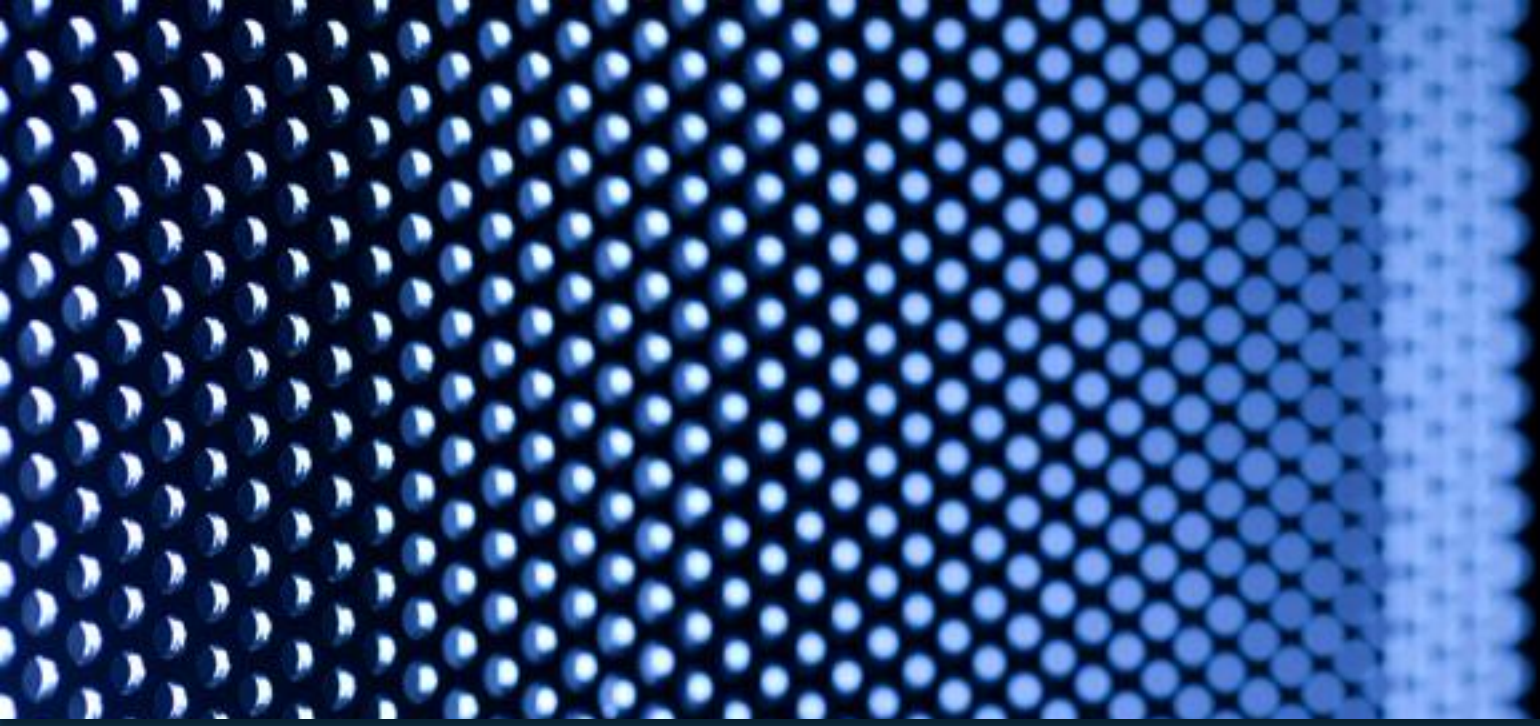


VERISIGN

Towards an end-to-end multilingual web

8th Multilingual Web Workshop, Riga 2015

Dennis Tan

Sr. Product Manager

IDN 101

A-Z 0-9 ‘-’

Internationalized Domain Names

Examples: **nestlé.com** **中文网.中国** **日本.jp**

IDNA2003

IDNA2008

xn--nestl-fsa.com
(ASCII compatible encoding)

