Internationalized Domain Names
-
Opportunities and Risks

Bill Jouris

A Little History

Domain Names

Repertoire
Conflicts

IDN
Numerous scripts/languages

Chinese	中文
Japanese	日本ご
Arabic	العَرَبِيَّة
Cyrillic (Russian)	Росси́я
Devanagari (Hindi)	हिन्दी
Georgian	ქართული
Lao	ລາວ
Ethiopic	አማርኛ

Risk 1- Communication

ᲙᲙᲗᲕᲘᲔ.com

ქართული ავტო.COM

The Latin Script is a Mess

- 210 languages
- 221 characters

Fonts

Times New Roman

Ariel

Courier New

a vs ɑ or g vs ɡ

c vs ƈ

Cases

- Simply Smaller X x Z z
- Very Similar P p Y y
- Totally Different A a *a* G g g

Upper Case and Lower Case

• Armenian	Բ բ	Ի ւ
• Cyrillic	В в	Е е
• Greek	Π π	Δ δ
• Latin	Z z	Q q

Diacritics

Grave Accent ◌̀

Acute Accent ◌́

Circumflex Above ◌̂

Tilde ◌̃

Macron ◌̄

Breve ◌̆

Caron ◌̌

Dot Above ◌̇

Diaeresis ◌̈

Double Acute ◌̋

Ring Above ◌̊

Hook Above

Horn ◌̛

Dot Below ◌̣

Comma Below ◌̦

Cedilla ◌̧

Ogonek ◌̨

Circumflex Below ◌̭

Macron Below ◌̱

A Small Slice

Any given language only uses a few diacritics.

- English: None? Don’t be naïve!
- Spanish: Tilde, Diaeresis, Acute
- French: Cedilla, Acute, Circumflex, Grave, Diaeresis

Variants:

Is cañon noticeably different from cañon, if you don’t know the Macron diacritic?

Is cafė noticeably different from cafè, if you don’t know the Dot Above diacritic?

30 variations of Letter O alone!

o	o̧	o̲	ò	ó	ô	õ	ö
ø	ō	ơ	ǒ	ő	ɔ	ȫ	o͜o
ọ	ọ̀	ọ́	ỏ	ố	ồ	ổ	ỗ
ộ	ớ	ờ	ở	ỡ	ợ		

Plus ð (eth)

Variants

“What you see is what you get” . . . only works if what you see is what you *think* you see.

What you want:

- As a user, go where you expect to go
- As a registrant, having your customers/users reliably come to your site, not go somewhere else

Variants vs Confusables vs Different
Why Do You Care?

Want to register something?

- Top Level Domain names (TLDs)
  - Variants – Blocked automatically
  - Confusables – Examined by the Similarity Review Team
  - Different – Just registers

- Second Level Domain Names

  (See later)

It’s not a Joke

Consider this domain name:

www.test.joƙe

Did you notice:

That the “K” isn’t just a K?

And the “J” isn’t actually a J at all?

ɟoƙe vs joke – not obviously different

Variants vs Confusables vs Different

What are they?

The “reasonably careful user”

.com	
.COM	ALL CAPS
.cом	Cyrillic
.coṃ	M with Dot below
.cơm	O with Horn
.çom	C with Cedilla
.cỗm	O with Circumflex and Tilde
.corn	C O R N

Cross-Script Variants
Related languages

- Cyrillic – Latin Variants 29 including:

Er	p	p	P
Es with descender	ҫ	ç	C with cedilla

- Greek – Latin Variants 18 including:

Nu	ν	v	V
Beta	ß	β	Sharp S

- Armenian – Latin Variants 7 including:

Seh	ց	g	G
Yiwn	ւ	ι	Iota

Cross-Script Variants
Generic Symbols

l	o	c	ɔ	Latin
l	o	c		Cyrillic
l	o			Hebrew
	o			Greek
	o			Armenian
◌ɿ	o	⊂	◌ɔ	Myanmar
		⊂	ↄ	Lao
◌l	O			Oriya

Can you distinguish .ooo from .ooo from .ooo from .ooo ?

Latin In-Script Variants

Schwa	ə	ə	Turned E
Iota	ɩ	ɪ	Dotless I
D with Caron	ď	ɗ	D with Hook
A with Breve	ă	ǎ	A with Caron
O with Diaeresis	ö	ő	O with Double Acute
Ligature AE	æ (*æ*)	œ	Ligature OE
K	k	ƙ	K with Horn

Underlining

www.example.com	No diacritics
www.example.com	L with Dot below
www.example.com	E with Macron below
www.example.test	S with Comma below
www.example.test	L with Circumflex below
www.example.test	M with Cedilla

Risk 2 – Confusion
Do you know the way to San Jose?

www.sanjose.gov

www.sanjosé.gov

Or

www.Munich.gov

www.München.gov

Risk 3 – DNS Abuse

The Easy Jet example

www.easyjet.com

www.easyiet.com

How Much Has This Happened?

- This is not a new problem

https://www.proofpoint.com/us/resources/white-papers/domain-fraud-report

www.FACEBO**O**K.com/31.13.77.36 : Facebook main site

Then there's:

1. www.FACEBO**O**Ķ.com www.xn--faceboo-bhb.com 50.63.202.25
2. www.FACEBO**Ò**K.com www.xn--facebok-i0a.com 198.54.117.212
3. www.FACEBO**Ó**K.com www.xn--facebok-q0a.com 185.53.179.6
4. www.FACEBO**Ô**K.com www.xn--facebok-y0a.com 208.73.211.178
5. www.FACEBO**Ö**K.com www.xn--facebok-f1a.com 78.108.52.211
6. www.FACEBO**Ø**K.com www.xn--facebok-v1a.com 198.89.124.40
7. www.FACEBO **O** Ķ.com www.xn--faceboo-bhb.com 50.63.202.25

Risk 4 – False Equivalence

Reference Label Generation Rulesets (LGRs) for the Second Level

What, exactly, is a variant?

DNS Abuse – A New Level

ᲙᲝᲙᲐᲘᲜ.com

ქართული ავტო.COM

Questions?