in Android in terms of UA related requirements?](#)
  - b. [What is the level of support for IDNs in iOS in terms of UA related requirements?](#)
4. General IDN/EAI protocol specific questions
  - a. [Are the IDNs case-sensitive?](#)
  - b. [Are the Internationalized Email Addresses case-sensitive?](#)
  - c. [How to detect language/script of a punycode?](#)
  - d. [Is there a way to avoid showing “xn--” for IDNs?](#)
  - e. [How to compare punycodes for visible similarity?](#)
  - f. [Is the whole URL required to be converted to punycode or only the domain name part?](#)
  - g. [Can I treat all the domain names \(IDNs and ASCII based\) to be IDNs and process them as such without any ill-effects in my code flow?](#)

- h. [While converting IDN to ASCII \(or vice-versa\) in any programming language, can the whole URL \(with protocol parameters such as HTTPS, Colons etc\) be submitted as it is?](#)
- i. [Are IDN Conversion Functions reversible?](#)
- j. [Just like there is a list of valid TLDs, where can I find a list of valid “Email TLDs”?](#)

## FAQs with Answers:

Proposed answer to the questions identified have been provided, wherever applicable, below.

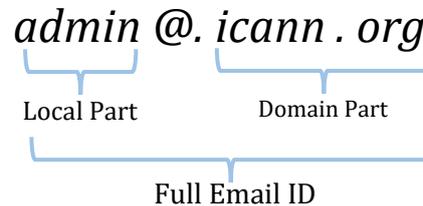
### 1. What is the simplest regular expression to validate emails to not accept them blindly?

#### Proposed Answer:

When it comes to validating an Email, there is no simple solution through the regular expression-based validation. Given the way the E-mail protocol is framed, technically speaking, the “regex” that will validate it fully will be huge and hence not feasible from the processing efficiency point of view. It is thus recommended to perform the validation by sending a mail to the Email ID to be validated with some personalized link or some code which the user has to click/input to verify himself/herself to be the rightful owner of the submitted E-mail ID.

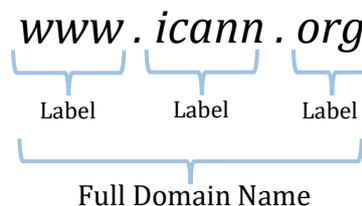
### 2. What is the maximum length of a valid email address?

*Typically, the maximum length of an email Local part is 64 while the maximum permissible length of the full email ID is 254. The length of the email local part is intrinsic to this limit.*



### 3. What is the maximum length of a valid domain name?

*As spelled out in [RFC 1035](#), the maximum length of a full domain name is 255 octets including the label separator i.e. “.”. Each label can have a maximum length of 63.*



### 4. Are the IDNs case-sensitive?

#### Proposed Answer:

The IDNs as a part of their processing to a punycode, undergo a [nameprep](#) operation which includes “case-folding”, i.e. converting all the characters to the lowercase. So, intrinsically, IDNs are case-insensitive.

## 5. Are the Internationalized Email Addresses case-sensitive?

### Proposed Answer:

An Email ID, comprises of two parts governed by two different sets of protocols. Following is the composition of an E-mail ID

`<email local part>@<domain name part>`

The `<email local part>` is governed by the mail administrator of the given email service and the email protocol does not explicitly bind the implementations to keep that in the case-insensitive form. Hence, though unwise, one can and expect `<email local part>` to be case-sensitive. For more elaborate discussion and common guidelines released by the Universal Acceptance Steering Group, refer to this link to [UASG 028](#).

The `<domain name part>` as a part of conversion of IDN to a punycode, undergo a [nameprep](#) operation which includes “case-folding”, i.e. converting all the characters to the lowercase. So, intrinsically, IDNs are case-insensitive.

## 6. How to detect language/script of a punycode?

### Proposed Answer:

Punycode is a language/script agnostic string and should be considered as such. However, if one wants to understand the language/script of the domain name that punycode represents, then conversion of the said punycode to its Unicode equivalent is a must. Once converted to Unicode, depending on the business requirement, appropriate deductions can be made as to the script/language of the given IDN label. For those interested in that process, refer to this [link](#) on the Unicode FAQs page.

## 7. Is there a way to avoid showing “xn--” for IDNs?

### Proposed Answer:

The “xn--” representation of an IDN pertains to indicate that the domain name is in its punycode format. To get rid of the same, depending on the programming language you are coding into, use the appropriate library to convert from “punycode to Unicode”. For the recent studies on the subject, follow the document [[UASG-018A](#)] for some of the common programming languages like: C, C#, Go, Java, Javascript, Python and Rust. In addition, for additional platforms like, iOS Swift, PHP and Android Kotlin, do follow the document [[UASG 037](#)].

## 8. How to compare punycodes for visible similarity?

### Proposed Answer:

Punycodes being a result of a mathematical algorithm, do not have direct bearing on the visual shape of the Unicode string they represent. For visual similarity comparison, one needs to convert the punycodes to their equivalent Unicode strings and rely either on the preferred string similarity assessment algorithms or resort to manual inspection.

