

# SaudiNIC:

## Supporting Arabic Domain Names

Raed Alfayez, SaudiNIC

ICANN60, Abu Dhabi, Oct 2017

# Agenda

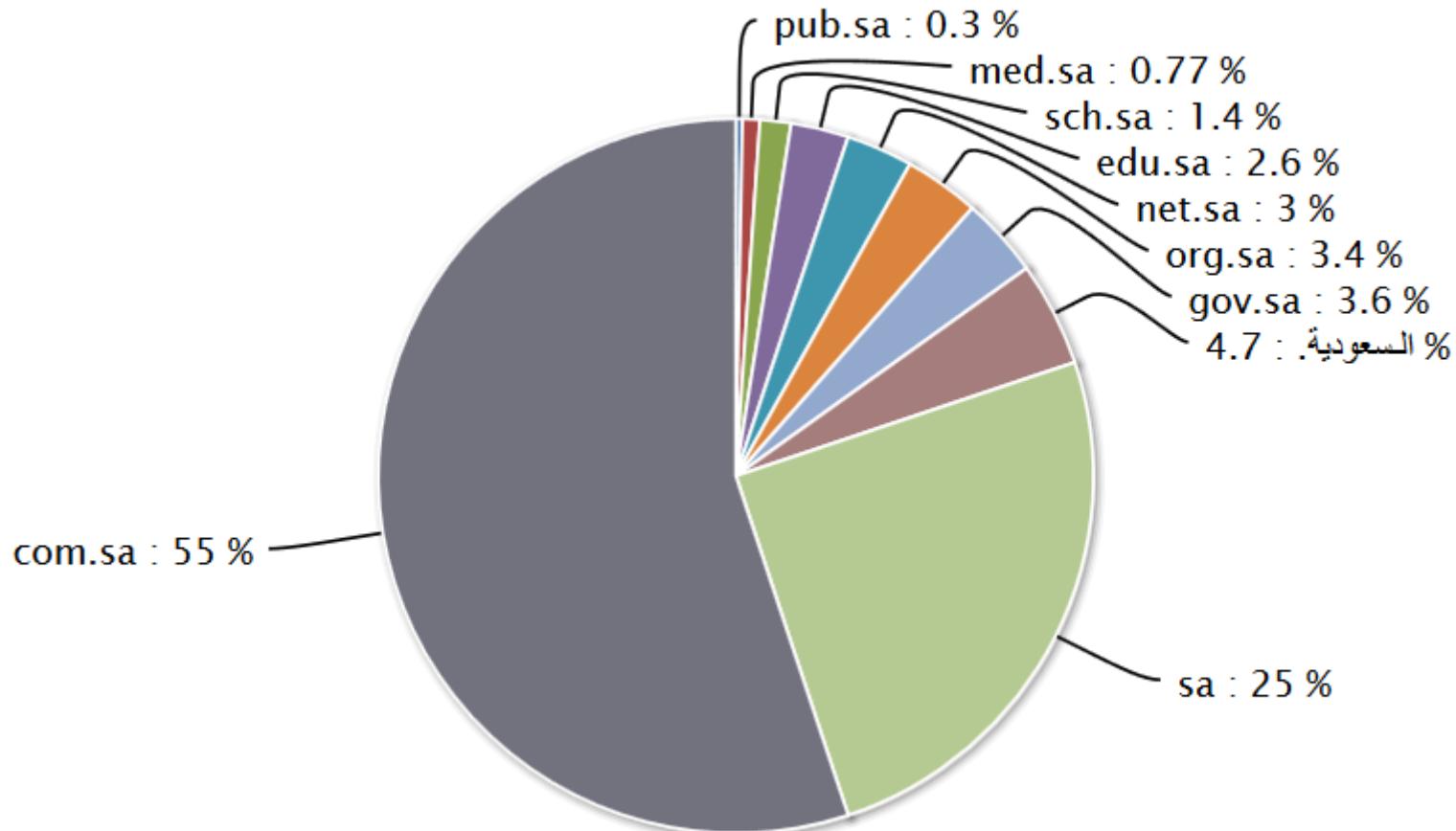
- About SaudiNIC
- Introduction
- SaudiNIC's major efforts
- What is missing?