DNS Server

Web Server

About us
Products
Customers
Contact us

Towards an end-to-end multilingual web

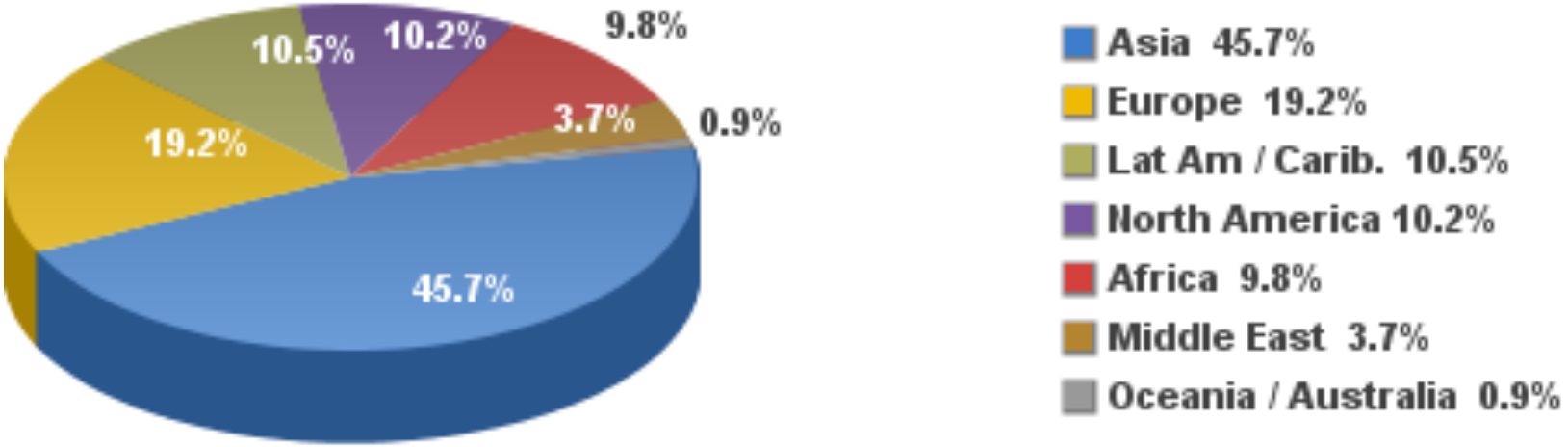
Status of multilingual online content

Support for linguistic diversity by popular web services

Name of service	Number of users	Languages supported[13]	Notes
Twitter	255 million active monthly	35+[14]	Network of 350 000 translators work through Twitter translation centre[15]
Google Translate		80	Statistical machine translation – based on patterns in large amounts of text, users are encourage users to contribute improved translations[16]
Facebook	1.3 billion active monthly	73[17]	Facebook also relies on a network of users who contribute translations[18]
Wikipedia	21 million[19]	287	Number reflects languages for which official Wikipedias have been created[20]. 9 languages have over 1 million Wikipedia articles.

Source: Word report on Internationalised Domain Names 2014, Unesco, Verisign, Eurid