Depending on the programming language you are coding into, use the appropriate library to convert from “punycode to Unicode”. For the recent studies on the subject, follow the document [[UASG-018A](#)] for some of the common programming languages like: C, C#, Go, Java, Javascript, Python and Rust. In addition, for additional platforms like, iOS Swift, PHP and Android Kotlin, do follow the document [[UASG 037](#)].

**9. Is the whole URL required to be converted to punycode or only the domain name part?**

**Proposed Answer:**

Technically speaking, only the domain part,

`https://www.थिंकट्रान्स.भारत/index.php`

i.e. only the “थिंकट्रान्स.भारत” part in the above URL needs to be subjected to the punycode conversion routine. Rest of the parts i.e. protocol and file-system elements should not be subjected to punycode conversion.

**10. Can I treat all the domain names (IDNs and ASCII based) to be IDNs and process them as such without any ill-effects in my code flow?**

**Proposed Answer:**

Given the way punycode to Unicode conversion algorithm is framed, subjecting non-IDN labels to it do not necessarily undergo any undesirable change. Hence it would be safe to say that one can process them without any ill-effects.

**11. While converting IDN to ASCII (or vice-versa) in any programming language, can the whole URL (with protocol parameters such as HTTPS, Colons etc) be submitted as it is?**

**Proposed Answer:**

Technically speaking, only the domain part,

`https://www.थिंकट्रान्स.भारत/index.php`

i.e. only the “थिंकट्रान्स.भारत” part in the above URL needs to be subjected to the punycode conversion routine. Rest of the parts i.e. protocol and file-system elements should not be subjected to punycode conversion.

**12. Are IDN Conversion Functions reversible?**

**Proposed Answer:**

The RFC 3491 and subsequently RFC 5891, recommend a set of operations on the IDN label which are not per-se fully reversible, case in point being Normalization and the case-folding routines. There could be labels which do not change even after undergoing these routines as they might already be in normalized and in lower case forms. However, as a whole, one can say that these functions may not always yield fully reversible strings.

**13. Are ASCII symbols prohibited altogether from IDNs?**

**Proposed Answer:**

The Internationalized Domain Names are defined as those domain names which by definition have “at least one character from the non-ASCII space”. This is to say that it has at least one character which is not in the Unicode character range of 0-128. So as long as “at least one character” clause is satisfied, rest of the characters, technically, can be from the ASCII space. Hence, ASCII symbols can be part of the IDNs.

#### **14. Just like there is a list of valid TLDs, where can I find a list of valid “Email TLDs”?**

##### **Proposed Answer:**

Typically speaking, such a list would be a list of TLDs which have an “mx” entry as a part of their DNS record. However, neither presence of an mx record ensures a full implementation of an e-mail server, nor absence of one indicates otherwise given the way DNS redirection mechanisms could work in tandem to redirect user queries. Having said this, Mail-servers often are subjected to attacks through various spamming agents for the bombardment of advertisements/offers/phishing-spoofing attacks. Hence, officially maintaining such a list would only amount to further abuse of the process and hence not advisable.

#### **15. What is the level of support for IDNs in Android in terms of UA related requirements?**

*There was a systematic study constituted by the UASG studying around 22 libraries across four of the major operating systems vis-à-vis support titled “UA-Readiness Evaluation of Programming Languages and Development Frameworks”. Eight critical libraries from Android-Kotlin were evaluated namely okHttp, HttpURLConnection, Retrofit, Fuel, Volley, HttpURLConnection, Apache HttpClient, Jakarta Mail, Email Intent. The support levels at the time of the study were found to be poor. The detailed study can be found in [presentation](#) form as well as the [document](#) form at the respective links.*

#### **16. What is the level of support for IDNs in iOS in terms of UA related requirements?**

*There was a systematic study constituted by the UASG studying around 22 libraries across four of the major operating systems vis-à-vis support titled “UA-Readiness Evaluation of Programming Languages and Development Frameworks”. Three critical libraries from iOS-Swift were evaluated namely MessageUI, Alamofire and URLSession. The support levels at the time of the study were found to be poor. The detailed study can be found in [presentation](#) form as well as the [document](#) form at the respective links.*

#### **17. I see a special character in an IDN which I want to type out. How should I type it?**

*Typically, all the operating systems have pre-defined set of language/script keyboards which enable users to input content in the respective languages. If the character you are looking for forms part of everyday usage of the language/script it belongs to, you can input the same using the respective language/script keyboard.*

*You can find the procedures of adding the language/script keyboard in some of the popular Operating Systems at links given. For Microsoft Windows, use [this](#). For macOS, use [this](#). For Android, use [this](#). For iOS use [this](#) and for Ubuntu Linux use [this](#).*

*In case you want to input the specific character without using the language/script inputting tools, there are some softwares that can be used which provide neat interface for searching and inputting the character. One such application is [BabelPad](#), for Windows platform. For those who are looking for platform independent solution, can make use of the [BabelMap](#) which is an online application and gives access to the entire Unicode repertoire.*

**18. Where can I find the compliance levels of various programming language libraries vis-à-vis UA Compliance?**

You can find the same at:

[UASG 018 Reviewing Programming Languages and Frameworks for Compliance with Universal Acceptance Good Practice EN](#)

[UASG 031 FAQs: UA Readiness of Programming Languages and Email Tools EN](#)

[UASG 037 UA-Readiness of Some Programming Language Libraries and Frameworks EN](#)