# 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)

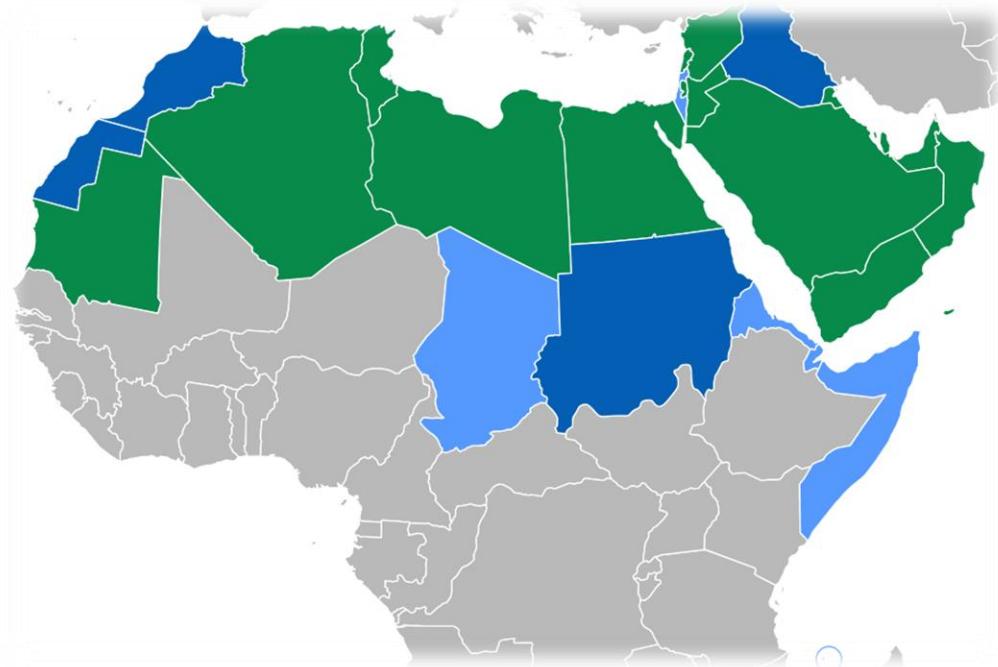
# About SaudiNIC

**50,813 Domain names  
2LD/3LD Domain Names Distribution %**



# Introduction: Arabic Language

- Ranked as the **5<sup>nd</sup>** language by native speakers in the world.
  - Native speakers: 295 million
- Considered as Official/Co-official language in **25** country



Source: [http://en.wikipedia.org/wiki/Arabic\\_script](http://en.wikipedia.org/wiki/Arabic_script)

# Introduction: Variants within the language

خ

Hā

ح

Hā

ج

Ǧim

ث

Tā

ت

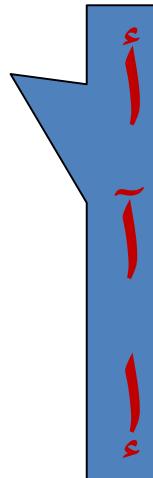
Tā

ب

Bā

ا

Alif



ص

Ṣād

ش

Šīn

س

Sīn

ز

Zāī

ر

Rā

ذ

Dāl

د

Dāl

ق

Qāf

ف

Fā

غ

Ǧain

ع

‘Ain

ظ

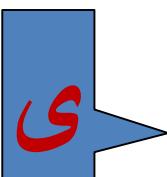
Zā

ط

Tā

ض

Dād



ي

Yā

و

Wāw

ه

Hā



ن

Nūn

م

Mīm

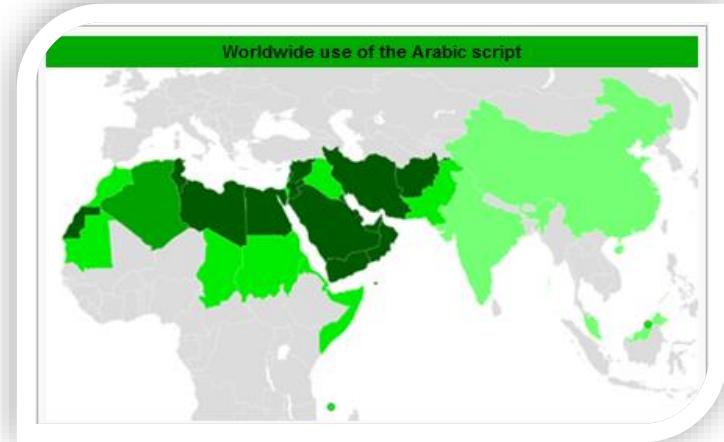
ل

Lām

ك

Kāf

# Introduction: Arabic Script



- The **2<sup>nd</sup>** 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.

# Arabic Script IDNs Major Issues

1. Combining Marks
2. Diacritics
3. Word/label separators (space, ZWNJ, ZWJ, hyphen)
4. Digits
5. Confusing similar characters  
(e.g. variant tables)
6. Bidirectional

Digit

1. European digits	U+0030 .. U+0039	(0123456789)
2. Arabic-Indic digits	U+0660 .. U0669	(۰۱۲۳۴۵۶۷۸۹)
3. Eastern Arabic-Indic digits	U+06F0 .. U+06F9	(۰۱۲۳۴۵۶۷۸۹)

Non-spacing Marks	
و	جُمِعَيْة

bidirectional



ZWNJ/ZWJ

Examples not using ZWNJ	Examples not using ZWNJ
طبل	طبل
input[0] = U+0637	input[0] = U+0637
input[1] = U+0628	input[1] = U+200c
input[2] = U+0644	input[2] = U+0628
input[3] = U+0644	input[3] = U+0644

Combining Marks

ى	+	ُ	=	ئِ نَسْتَرِي	is confusing with	ئِ نَسْتَرِي
U+0649		U+0654		U+0649 U+0654		U+0626
Description: Alef Maksura + Hamza Above <> Yeh With Hamza Above						
Comments: This is a Unicode confusable!						

