SaudiNIC: Supporting Arabic Domain Names

Raed Alfayez, SaudiNIC
Middle East DNS Forum, Ankara, Apr 2018
About SaudiNIC

Introduction

SaudiNIC’s major efforts

What is missing?
About SaudiNIC

• Administering the domain name space under:
  – (.sa) since 1995
  – (السعودية) since 2010.

• Operated by a government organization:
  – CITC (Communication and Information Technology Commission)

• Coordinating with regional and international bodies in order to present the local community needs

• Leading the local and regional communities efforts towards supporting Arabic language in Domain Names since 2001 (more than 15 years of experience)
About SaudiNIC

54,813 Domain names
2,521 Arabic domains under (السعودية)
465 Variant
2LD/3LD Domain Names Distribution %

- com.sa: 54%
- sa: 27%
- pub.sa: 0.29%
- med.sa: 0.76%
- sch.sa: 1.4%
- edu.sa: 2.6%
- net.sa: 3%
- org.sa: 3.3%
- gov.sa: 3.5%
- السعودية: 4.6%
Introduction: Arabic Language

• Ranked as the 5\textsuperscript{nd} language by native speakers in the world.
  – Native speakers: 295 million

• Considered as Official/Co-official language in 25 countries

• 28 letters in the Arabic alphabet

Source: http://en.wikipedia.org/wiki/Arabic_script
Introduction: Arabic Alphabet


<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>أ</td>
<td>A</td>
</tr>
<tr>
<td>ب</td>
<td>B</td>
</tr>
<tr>
<td>ت</td>
<td>T</td>
</tr>
<tr>
<td>ث</td>
<td>Th</td>
</tr>
<tr>
<td>ج</td>
<td>G</td>
</tr>
<tr>
<td>ح</td>
<td>H</td>
</tr>
<tr>
<td>خ</td>
<td>Kh</td>
</tr>
<tr>
<td>د</td>
<td>D</td>
</tr>
<tr>
<td>ذ</td>
<td>Zh</td>
</tr>
<tr>
<td>ر</td>
<td>R</td>
</tr>
<tr>
<td>ز</td>
<td>Z</td>
</tr>
<tr>
<td>ص</td>
<td>Sh</td>
</tr>
<tr>
<td>ض</td>
<td>St</td>
</tr>
<tr>
<td>ش</td>
<td>Sh</td>
</tr>
<tr>
<td>ط</td>
<td>Tt</td>
</tr>
<tr>
<td>ض</td>
<td>St</td>
</tr>
<tr>
<td>ع</td>
<td>U</td>
</tr>
<tr>
<td>غ</td>
<td>Gh</td>
</tr>
<tr>
<td>ف</td>
<td>F</td>
</tr>
<tr>
<td>ق</td>
<td>Q</td>
</tr>
<tr>
<td>إ</td>
<td>E</td>
</tr>
<tr>
<td>ئ</td>
<td>E</td>
</tr>
<tr>
<td>ي</td>
<td>Y</td>
</tr>
<tr>
<td>و</td>
<td>W</td>
</tr>
<tr>
<td>ه</td>
<td>H</td>
</tr>
<tr>
<td>ة</td>
<td>Th</td>
</tr>
<tr>
<td>م</td>
<td>M</td>
</tr>
<tr>
<td>ن</td>
<td>N</td>
</tr>
<tr>
<td>ل</td>
<td>L</td>
</tr>
<tr>
<td>ك</td>
<td>K</td>
</tr>
<tr>
<td>ميم</td>
<td>M</td>
</tr>
<tr>
<td>لام</td>
<td>L</td>
</tr>
<tr>
<td>كاف</td>
<td>K</td>
</tr>
<tr>
<td>وو</td>
<td>W</td>
</tr>
<tr>
<td>ها</td>
<td>H</td>
</tr>
<tr>
<td>نن</td>
<td>N</td>
</tr>
</tbody>
</table>
Introduction: Arabic Script

- The 2nd most widely used alphabetic writing system in the world
- Used by many languages such as:
  - Arabic, Urdu, Persian, Turkish, Kurdish, Pashto, ... etc
- It is widely used by more than 43 countries
  - more than one billion potential users could be concerned in using Arabic script domain names.