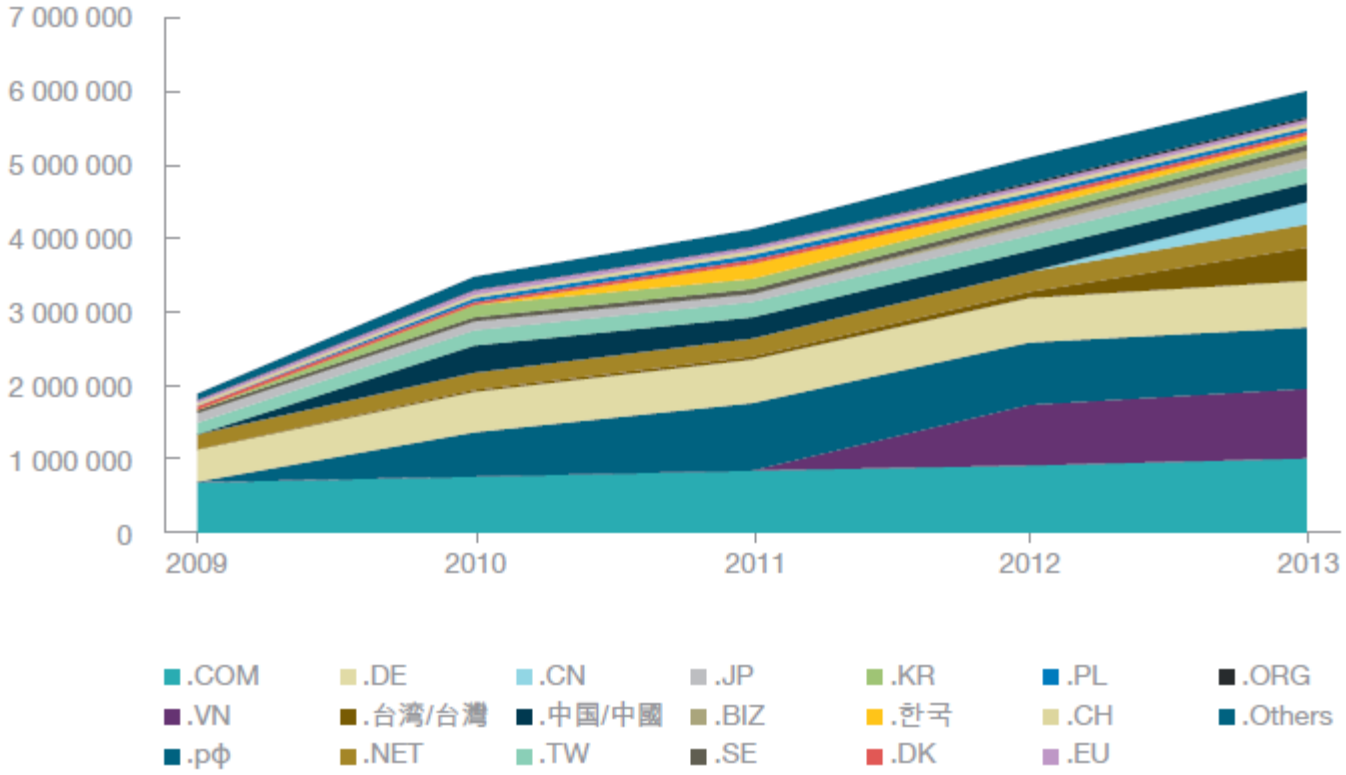
Internet Users in the World
Distribution by World Regions - 2014 Q2



Source: Internet World Stats - www.internetworldstats.com/stats.htm
Basis: 3,035,749,340 Internet users on June 30, 2014

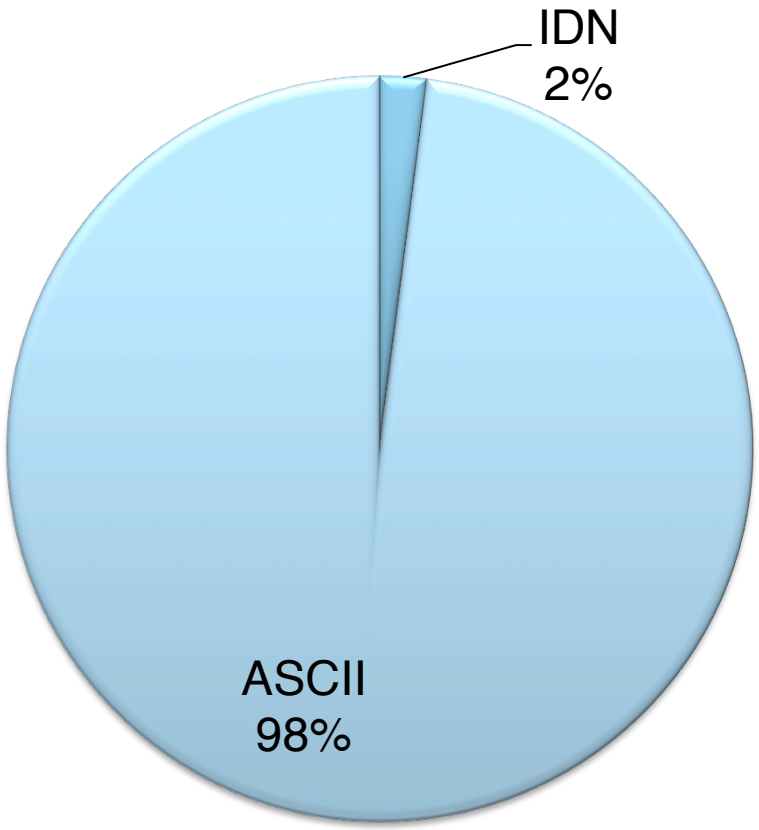
Source: http://internetworldstats.com/stats.htm, accessed 5 April 2015