ى	+	ُ	=	ئِ نَسْتَرِي	is confusing with	ئِ نَسْتَرِي
U+06cc		U+0654		U+06cc U+0654		U+0626
Description: Farsi Yeh + Hamza Above <> Yeh With Hamza Above						
Comments: This is Unicode confusable!						

# Main issues: Confusing Similar Characters

- There are a number of **groups** of characters that have the **same shapes** (**Homoglyph**), eg.:
  - Kaf group,
  - Heh group,
  - Yeh group,
  - Alef group
  - ...

Arabic																
0600		06FF														
0	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	۰۶۰۰	۰۶۱۰	۰۶۲۰	۰۶۳۰	۰۶۴۰	۰۶۵۰	۰۶۶۰	۰۶۷۰	۰۶۸۰	۰۶۹۰	۰۶۰۱	۰۶۰۲	۰۶۰۳	۰۶۰۴	۰۶۰۵	۰۶۰۶
۱	۰۶۱۱	۰۶۲۱	۰۶۳۱	۰۶۴۱	۰۶۵۱	۰۶۶۱	۰۶۷۱	۰۶۸۱	۰۶۹۱	۰۶۰۲	۰۶۰۳	۰۶۰۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸
۲	۰۶۱۲	۰۶۲۲	۰۶۳۲	۰۶۴۲	۰۶۵۲	۰۶۶۲	۰۶۷۲	۰۶۸۲	۰۶۹۲	۰۶۰۳	۰۶۰۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹
۳	۰۶۱۳	۰۶۲۳	۰۶۳۳	۰۶۴۳	۰۶۵۳	۰۶۶۳	۰۶۷۳	۰۶۸۳	۰۶۹۳	۰۶۰۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۰
۴	۰۶۱۴	۰۶۲۴	۰۶۳۴	۰۶۴۴	۰۶۵۴	۰۶۶۴	۰۶۷۴	۰۶۸۴	۰۶۹۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۱	۰۶۱۲
۵	۰۶۱۵	۰۶۲۵	۰۶۳۵	۰۶۴۵	۰۶۵۵	۰۶۶۵	۰۶۷۵	۰۶۸۵	۰۶۹۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۰	۰۶۱۱	۰۶۱۲
۶	۰۶۰۶	۰۶۱۶	۰۶۲۶	۰۶۳۶	۰۶۴۶	۰۶۵۶	۰۶۶۶	۰۶۷۶	۰۶۸۶	۰۶۹۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۰	۰۶۱۱	۰۶۱۲
۷	۰۶۰۷	۰۶۱۷	۰۶۲۷	۰۶۳۷	۰۶۴۷	۰۶۵۷	۰۶۶۷	۰۶۷۷	۰۶۸۷	۰۶۹۷	۰۶۰۸	۰۶۰۹	۰۶۰۱	۰۶۰۲	۰۶۰۳	۰۶۰۴
۸	۰۶۰۸	۰۶۱۸	۰۶۲۸	۰۶۳۸	۰۶۴۸	۰۶۵۸	۰۶۶۸	۰۶۷۸	۰۶۸۸	۰۶۹۸	۰۶۰۹	۰۶۱۰	۰۶۱۱	۰۶۱۲	۰۶۱۳	۰۶۱۴
۹	۰۶۰۹	۰۶۱۹	۰۶۲۹	۰۶۳۹	۰۶۴۹	۰۶۵۹	۰۶۶۹	۰۶۷۹	۰۶۸۹	۰۶۹۹	۰۶۰۱	۰۶۰۲	۰۶۰۳	۰۶۰۴	۰۶۰۵	۰۶۰۶
A	۰۶۰۱	۰۶۱۱	۰۶۲۱	۰۶۳۱	۰۶۴۱	۰۶۵۱	۰۶۶۱	۰۶۷۱	۰۶۸۱	۰۶۹۱	۰۶۰۲	۰۶۰۳	۰۶۰۴	۰۶۰۵	۰۶۰۶	۰۶۰۷
B	۰۶۰۲	۰۶۱۲	۰۶۲۲	۰۶۳۲	۰۶۴۲	۰۶۵۲	۰۶۶۲	۰۶۷۲	۰۶۸۲	۰۶۹۲	۰۶۰۳	۰۶۰۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸
C	۰۶۰۳	۰۶۱۳	۰۶۲۳	۰۶۳۳	۰۶۴۳	۰۶۵۳	۰۶۶۳	۰۶۷۳	۰۶۸۳	۰۶۹۳	۰۶۰۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹
D	۰۶۰۴	۰۶۱۴	۰۶۲۴	۰۶۳۴	۰۶۴۴	۰۶۵۴	۰۶۶۴	۰۶۷۴	۰۶۸۴	۰۶۹۴	۰۶۰۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۰
E	۰۶۰۵	۰۶۱۵	۰۶۲۵	۰۶۳۵	۰۶۴۵	۰۶۵۵	۰۶۶۵	۰۶۷۵	۰۶۸۵	۰۶۹۵	۰۶۰۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۱	۰۶۱۲
F	۰۶۰۶	۰۶۱۶	۰۶۲۶	۰۶۳۶	۰۶۴۶	۰۶۵۶	۰۶۶۶	۰۶۷۶	۰۶۸۶	۰۶۹۶	۰۶۰۷	۰۶۰۸	۰۶۰۹	۰۶۱۰	۰۶۱۱	۰۶۱۲

# 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 (كلى) vs. Urdu (کلی)!
  - Arabic (إنترنت) vs Arabic (انترنت)

## Example of ASCII Variants

Google.com  
gOogle.com  
goOgle.com  
gooGle.com  
GooGLE.com  
GooglE.com  
...etc.