Source: http://en.wikipedia.org/wiki/Arabic_script
Arabic Script IDNs Major Issues

1. Combining Marks
2. Diacritics
3. World/label separators (space, ZWNJ, ZWJ, hyphen)
4. Digits
5. Confusing similar characters (e.g. variant tables)
6. Bidirectional
There are a number of groups of characters that have the same shapes (Homoglyph), eg.:

- Kaf group,
- Heh group,
- Yeh group,
- Alef group
- ...
Main issues: Variants

• There are 64 “variants” for “Google.com” domain due to lower/upper case of ASCII letters.
  – If you type any of them you will reach the same site
  – The solution was done by DNS protocols
  – All are allocated and delegated

• But this is not the case for other languages!
  – Arabic (کلی) vs. Urdu (کلی)!
  – Arabic (إنترنت) vs Arabic (انترننت)

Example of ASCII Variants

Google.com
gOogle.com
goOgle.com
gooGle.com
GooGle.com
GooglE.com
...etc.
Main issues: Variants

- Exponential number of variants!!!

<table>
<thead>
<tr>
<th>Label</th>
<th>Approximately # of variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>اتصال</td>
<td>300</td>
</tr>
<tr>
<td>اتصالات</td>
<td>6,000</td>
</tr>
<tr>
<td>الاتصالات</td>
<td>60,000</td>
</tr>
<tr>
<td>هيئة-الاتصالات</td>
<td>2,879,999</td>
</tr>
<tr>
<td>هيئة-الاتصالات-تقنية-المعلومات</td>
<td>82,944,000,000</td>
</tr>
</tbody>
</table>
SaudiNIC’s Major Efforts

Arabic IDN pilot projects
- Arab League (2005 - 2009)
- Language & Variant Tables

Tools, algorithms and solutions to manage variants:
- Master Key Algorithm
- Filters
- Variant Management System (VMS)

IDN Assessment Reports

Arabic Email Project (Raseel)

SaudiNIC’s Major Efforts
Arabic IDN pilot projects

• RFC: Linguistic Guidelines for the Use of the Arabic Language in Internet Domains
  – https://www.rfc-editor.org/rfc/rfc5564.txt

• For more information
Arabic IDN pilot projects

- Language & Variant Tables

This document provides the IDN (Internationalized Domain Names) Language Table and guideline rules to be used for writing and registering Arabic Domain names in accordance with the recommendations outlined in the RFC 5564: "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains". This can be found in the following URLs: arabic-language.tld, Saudi Network Information Center Raed Al Fazy (raed@tld), Saudi Network Information Center Abdulrahman L Al-Ghadir (aghadir@tld).

References:
- The Unicode Standard, Version 6.3.0
- Guideline rules for writing Arabic IDNs under the IDN ccTLD (Arabic). http://nic.net.sa/docs/Guidelines_for_writing_Arabic_IDNs_under_the_IDNccTLD_v1.2-en.pdf
SaudiNIC’s Major Efforts

Arabic IDN pilot projects
- Arab League (2005-2009)
- Language & Variant Tables

Tools, algorithms and solutions to manage variants:
- Master Key Algorithm
- Filters
- Variant Management System (VMS)

IDN Assessment Reports

Arabic Email Project (Raseel)
Tools and solutions: Compare Characters

- Display all code points of the whole Arabic script in one page
- Give the ability to compare code points based on their position
- It helped us to study the behavior of the code points and compare them against each other, in order to build our LT and VT.
SaudiNIC’s VMS

• An easy and stable variant management system:
  • No language mixing (utilizing the powerful tools: Language tables)
    – control input via the user interface
    – help identifying “must-be-allocated” variants for reachability purposes.
    – tremendously reduce the number of unnecessary allocateable variants
    – protect the TLD-space.
  • Master Key algorithm
    – Easily manage the whole variants list with one unique identifier
    – Speed up the lookup process
    – Eliminate the need of saving all possible variants
  • Must be allocated variants
    – For reachability purposes, “must-be-allocated” variants should be generated and activated automatically by the registry, so that: registered domain name is accessed regardless of the input devices (language table) being used by the navigator users.
  • Filters
    – To identify desired allocatable variants
SaudiNIC’s VMS: Registrant will use his/her keyboard

U+0645 U+0643 U+0629
U+0645 U+0643 U+06A9
U+0645 U+0647
U+0645 U+06C1
## SaudiNIC’s VMS: blocking quality??