Top 20 IDN registries: market share over time



Source: Word report on Internationalised Domain Names 2014, Unesco, Verisign, Eurid

Domain names: ASCII and IDN



Source: Word report on Internationalised Domain Names 2014, Unesco, Verisign, Eurid

IDNs are drivers of multilingualism

- IDNs enhance linguistic diversity in cyberspace
- Websites using IDNs are better predictors of language content than websites with an ASCII domain name

World report on Internationalised Domain Names 2014, Unesco, Verisign, Eurid

Status of IDN support in popular browsers

	Google Chrome	Microsoft Internet Explorer 11	Mozilla Firefox	Opera	Safari
Can be forced to always show the IDN URL?	No	No	No	No	No
Decides whether to show the IDN URL as a whole or label by label?[32]	Label by label	Label by label	Label by label	Based on the TLD only	Based on the script only
Contains a blacklist of characters that will prevent display of the IDN URL?	Yes	No	Yes	No	No
Has a configurable list that will allow display in specific languages?	Yes	Yes	No	No	No
Has a whitelist of TLDs and will only show the IDN for TLDs in the list?	No	No	Yes	Yes	No
Has a whitelist of scripts and will only show the IDN for scripts in the list?	No	No	Yes, with algorithmic exceptions	No	Yes
Allows for hybrid IDNs such as http://www.research.онлайн?	Yes	Yes	Yes	Yes	Yes

Source: World report on Internationalised Domain Names 2014, Unesco, Verisign, Eurid

Support for internationalized email addresses in top ten email clients across all platforms