کلی

input[0] = U+0643  
input[1] = U+0644  
input[2] = U+0649

کلی

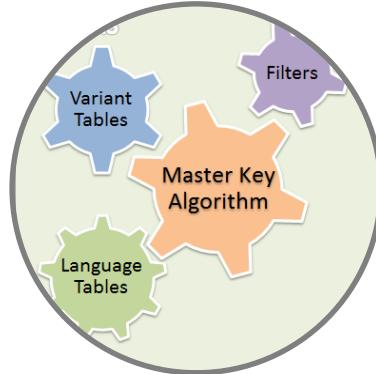
input[0] = U+06a9  
input[1] = U+0644  
input[2] = U+06cc

# SaudiNIC's Major Efforts



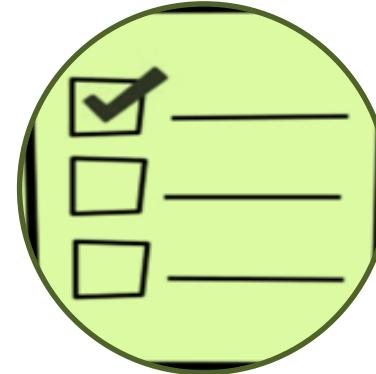
## Arabic IDN pilot projects

- 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)

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
  - <http://arabic-domains.org/en/>

# Arabic IDN pilot projects

- Language & Variant Tables

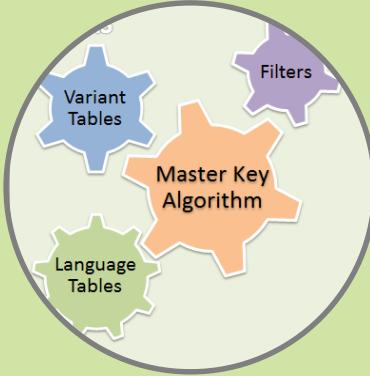
```
- <igr>
- <meta>
<version>2</version>
<date>2016-06-05</date>
<language>ar</language>
<scope type="domain">xn--mgberp4a5d4ar</scope>
- <description type="text/plain">
  This document provides the IDN (Internationalized Domain Names) Language Table and guideline rules to be used for writing and registering 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", that can be found in the following URLs:
  (azoman[at]cict.gov.sa) , Saudi Network Information Center Raed Al-Fayez (rfayez[at]cict.gov.sa) , Saudi Network Information Center Abdulrahman I. AL-Ghadir (aghadi[...]
</description>
<validity-start>2014-12-07</validity-start>
<validity-end>2020-12-07</validity-end>
<unicode-version>6.3.0</unicode-version>
- <references>
  <reference id="0">The Unicode Standard, Version 6.3.0</reference>
- <reference id="1">
  RFC 5564: "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains", http://tools.ietf.org/html/rfc5564
</reference>
- <reference id="2">
  Guideline Rules for writing Arabic IDNs under the IDN ccTLD (السعودية), http://nic.net.sa/docs/Guidelines_for_writing_Arabic_IDNs_under_the_IDN_ccTLD_V1.2-en.pdf
</reference>
</references>
</meta>
- <data>
  <char cp="0621" ref="0 1 2"/>
- <char cp="0622" ref="0 1 2">
  <var cp="0623" type="allocate" when="arabic-final-right-join" comment="Language variant"/>
  <var cp="0623" type="allocate" when="arabic-isolated-right-join" comment="Language variant"/>
  <var cp="0625" type="allocate" when="arabic-final-right-join" comment="Language variant"/>
  <var cp="0625" type="allocate" when="arabic-isolated-right-join" comment="Language variant"/>
  <var cp="0627" type="activate" when="arabic-final-right-join" comment="Language variant"/>
  <var cp="0627" type="activate" when="arabic-isolated-right-join" comment="Language variant"/>
  <var cp="0671" type="block" when="arabic-final-right-join" comment="Typo variant"/>
  <var cp="0671" type="block" when="arabic-isolated-right-join" comment="Typo variant"/>
  <var cp="0672" type="block" when="arabic-final-right-join" comment="Typo variant"/>
  <var cp="0672" type="block" when="arabic-isolated-right-join" comment="Typo variant"/>
  <var cp="0675" type="block" when="arabic-final-right-join" comment="Typo variant"/>
```

# 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)

SaudiNIC's Major Efforts

# 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.

The screenshot shows two windows related to character comparison:

- Left Window (Compare Characters):** A grid-based interface for comparing characters. The columns are labeled with code points: 060, 061, 062, and 063. The rows are labeled with numbers 0, 1, 2, 3, and 4. Each cell contains a small square icon representing a character, with a larger version of the character shown below it. Row 0 contains the characters ي (Yeh), ؤ (Tatweel), ي (Yeh), and ذ (Zay). Row 1 contains ي (Yeh), ؤ (Tatweel), و (Waw), and ر (Reh). Row 2 contains ي (Yeh), ؤ (Tatweel), آ (Alef Maqnoon), and ز (Zay). Row 3 contains ي (Yeh), ؤ (Tatweel), أ (Alef), and س (Sin). Row 4 contains ي (Yeh), ؤ (Tatweel), ف (Fathah), and ش (Shay).
- Right Window (Arabic-Script-IDN-Page2):** A web-based application for comparing characters. It includes a search bar, a "Select joiner" dropdown set to "Tatweel", and a "Add Combining Character" input field. Below these are sections for "Representative shape in code chart" (showing ي (Yeh) at 064A and LETTER YEH), "Possible shapes in context [isolated, final, medial, initial]" (showing various forms of ي (Yeh)), and "Font" (set to "Arabic Typesetting").

# Tools and solutions: Master Key Algorithm

- Secures the domain name space for the registry, **speeds up lookup** process and **minimizes storage** space:
  - Generates a **unique key** for a domain name label and all of its possible variants
  - the key can be used in the **lookup process** for both:
    - Domain name availability
    - Variants generation and allocation
- Supports multiple languages in a registry and it is easy to add a new language in the future
  - It requires a Language table (LT) and a Variant table (VT) for each supported language
- Provides automatic blocking of variants due to language mixing
- Supports defining variants based on character position
- Classify the relationship between variants (Exact /Typo/InterReach)
- ...etc

**G41B G42M G43F**

represents  
18 variants

⊗ ﻷ (U+0643) (U+0644) (U+0649)

⊗ ﻷ (U+06A9) (U+0644) (U+0649)

⊗ ﻷ (U+06A9) (U+0644) (U+06CC)

⊗ ﻷ (U+0643) (U+0644) (U+064A)

⊗ ﻷ (U+06A9) (U+0644) (U+064A)

⊗ ﻷ (U+06A9) (U+0644) (U+06CD)

⊗ ﻷ (U+06A9) (U+0644) (U+06D2)

⊗ ...etc ,

# Tools and solutions: Master Key Algorithm

- Exponential number of variants!!!

Label	Approximately # of variants
اتصال	300
اتصالات	6,000
الاتصالات	60,000
هيئة-الاتصالات	2,879,999
هيئة-الاتصالات-وتقنية-المعلومات	82,944,000,000

# Tools and solutions: Filters (language based)

- **Goal:**
  - To reduce the huge size of allocate-able variants by intelligently identify and displaying only the desired variants
- **How?**
  - Linguistically we study words in the Arabic language to find some rules to help identifying desired variants:
    - We used N-grams model to statically study the repetitive patters in Arabic words
      - An example of 2-gram for the word “cars”: “c”, “ca”, “ar”, “rs”, “s”
      - We studied 2, 3 and 4-grams for more than 7 million non-repetitive words in the Arabic language
      - Source: Books, Newspapers, Refereed Academic Journals.. Etc. (**KACST Arabic Corpus**)
    - We studied high-frequency patterns and then built some rules/filters based on them: (\*أَلْ, \*أَلْ, \*أَلْ, ... etc.)
  - We developed later a ranking system to arrange allocate-able variants based on weight given by each rule.
  - We have confirmed our findings with linguists and researchers.

# Tools and solutions: Filters (language based)

- Sample of our variant rules ( 21+ rules):
  - AlefMadaEnd
    - Input: خطأ-ظماً
    - Filtered out: خطأ-ظما، خطأ-ظما، خطأ-ظما ..etc
  - AlefHamzaDownEnd
    - Input: خطأ-ظما
    - Filtered out: خطأ-ظما، خطأ-ظما، خطأ-ظما ..etc
  - Alf-Altareef:
    - Input: القرآن
    - Filtered out: القرآن، القرآن، القرآن
  - Alef-letter-Alef
    - Input: رأيات
    - Filtered out: رأيات، رأيات، رأيات
  - .. etc.

## Note

Filtered out variants are still  
can be allocated manually  
after some verification

label: مكة المكرمة

Filtered Variants

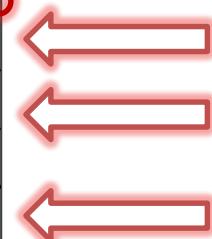
All Variants

## Results:

**Master Key:** G43B G41M G18F G14I G42B G43M G41M G26F G43B G18F

## Statistics Summary:

Total Variants	3239
I. Must be Allocated Variants (International Reachability )	2
II. Desired Variants	4
III. Not desired Variants	28
IV. Blocked Variants	3205



## I. Input:

LANGUAGE	UNICODE	LABEL
Arabic	(U+0645) (U+0643) (U+0629) (U+002D) (U+0627) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0629)	مكة المكرمة