<table>
<thead>
<tr>
<th>IDN</th>
<th>Total Variants</th>
<th>Allocatable</th>
<th>Blocked</th>
<th>Blocked due to Language Mixing</th>
</tr>
</thead>
<tbody>
<tr>
<td>مكة-المكرمة</td>
<td>3239</td>
<td>34</td>
<td>3205</td>
<td>3181 (99.25%)</td>
</tr>
<tr>
<td>القرآن-الكريم</td>
<td>11999</td>
<td>111</td>
<td>11888</td>
<td>11836 (99.56%)</td>
</tr>
<tr>
<td>هيئة-الإعلام</td>
<td>47999</td>
<td>81</td>
<td>47918</td>
<td>47764 (99.68%)</td>
</tr>
<tr>
<td>كهف-الياسمين</td>
<td>28799</td>
<td>65</td>
<td>28734</td>
<td>28680 (99.81%)</td>
</tr>
<tr>
<td>كهف-اكيا</td>
<td>21599</td>
<td>47</td>
<td>21552</td>
<td>21534 (99.92%)</td>
</tr>
</tbody>
</table>
SaudiNIC’s Major Efforts

Arabic IDN pilot projects
• GCC Pilot Project (2004-2005)
• Arab League (2005 - 2009)
• Language & Variant Tables

Tools, algorithms and solutions to manage variants:
• Master Key Algorithm
• Filters
• Variant Management System (VMS)

IDN Assessment Reports

Arabic Email Project (Raseel)
IDN Assessment Reports

Conducted and Published a number of IDN Assessment Reports:

2007
- **IDN Top Level Domain Evaluations and Testing Report**
- with the cooperation of the Arabic Domain Name Pilot Project Team.

2010
- Arabic IDN Test Results for Browsers
- Mozilla Firefox & **Microsoft IE**

2014
- **IDN Assessment Report**
SaudiNIC’s Major Efforts

Arabic IDN pilot projects
- Arab League (2005-2009)
- Language & Variant Tables

Tools, algorithms and solutions to manage variants:
- Master Key Algorithm
- Filters
- Variant Management System (VMS)

IDN Assessment Reports

Arabic Email Project (Raseel)

SaudiNIC’s Major Efforts
Raseel: An Arabic Email System

• Phase I (2010~2013):
  – A pilot project to test Arabic email addresses
  – Built before the EAI RFCs
    • Using a hack: convert the user part of the email address to Punycode
    • Implemented plugins for Outlook and Roundcube to display the Arabic addresses correctly.
  – Work with existing Email Servers and old RFCs.
Raseel: An Arabic Email System

• Phase II (2016+):
  – Built based on the new EAI RFCs using standard EAI addresses
    • Postfix, Horde/Roundcube and Archiveopteryx
  – Still in a beta version and not open for public.
  – Successful test internally and with Gmail and MS Outlook.
  – No need for plugins.
Raseel: An Arabic Email System
Almost 6+ years since the EAI RFCs were published and until now there are almost very limited support in:
- Email servers (SMTP, IMAP, POP),
- Email providers (Hotmail, Yahoo)
- Emails clients (Webmail, Application)

Need to have a protection mechanism for the user part of the emails addresses (similar to IDN variants)

<table>
<thead>
<tr>
<th>Farsi Yeh (U+06CC)</th>
<th>Arabic Yeh(U+064A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>البريد@رسيل.السعودية</td>
<td>البريد@رسيل.السعودية</td>
</tr>
</tbody>
</table>

Automatic tools to configure and manage variants (Domain, User Accounts).

Boosting the adoption of the new EAI RFC by ISP and service/hosting providers.
WHAT IS MISSING?
What is missing?

Main Players

- Application Vendor
- Registrant/User
- Service Provider
- Registry/Registrar
Usability
- The proper selection of available variants to be enabled at the language or script levels.
- Why allocatable variants are not the same across different IDN-TLD registries?