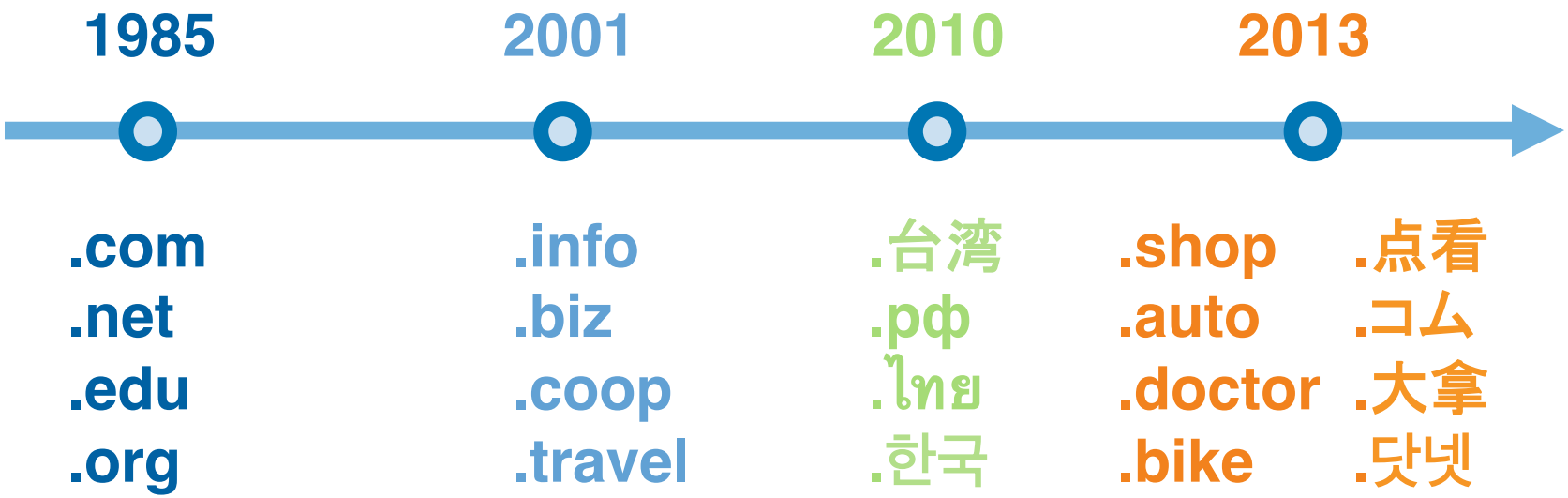
福祥@邓福祥.com

			Supports	
Market position	Client name	Share of market	International Email Addresses?[46]	Sending of International Email?[47]
1	Apple iPhone	26%	No	Yes
2	Outlook	14%	No	No
3	Apple iPad	12%	No	Yes
4	Gmail	12%	Yes	Yes
5	Apple Mail	8%	No	Yes
6	Google Android	6%	No	Yes
7	Outlook.com	6%	No	No
8	Yahoo! Mail	5%	No	No
9	Windows Live Mail	2%	No	No
10	Windows Mail	2%	No	No

Source: World report on Internationalised Domain Names 2014, Unesco, Verisign, Eurid