## II. Must be Allocated Variants (2):

LANGUAGE	UNICODE	TYPO/EXACT	LABEL
Persian, Malay, Pashto	(U+0645) (U+06A9) (U+0629) (U+002D) (U+0627) (U+0644) (U+0645) (U+06A9) (U+0631) (U+0645) (U+0629)	E	مکة-المكرمة
Urdu	(U+0645) (U+06A9) (U+06C3) (U+002D) (U+0627) (U+0644) (U+0645) (U+06A9) (U+0631) (U+0645) (U+06C3)	T	مکة-المكرمة

### III. Desired Variants (4):

RULES	WEIGHT	LANGUAGE	UNICODE	TYPO/EXACT	LABEL
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	4	Arabic	(U+0645) (U+0643) (U+0647) (U+002D) (U+0627) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0647)	T	مکہ-المرکمة
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	2	Persian, Malay, Pashto	(U+0645) (U+06A9) (U+0647) (U+002D) (U+0627) (U+0644) (U+0645) (U+06A9) (U+0631) (U+0645) (U+0647)	T	مکہ-المرکمة
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	0	Arabic	(U+0645) (U+0643) (U+0629) (U+002D) (U+0627) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0647)	T	مکہ-المرکمة
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	0	Arabic	(U+0645) (U+0643) (U+0647) (U+002D) (U+0627) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0629)	T	مکہ-المرکمة

#### IV. Not Desired Variants (28)

RULES	WEIGHT	LANGUAGE	UNICODE	TYPO/EXACT	LABEL
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	-98	Arabic	(U+0645) (U+0643) (U+0629) (U+002D) (U+0622) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0629)	T	مكـةـ المـكرـةـ
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	-100	Arabic	(U+0645) (U+0643) (U+0647) (U+002D) (U+0622) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0629)	T	مـكـةـ المـكرـةـ
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	-98	Arabic	(U+0645) (U+0643) (U+0629) (U+002D) (U+0623) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0629)	T	مـكـةـ المـكرـةـ
<code>AllocatableInInputLanguage,</code> <code>SameLanguage, AlAtarif,</code> <code>Amlat, Sync, RuleOne,</code> <code>RuleTwo, RuleThree,</code> <code>RuleFour, RuleFive, RuleSix,</code> <code>RuleSeven, RuleEight,</code> <code>RuleNine, RuleTen</code>	-100	Arabic	(U+0645) (U+0643) (U+0647) (U+002D) (U+0623) (U+0644) (U+0645) (U+0643) (U+0631) (U+0645) (U+0629)	T	مـكـةـ المـكرـةـ
	22	Arabic	(U+0645) (U+0643)	T	مـكـةـ المـكرـةـ

# 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: international reachability

- For reachability purposes, variants should be addressed to be activated automatically by the registry, so that:
  - A registered domain name is accessed regardless of the input devices (language table) being used by the navigator users.
  - For example:
    - A user registered the domain “مكة” (all characters from the Arabic language)
    - if another user try to reach that domain name from an Internet café in Pakistan he/she will type “مڪّه” (all characters from the Urdu language)
    - If the “must-be-allocated” variants were not allocated, delegated and hosted then the domain name will not be reachable.



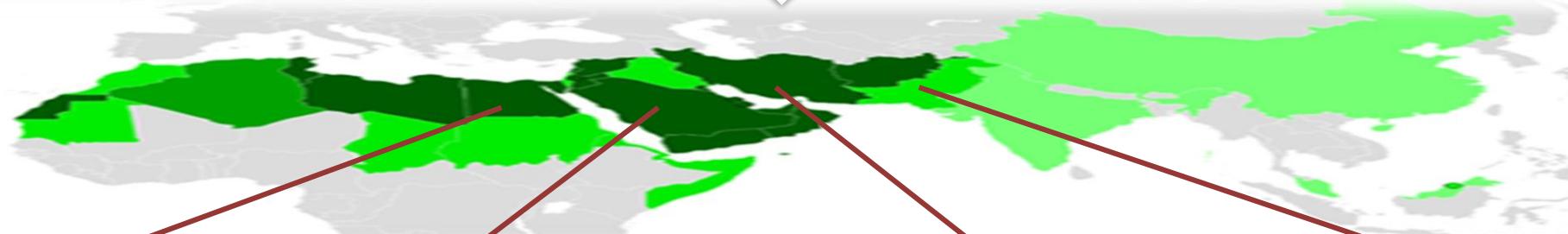
LANGUAGE	UNICODE	LABEL
Arabic	(U+0645) (U+0643) (U+0629)	مكة
Persian, Malay, Pashto	(U+0645) <b>(U+06A9)</b> (U+0629)	مڪّه
Urdu	(U+0645) <b>(U+06A9)</b> <b>(U+06C3)</b>	مڪّه



Hence, reachability issue (based on input devices used by other language communities) should be carefully considered when defining variants (by language communities).