Lack of awareness
- IDN is not well-known or recognized
- Misconception about IDNs
  - No need for IDNs!
  - No need for ASCII domain names!!
  - IDNs Benefits are not realized
  - IDN will make isolated internets
- IDN variants are not understood or known
- Variant tables are not known and universal
- Variant management challenges
- Hosting and using IDNs
- Hosting and using IDN variants
- Registrant will use his/her keyboard based on his/her language knowledge.

Technical issues
- Working with IDNs and ASCII equivalents (i.e., “xn--” versions of IDN domain names)
- Working or editing IDN’s zone files.
- Different input devices (keyboards)
- No standard or common way to work with IDN and its variants
- Language specific issues (Right-to-left direction, display, interface)
Full IDN support
- Ignoring the fact that IDNs are generated using input devices (keyboards) that are language based
- Therefore, some (registries) are adopting only one huge script-wide table (with mixed languages)
- No easy or direct way to enable variants
- There should be one key for all variants
- For enhancing and decreasing the number of variants they should be based on character positions
- International Reachability
- Study the whole Script
- No easy interface for variant management
- Consistent variant list
- Whois should support IDN and Variants

Lack of IDN support as part of registration information, e.g.:
- Registration system (web-interface and EPP)
- Refuse IDN Email (EAI)
- Refuse IDN Name servers
- Tools provided by the registry (whois, DNS checker, etc.).
- Sending and receiving IDN emails (EAI)

Security issues
- Need protection to the TLD name space
- Missing some code points when defining variants
- Variants for the reserve list

Cross Registries
- Unified terminologies
- There should be a common (agreed) protection to secure any name space.
Hosting
- Does not accept IDN
- How to point all variants to the main domain (e.g. web/mail)?
- The absence of technical solutions to host variants automatically.
- Create account and login
- Accept IDN Email (EAI)
- Accept IDN Name servers

Web Email Services
- Sending and receiving EAI
- Display IDN and EAI correctly

Others
- Create account and login
- IDN Email

DNS Hosting
- Create account and login
- Accept IDN Email (EAI)
- Accept IDN Name servers
Email clients
• Handling IDN and EAI in the View mode
• Handling IDN and EAI in the Edit mode
• Handling IDN and EAI when doing copy/paste
• Handling IDN and EAI in All fields
  (To, CC, BCC, body, etc.)
• Handling IDN and EAI in Address book

Social Media
• Create account
• IDN email
• A-label email

Software Tools
• IDN hosts

Others
• Webmaster
• Content Management
• Security

Search Engine
• display results
• site search

Browsers
• Reachability
• Visibility
• Ability to edit
• Auto completion
• Ability to copy and paste
• Text alignment
• SSL information
• Protocol representation
• URL auto recognition
• Ability to click

Office Applications
• Clickable
• Auto recognition
• Copy and paste

Email clients
• Handling IDN and EAI in the View mode
• Handling IDN and EAI in the Edit mode
• Handling IDN and EAI when doing copy/paste
• Handling IDN and EAI in All fields
  (To, CC, BCC, body, etc.)
• Handling IDN and EAI in Address book
Variants enablement must be done in every level

Registry

DNS Hosting

Email Services

Web Hosting

Register and enable variants:
- مكة
- مكة
- مكة

Configure DNS & add need RRs (e.g. NS & A & CNAME) for:
- xn--ogb5cf
- xn--ogb9c4p
- xn--hhb4rwc

Configure Email account and email aliases:
- رائد@مكة
- رائد@مكة
- رائد@مكة

Configure web-server and account and aliases:
```
<VirtualHost 10.10.10.10>
    DocumentRoot /makkah
    ServerName xn--ogb5cf
    ServerAlias xn--ogb9c4p
    ServerAlias xn--hhb4rwc
</VirtualHost>
```
Gift

- Published “SaudiNIC’s Best Practices in Supporting and Managing Arabic Domain Names”

http://www.nic.sa/docs/SaudiNIC_ADNBP.pdf
Thank you

للمزيد من المعلومات يمكنكم زيارة:
For more information you can visit:

سجل.السعودية

nic.sa

هيئة الاتصالات وتقنية المعلومات
Communications and Information Technology Commission

citic.gov.sa