

Internet guardians want to break web's language barriers

December 10 2021, by Katy Lee



ICANN says the internet could be made much more accessible for people whose languages do not use the Latin alphabet.

When website addresses using writing systems like Chinese and Arabic were introduced back in 2009, it was hailed as a step that would

transform the internet.

But 12 years later, the vast majority of the web remains wedded to the Roman alphabet—and ICANN, the organisation in charge of protecting the internet's infrastructure, is on a mission to change it.

"The truth of the matter is that even if half the world's population uses the internet today, it's the elite of the world—mainly those living in cities, mainly those with a good income," Goran Marby, head of the US-based non-profit, told AFP in an interview.

"Shouldn't we give people the opportunity to use their own scripts, their own keyboards, their own narratives?"

It's thanks to ICANN—the Internet Corporation for Assigned Names and Numbers—that when you type an address at the top of the screen, your computer can find the web page you're looking for.

These days it's theoretically possible to type an address in more than 150 scripts, including obscure ones like ancient Egyptian hieroglyphs, and watch the page load.

But large parts of the internet remain incompatible with writing systems other than the Latin alphabet. Many US websites, for example, would not allow you to make a purchase or subscribe to their newsletter if you entered an email address in Tamil or Hebrew.

When a group of bodies including ICANN tested the world's top 1,000 websites last year, only 11 percent accepted a Chinese or Arabic email address when trying, for example, to contact them via an online form.



Goran Marby is on a mission to defend 'the original idea of the internet—to connect people'

Western-centric since conception

One of ICANN's priorities for the coming years is to ensure that 28 commonly used writing scripts are usable across the internet.

The problem isn't restricted to the West: in China even WeChat, the country's most popular messaging app, does not recognise email addresses written in Chinese characters.

Chinese web addresses often use strings of numbers, like the dating site 5201314.com.

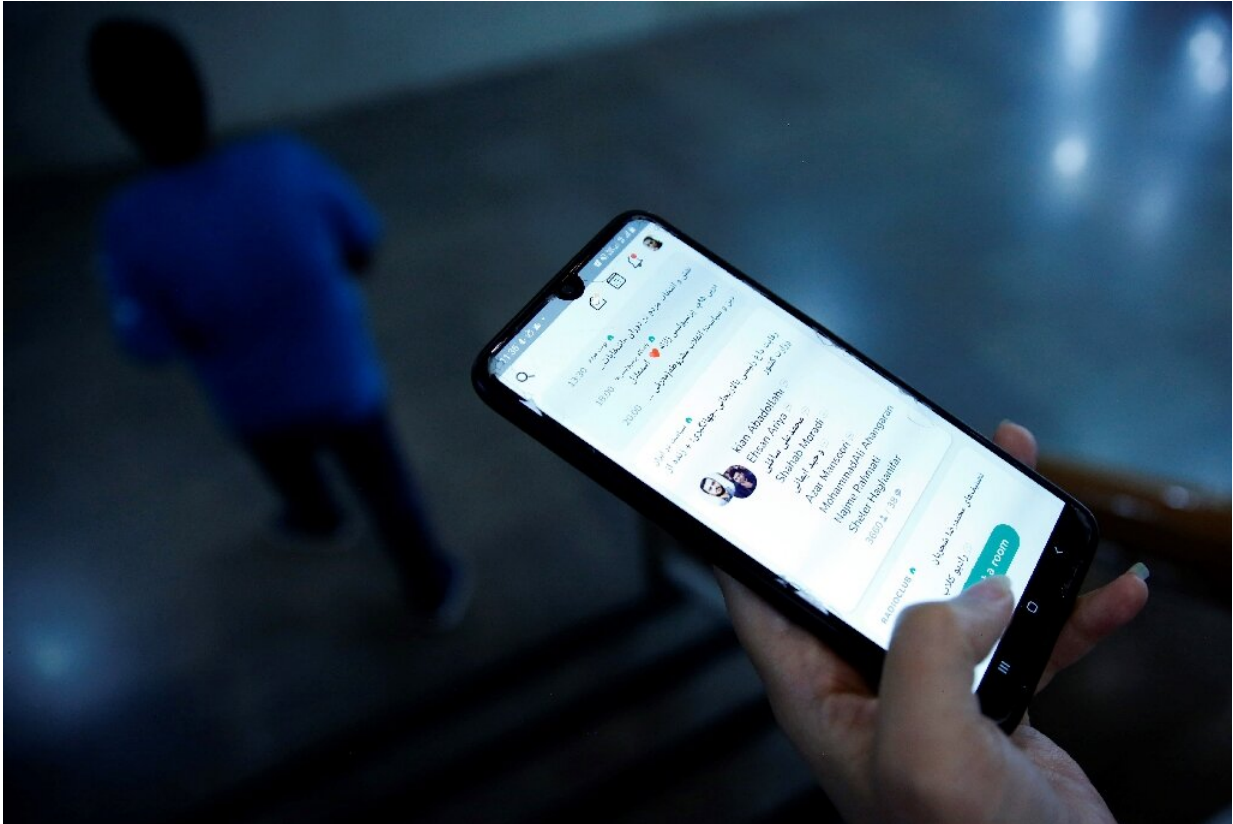
This is partly because it can be hard to remember how to spell a web address in pinyin, the romanised version of Chinese, and partly because number-based puns work well in Mandarin ("520" sounds like "I love you").

In many parts of the world, people have simply tried to adapt to an internet that doesn't speak their language.

"It never even crossed my mind," said Cairo finance worker Hadeer al-Shater, when asked whether she had considered setting up an Arabic-script email account.

"The whole point is to be able to communicate with the rest of the world. And unfortunately, Arabic is not very practical on the internet," she said.

Marby points out that the internet of today grew largely out of the work of US and European computer scientists. As a result, it has disproportionately benefited those who can read and write in Latin-based scripts.



One of ICANN's priorities for the coming years is to ensure that 28 commonly used writing scripts are usable across the internet.

An estimated 37 percent of the world's population—2.9 billion people—have still never used the internet, according to the UN's International Telecommunication Union, 96 percent of them in developing countries.

And Marby argues that this number will remain high if people who don't use the Latin alphabet are locked out.

"We think it's very important to make sure that the original idea of the internet—to connect people—doesn't get forgotten," he said.

Good for business?

Progress has been made on some fronts. Users of Google's popular Gmail service, for example, have been able to exchange messages with people whose email addresses use non-Latin characters since 2014.

In Russia, about 40 percent of companies have a version of their website that uses a Cyrillic "top-level domain" (TLD) – meaning the suffix of a website, such as ".com" or ".org".

But while most Russian hosting services allow email addresses with a Cyrillic domain, the part before the "at" symbol remains in the Latin alphabet, the Russian Coordination Center for TLDs explains on its website.

Marby stresses that "universal acceptance"—the idea of all scripts being usable across the [internet](#)—will never happen unless companies help fix the issue on their end.

"We have to continue to work with software developers and manufacturers to make sure they actually do this," he said.

In the long-run, he argues, universal acceptance will be good for business by allowing companies to reach new markets.

"But this is not something that we're doing over the next six months," he said. "This is going to take years."

© 2021 AFP

Citation: Internet guardians want to break web's language barriers (2021, December 10) retrieved 6 May 2024 from <https://techxplore.com/news/2021-12-internet-guardians-web-language-barriers.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.