Universal Acceptance

Universal acceptance

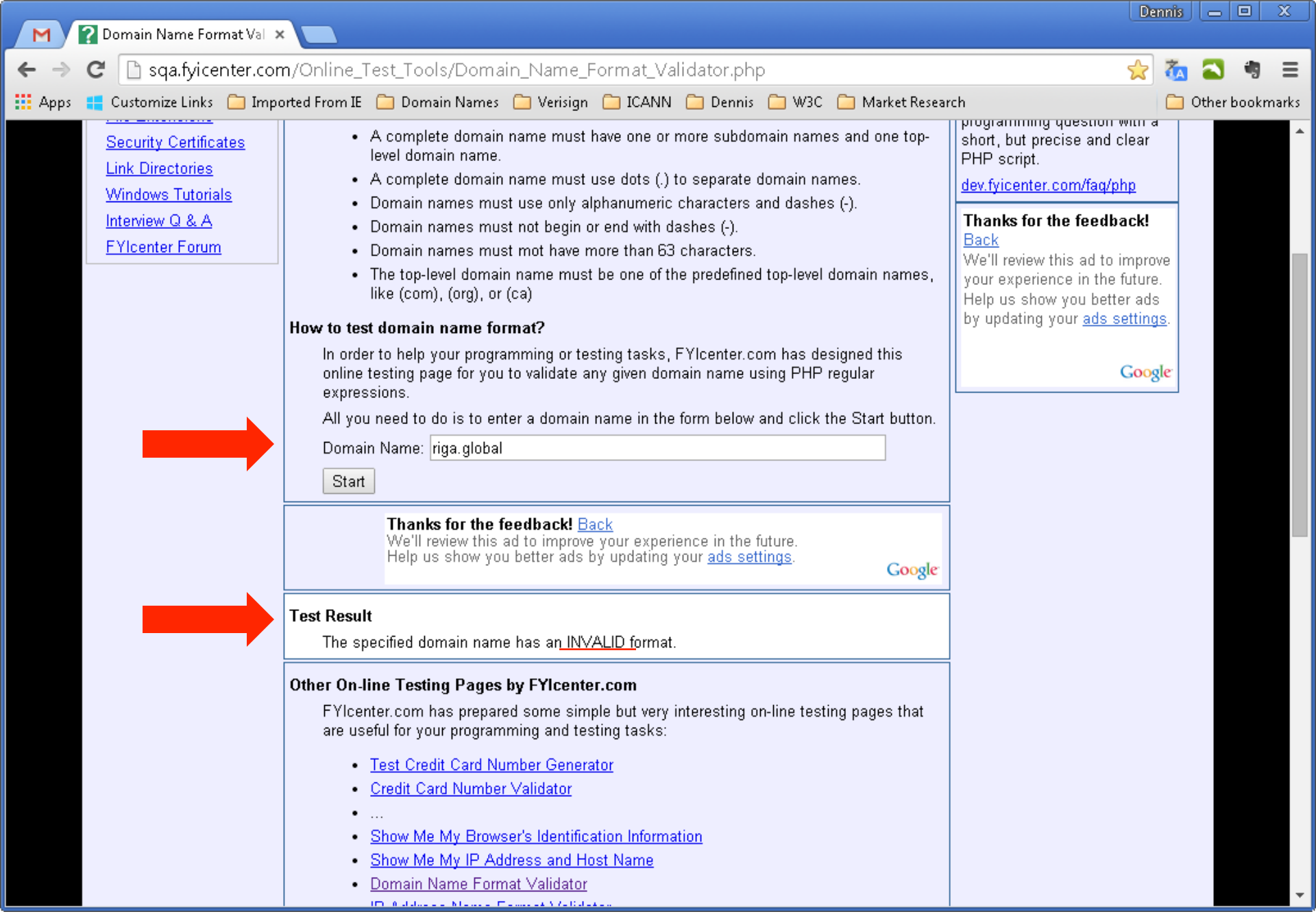


Software assumptions:
- TLD <= 3 characters
- ASCII only
- Hardcoded

Software assumptions:
- ~~TLD <= 3 characters~~
- ASCII only
- Hardcoded

Software assumptions:
- ~~TLD <= 3 characters~~
- ~~ASCII only~~
- ~~Hardcoded~~

Software assumptions:
- ~~TLD <= 3 characters~~
- ~~ASCII only~~
- ~~Hardcoded~~


Domain Name Format Val
sqa.fyicenter.com/Online_Test_Tools/Domain_Name_Format_Validator.php
Apps
Customize Links
Imported From IE
Domain Names
Verisign
ICANN
Dennis
W3C
Market Research
Other bookmarks
Security Certificates
Link Directories
Windows Tutorials
Interview Q & A
FYIcenter Forum
A complete domain name must have one or more subdomain names and one top-level domain name.
A complete domain name must use dots (.) to separate domain names.
Domain names must use only alphanumeric characters and dashes (-).
Domain names must not begin or end with dashes (-).
Domain names must mot have more than 63 characters.
The top-level domain name must be one of the predefined top-level domain names, like (com), (org), or (ca)
How to test domain name format?
In order to help your programming or testing tasks, FYIcenter.com has designed this online testing page for you to validate any given domain name using PHP regular expressions.
All you need to do is to enter a domain name in the form below and click the Start button.
Domain Name: riga.global
Start
Thanks for the feedback! Back
We'll review this ad to improve your experience in the future.
Help us show you better ads by updating your ads settings.
Google
Test Result
The specified domain name has an INVALID format.
Other On-line Testing Pages by FYIcenter.com
FYIcenter.com has prepared some simple but very interesting on-line testing pages that are useful for your programming and testing tasks:
Test Credit Card Number Generator
Credit Card Number Validator
...
Show Me My Browser's Identification Information
Show Me My IP Address and Host Name
Domain Name Format Validator
short, but precise and clear PHP script.
dev.fyicenter.com/faq/php
Thanks for the feedback!
Back
We'll review this ad to improve your experience in the future.
Help us show you better ads by updating your ads settings.
Google