# SaudiNIC's VMS: Registrant will use his/her keyboard

مكة



مكة
U+0645
U+0643
U+0629

مكة
U+0645
U+0643
U+0647

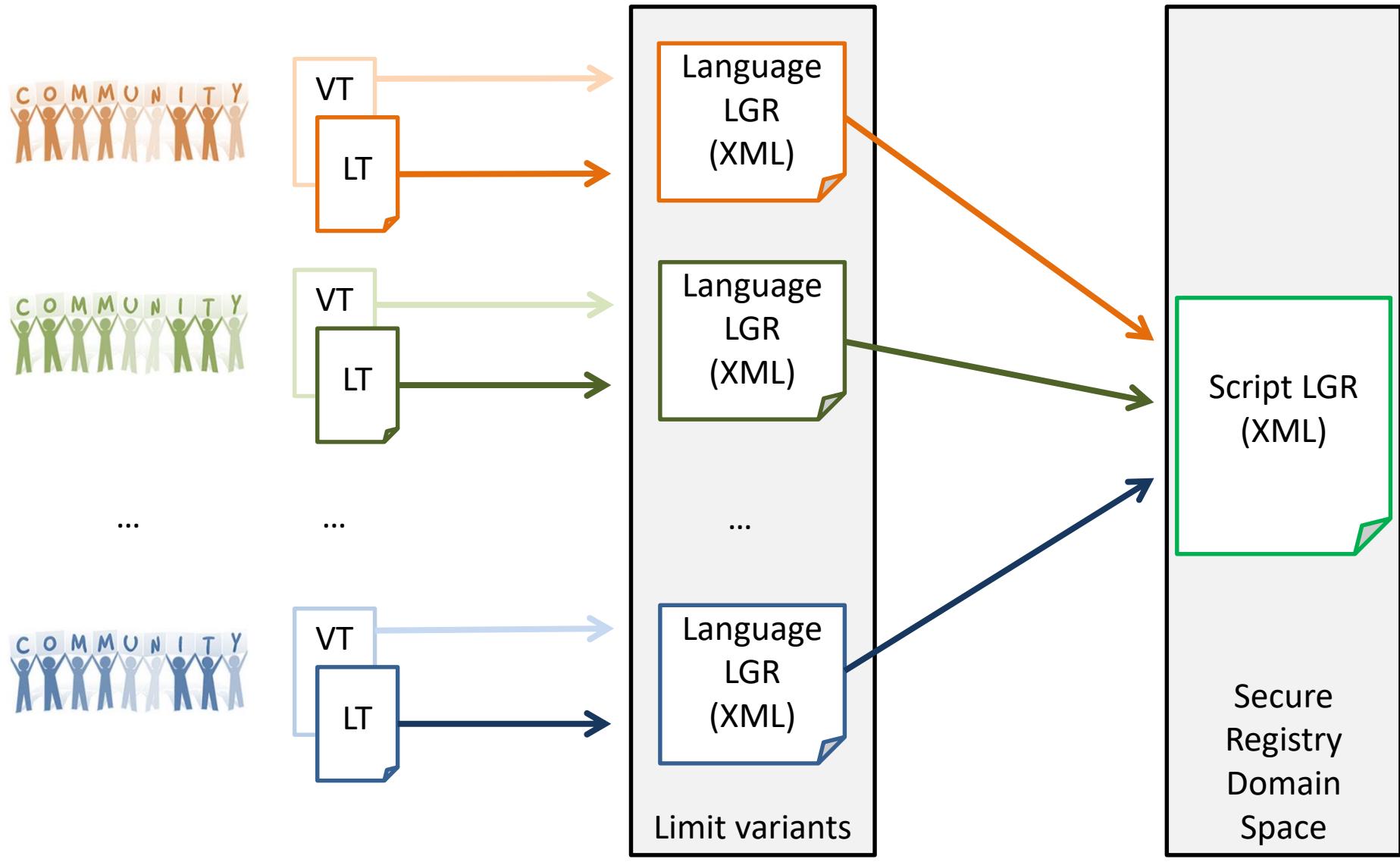
مكة
U+0645
U+06A9
U+0647

مكة
U+0645
U+06A9
U+06C1

# SaudiNIC's VMS: blocking quality??

IDN	Total Variants	Allocatable	Blocked	Blocked due to Language Mixing
مكة-المكرمة	3239	34	3205	3181 (99.25%)
القرآن-الكرييم	11999	111	11888	11836 (99.56%)
هيئة-الإعلام	47999	81	47918	47764 (99.68%)
كهف-الياسمين	28799	65	28734	28680 (99.81%)
كهف-اكيا	21599	47	21552	21534 (99.92%)

# SaudiNIC's VMS: Language LGR and Script LGR



# SaudiNIC's VMS: Easy interface for registrants

The screenshot shows the Variant Management Form for the domain `مكة.السعودية`. The interface is in Arabic and includes the following sections:

- Domain information:** Shows the domain name `مكة.السعودية`.
- Must be allocated:** Lists three variants:
  - Variant 1: `مكة` (Arabic) -
  - Variant 2: `مكة` (Persian, Malay, Pashto) -
  - Variant 3: `مكة` (Urdu) -
- Desired Variants:** Lists two desired variants:
  - `مكة` -
  - `مكة` -
- Other Variants:** A text input field for Variant 1 with a minus and plus button.

