

Domain Name Basics

Internationalized Domain Names

Tobias Sattler
tobiassattler.com

What are Internationalized Domains?

An Internationalized Domain Name (**IDN**) is an Internet domain name **containing** at least one label displayed in a **language-specific script** or **alphabet**, such as Arabic, Chinese, Cyrillic, etc.

That enables people around the world to use domain names in local languages and scripts, such as 例.com, пример.рф, or nic.vermögensberater.



Why are they important?

The number of Internet users is **growing** steadily, and for many, it isn't easy to use ASCII characters.

Far **more people** who do **not** use **Latin** languages and scripts are using the Internet than those who do.

The total number of IDNs is continually increasing and is now over 9 million.



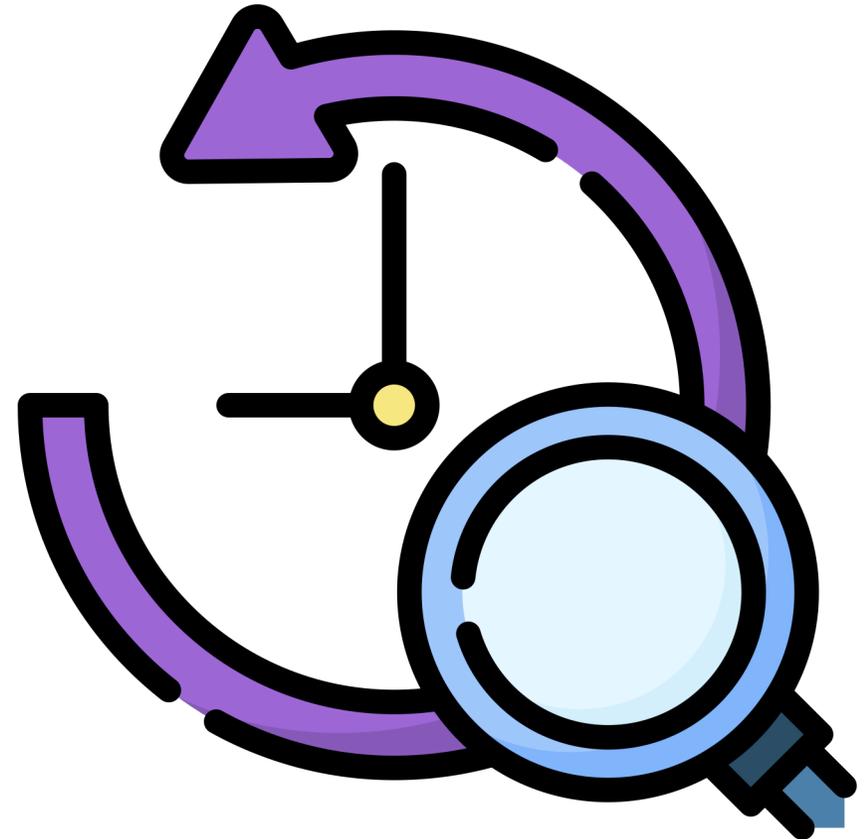
History #1

In **1996** Martin Dürst **proposed** the first usage of IDN.

From 1998 to 1999, Multiple organizations around the world researched on IDN and developed first testbeds.

In **1999** i-DNS.net launched the **first commercial IDN** under the top-level IDN domain .gongsi.

IETF formed the IDN working group in 2000, and ICANN followed a year later.



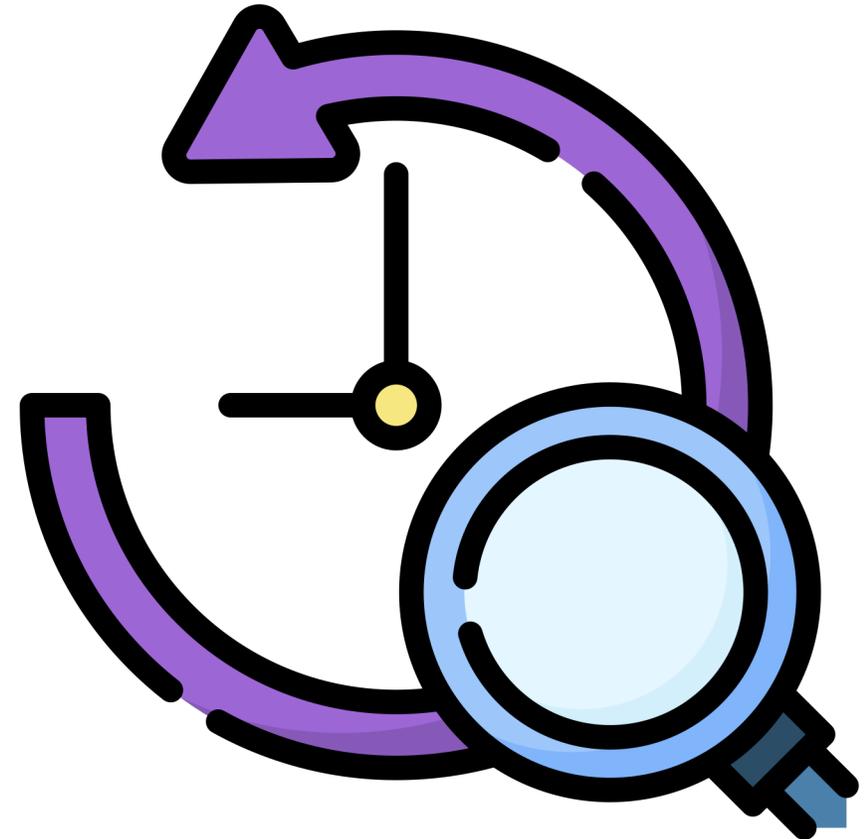
History #2

In 2003 the IDN **RFCs** were **published** (IDNA 2003).