Universal acceptance

- **Ubiquity**

✓ **Consistent**
✓ **Predictable**

Universal acceptance

- Ubiquity
- **Vicious circle**

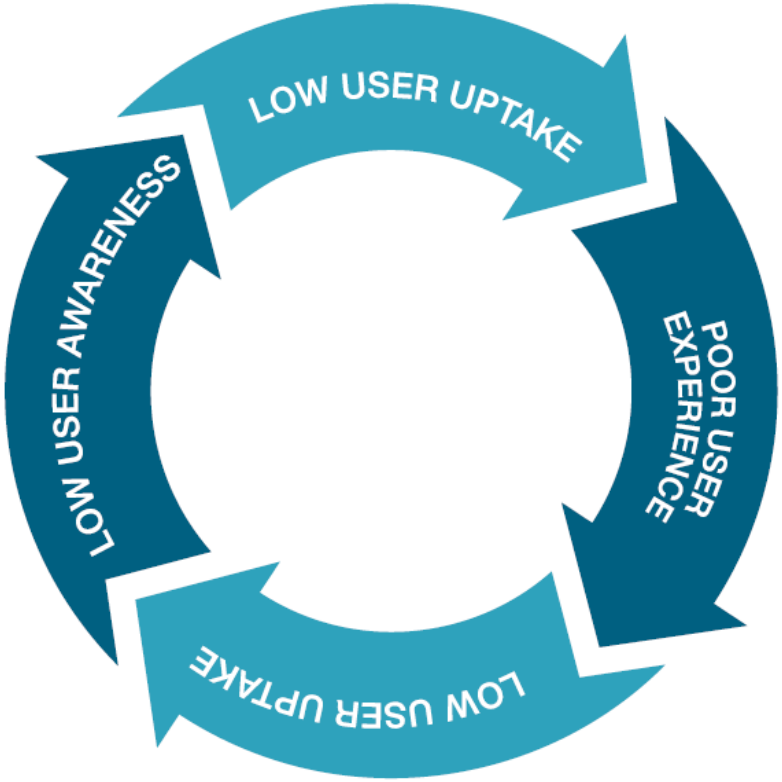


Image source: World Report on IDN Deployment 2013, Unesco, Verisign, Eurid

Universal acceptance

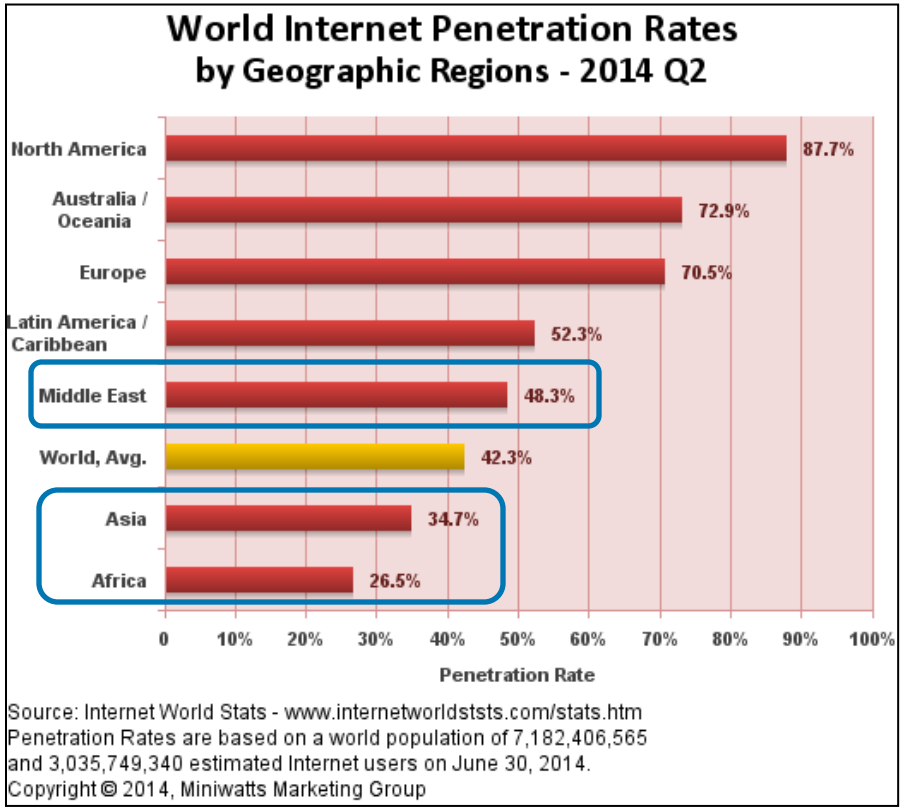
- Why is important?