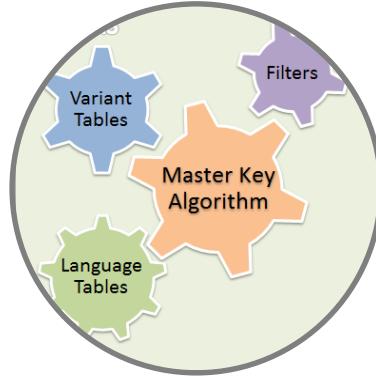
A sidebar on the right contains a vertical scrollable list of navigation items in Arabic, such as "معلومات عن", "مثال لتشبيهات", "اسم النطاق المس", "المثال الأول", "المثال الثاني", "المثال الثالث", "المثال الرابع", "قائمة التشبيهات", "حذف حمير", "الـ", "الـ", and "الـ".

# 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)

SaudiNIC's Major Efforts

# 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

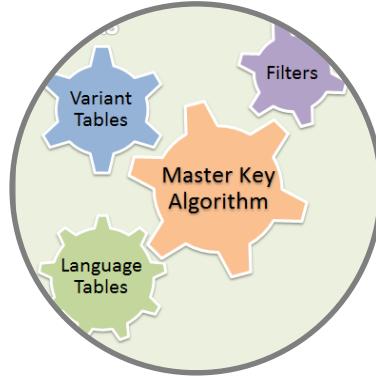
- IDN Assessment Report

# SaudiNIC's Major Efforts



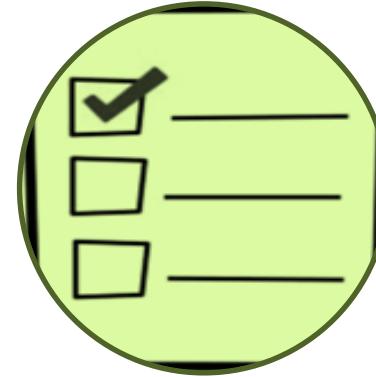
## Arabic IDN pilot projects

- 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)

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

The screenshot shows a Gmail inbox interface with Arabic language support. The top navigation bar includes tabs for 'rae...' ( التجربة للنص العربي ) and 'بريد رسيل العربي :: الوارد' ( Raseel Arabic Mail :: Incoming ). The search bar contains the text 'تجربة للنص العربي'. The inbox header includes standard controls like back, forward, search, and a compose button.

The inbox list displays two messages:

- تجربة للنص العربي** (rae...@rseel.com.sa) - 1:03 PM (1 hour ago) ☆  
Raed Al-Fayez <raed.alfayez@gmail.com> 1:04 PM (1 hour ago) ☆  
to رائد  
تمام التمام يا باتنا
- 2016-02-29 13:03 GMT+03:00 <rae...@rseel.com.sa>  
تجربة للنص العربي  
میروووووووک

The bottom of the inbox shows a message placeholder: 'Click here to Reply or Forward'.

On the left sidebar, there are links for 'Inbox', 'Starred', 'Important', 'Sent Mail', and 'Drafts (1)'. The user profile 'Raed' is also listed.

# Raseel: An Arabic Email System

- Almost **5 years** since the EAI RFCs were published and until now there are **almost no support** (or very limited) in:
  - Email servers (SMTP, IMAP, POP),
  - Email providers (Gmail, Hotmail, Yahoo)
  - Emails clients (Webmail, Application)
- Need to have a **protection** mechanism for the **user part** of the emails addresses (similar to IDN variants)

بريد@رسيل.ال سعودية	بريد@رسيل.ال سعودية
Farsi Yeh (U+06CC)	Arabic Yeh(U+064A)

- 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?**

# Variants enablement must be done in every level

Register and enable variants:

مكة

مكة

مكة

Registry

Configure DNS & add need RR  
(e.g. NS & A & CNAME) for:

xn--ogb5cf

xn--ogb9c4p

xn--hhb4rwc

DNS Hosting

Configure Email account and email aliases:

رائد@مكة

رائد@مكة

رائد@مكة

Email Services

Configure web-server and account and aliases:

```
<VirtualHost 10.10.10.10>
```

```
DocumentRoot "/makkah"
```

```
ServerName xn--ogb5cf
```

```
ServerAlias xn--ogb9c4p
```

```
ServerAlias xn--hhb4rwc
```

```
</VirtualHost>
```

Web Hosting

IDN + Variants

# Gift

- Published “**SaudiNIC’s Best Practices in Supporting and Managing Arabic Domain Names**”
  - [http://www.nic.sa/docs/SaudiNIC\\_ADNP.pdf](http://www.nic.sa/docs/SaudiNIC_ADNP.pdf)

*Thank you*

شكراً

للمزيد من المعلومات يمكنكم زيارة:

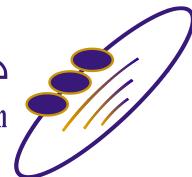
For more information you can visit:



سجل.السعودية

[nic.sa](http://nic.sa)

هيئة الاتصالات وتقنية المعلومات  
Communications and Information Technology Commission



هيئة-الاتصالات.السعودية

[citc.gov.sa](http://citc.gov.sa)