During 2008 and 2010, **IDNA 2008** was developed, and the normalization is no longer part of IDNA, such as β became ss or ζ became σ .

IDNA 2008 disallows about 8,000 characters that used to be valid, including all uppercase characters, full/half-width variants, symbols, and punctuation.

It is noteworthy that many registries don't allow mixing scripts.



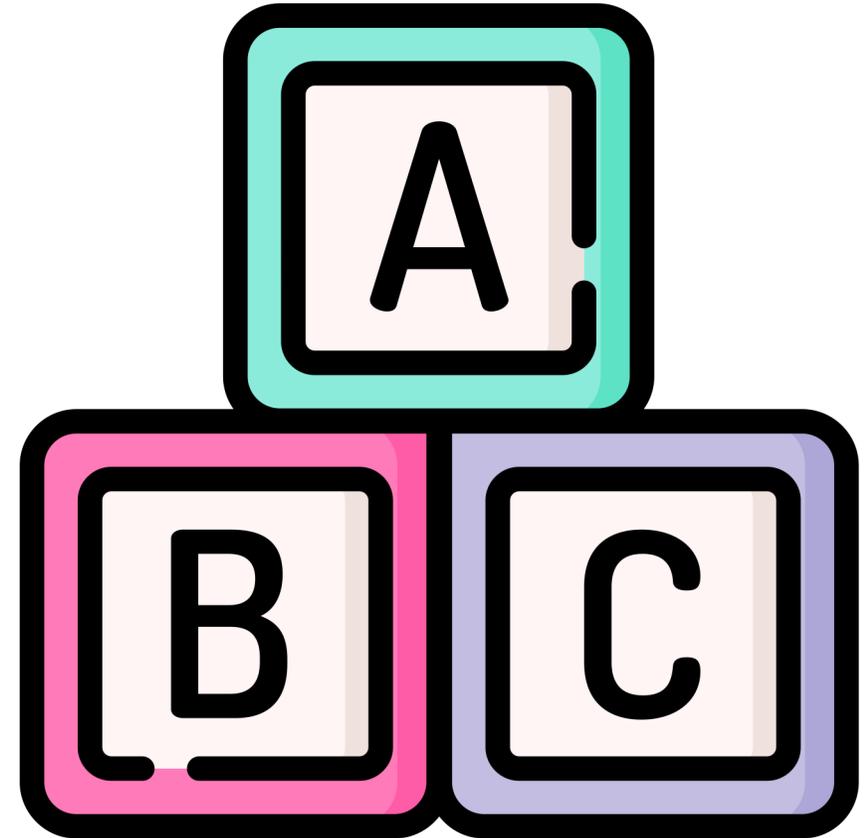
RACE

Row-based ASCII compatible encoding (RACE) is a method for **encoding foreign languages** that use non-English characters.

It was developed in 2000 and considered the prepend character `bq--` to the encoded string.

Some registries, such as Verisign and PIR, had a testbed for RACE domains.

Today, it is mostly **abandoned** in favor of **Punycode**.