  For the next generation of internet users

- Join the conversation

  icann.org/universalacceptance

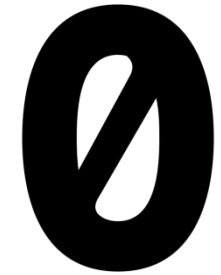
- Be part of the solution

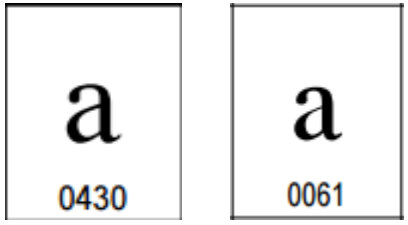


Confusables

What are confusable characters?

- Visually similar glyphs
  - Same script
    - Capital letter ‘I’, Lowercase letter ‘l’, Latin number ‘1’
    - Capital letter ‘O’, Latin number ‘0’
  - Different scripts
    - Cyrillic small letter ‘а’, Latin small letter ‘a’
    - Greek small letter ‘η’, Latin small letter ‘n’
    - Greek small letter ‘ρ’, Latin small letter ‘p’





Cyrillic Latin

The homograph issue

"... a way a malicious party may deceive computer users about what remote system they are communicating with, by exploiting the fact that many different characters look alike."

Wikipedia.org

rnicrosoft

coca-cola

Cyrillic and Greek characters

xn----gmb96ac0ecrbc

paypal

Cyrillic characters

xn--pypl-53dc

Preventing measures

Browser implementation techniques:

- Expose the ascii label (i.e. punycode) in all cases → bad user experience
  - Example:
    - user input = автосалон.com
    - Browser output = xn--80aaf1bkdcvh.com
- Show IDN if user language settings match IDN's script, show ascii label otherwise → better user experience
  - Example:
    - User input = 汽车经销商.net
    - Browser output = 汽车经销商.net

Preventing measures

Registry techniques:

- IDNA2008 (RFC's 5890, 5891, 5892, 5893)
  - Allowed code points in a domain name
- Unicode Technical Standard #39 (www.unicode.org/reports/tr39/)
  - Restriction-level detection
    - ASCII only
    - Single script
    - Highly restrictive (Latin/ Han, Hiragana, Katakana/Han, Bopomofo/ Han, Hangul)
    - Moderately restrictive (allow Latin with other, except Cyrillic and Greek)
    - Minimally restrictive (allow any mixture of scripts)
    - Unrestricted (allow any valid identifier, including symbols)

IDN conversion API and analysis tool

https://mctapi.verisign-grs.com

- A-label to U-label conversion and vice-versa
- Script analysis by label and code-point
- JSON format
- Available to public

```
[{
"input":"ascii",
"ascii":"xn--rsum-bsad.com",
"unicode":"résumé.com",
"unicodeLabels":["résumé","com"],
"codePoints":[["U+0072","U+00E9","U+0073","U+0075","U+006D","U+00E9"],["U+0063","U+006F","U+006D"]],
"scripts":[["Latin","Latin","Latin","Latin","Latin","Latin"],["Latin","Latin","Latin"]],
"scriptCombination":[["Latin"],["Latin"]],
"success":true
} ]
```

```
[{
"input":"unicode",
"ascii":"xn--jjz7c561f.com",
"unicode":"邓福祥.com",
"unicodeLabels":[" 邓福祥 ","com"],
"codePoints":[["U+9093","U+798F","U+7965"],["U+0063","U+006F","U+006D"]],
"scripts":[["Han","Han","Han"],["Latin","Latin","Latin"]],
"scriptCombination":[["Han"],["Latin"]],
"success":true
}]
```

```
[{
"input":"unicode",
"ascii":"xn--pypl53dc",
"unicode":"paypal",
"unicodeLabels":["paypal"],
"codePoints":[["U+0070","U+0430","U+0079","U+0070","U+0430","U+006C"]],
"scripts":[["Latin","Cyrillic","Latin","Latin","Cyrillic","Latin"]],
"scriptCombination":[["Latin","Cyrillic"]],
"success":true
}]
```

powered by

VERISIGN®