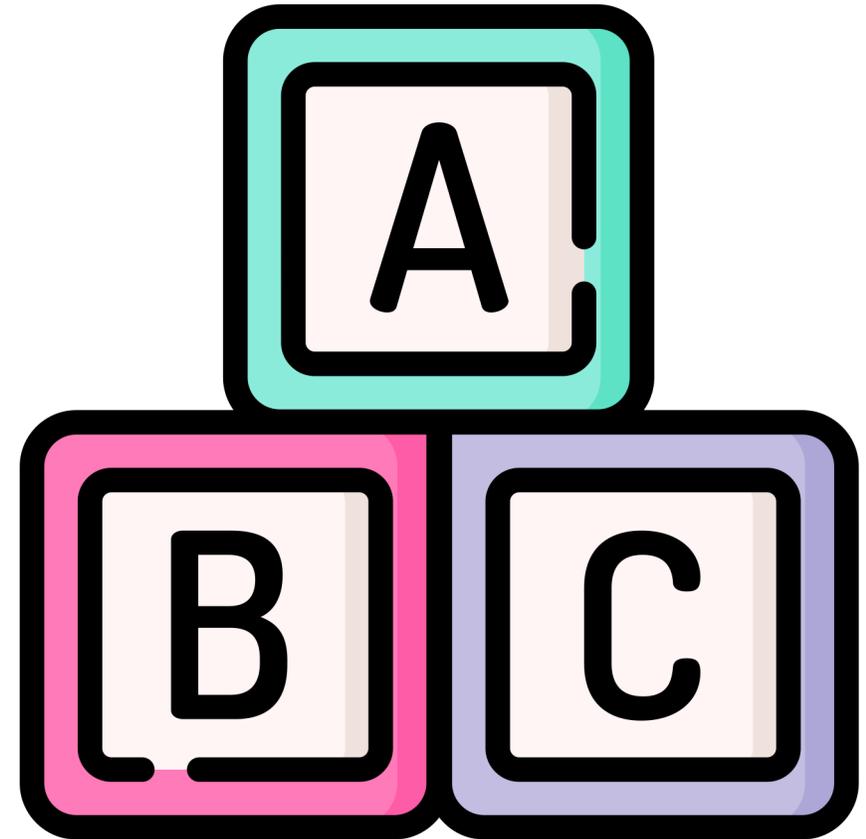


Punycode

Punycode is a **standardized encoding** for converting Unicode into ASCII-compatible character strings, published in RFC 3492 in 2003.

It is designed to uniquely and reversibly represent IDNs consisting of Unicode characters using ASCII characters.

To prevent non-IDN containing hyphens from being accidentally interpreted as Punycode, IDN Punycode sequences have a so-called ASCII Compatible Encoding (**ACE**) prefix **xn--**.



How does Punycode work?

In the following, the essential **characters** are the **letters** a to z and the **numbers** 0 to 9. Together with the hyphen as the separator, these **37** characters are the only valid characters in a Punycode encoded text.

The non-base characters are first sorted by their numerical value. The difference between the values of the individual characters is encoded into a number together with the respective position in the original character string. This number is then represented by the 36 basic characters and appended to the text.



First IDN Top-Level Domains

ICANN **approved** a **working group** within the ccNSO in 2006.

In 2009, the first countries applied for IDN TLDs, such as Egypt, Saudi Arabia, the United Arab Emirates, and Russia.

In **May 2010** the **first** implementations were activated, and TLDs, such as مصر, السعودية, السعودية, and السعودية went **live**.

By now, more than 60 IDN TLDs are approved.

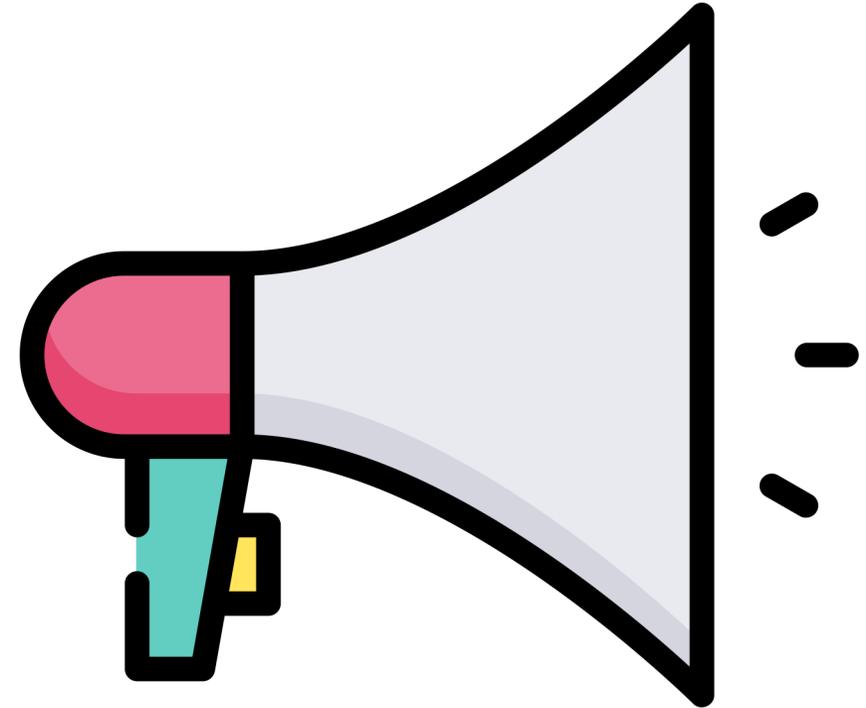


Adoption

Currently, the DNS **root contains** more than **150 IDN TLDs**, with more than 3M domain names under management.

More than **500 TLDs** are **supporting IDNs**, and have more than 7M domain names under management.

The numbers are **steadily growing**; however, considering a total volume of 360M domain names, it will be a long way.



Challenges

Using the Internet in your native language is essential because it makes it much easier for many people.

Therefore, **not** only **domain** names **but** also **email** addresses must be **usable** in any application regardless of the written script.

Unfortunately, this is not yet the case everywhere. Please refer to my presentation on Universal Acceptance to see more on that.



Future

Studies on IDN show that the total number is **steadily increasing**. However, it will take its time, and it is unclear if they will ever surpass ASCII domain names. Nevertheless, it is essential to **support the Universal Acceptance** to prevent unnecessary hurdles.



Thank you!