هيئة الاتصالات وتقنية المعلومات
Communications and Information Technology Commission

Proposed Solution for

Writing Domain Names in Different Arabic Script Based Languages

TF-AIDN, June / 2014

Presented by …

AbdulRahman I. Al-Ghadir
Researcher in SaudiNIC

Content

- What we have done so far?
- Problem definition
- Proposed solution
- Requirements for Languages
- Requirements for Registries
- GVT as Open Source Product
- Conclusion

What we have done so far?

- **The work was done based on the following methodology:**

1. Identifying problems & areas of contributions
2. Participating and initiating interest groups & task forces
3. Conducting web surveys
4. Publishing reports & papers
5. Meeting linguists (face to face)
6. Disseminating information to public
7. Testing and building local experiences

What we have done so far?
Linguistic Recommendations

Tashkeel (Diacritics)	*Tashkeel should not be allowed.* *However, if there is a need to allowed users to entered it as part of a domain name then it should be stripped off by nameprep*
Kasheeda	Kasheeda should be disallowed
Character folding: Teh Marbuta + Heh different forms of Hamzah Alif Maqsura+YaNumbers	Folding should not be allowed
Numbers (numerical digits)	If it is technically possible, it is preferred to support both (Latin and Arabic) sets with folding to one set. Otherwise, Latin set is sufficient
Connecting Multiple Words	It is recommended that multiple words are separated by the character "-".
Mixing Latin and Arabic Characters	It is recommended that Arabic domain names be pure Arabic and they should not be mixed with other languages.
Special Characters (e.g., @, #, $, %, ...)	It is recommended that Arabic domain names should follow the standard with respect to the use of special characters.

What we have done so far?
ADNPP Members so far

- **Participated Countries:**
  - United Arab Emirates
  - Saudi Arabia
  - Qatar
  - Oman
  - Palestine
  - Egypt
  - Tunisia
  - Syria
  - Jordan
  - Morocco
  - Libya

United Arab Emirates
United Arab Emirates Network Information Center

Kingdom of Saudi Arabia
Saudi Network Information Center

State of Qatar
Internet Qatar

Sultanate of Oman
Oman Telecommunications Company

State Of Palestine
Ministry of Telecom and IT

Arab Republic Of Egypt
Ministry of Communications & Information Technology

Republic Of Tunisia
Tunisian Internet agency

Syrian Arab Republic
Tunisian Internet agency

Hashemite Kingdom of Jordan
National information technology center

Kingdom of Morocco
www.anrt.ma

Great Socialist People's Libyan Arab Jamahiriya
http://www.ltt.ly

.sa

What we have done so far?
Accepted Character Set Table

- Characters from Unicode Arabic Table (0600–06FF)

- 0621 (ء) Arabic Letter HAMZA
- 0622 (آ) Arabic Letter ALEF with MADDA above
- 0623 (أ) Arabic Letter ALEF with HAMZA above
- 0624 (ؤ) Arabic Letter WAW with HAMZA above
- 0625 (إ) Arabic Letter ALEF with HAMZA below
- 0626 (ئ) Arabic Letter YEH with HAMZA above
- 0627 (ا) Arabic Letter ALEF
- 0628 (ب) Arabic Letter BEH
- 0629 (ة) Arabic Letter TEH MARBUTA
- 062A (ت) Arabic Letter TEH
- 062B (ث) Arabic Letter THEH
- 062C (ج) Arabic Letter JEEM
- 062D (ح) Arabic Letter HAH
- 062E (خ) Arabic Letter KHAH
- 062F (د) Arabic Letter DAL
- 0630 (ذ) Arabic Letter THAL
- 0631 (ر) Arabic Letter REH
- 0632 (ز) Arabic Letter ZAIN
- 0633 (س) Arabic Letter SEEN
- 0634 (ش) Arabic Letter SHEEN
- 0635 (ص) Arabic Letter SAD
- 0636 (ض) Arabic Letter DAD
- 0637 (ط) Arabic Letter TAH
- 0638 (ظ) Arabic Letter ZAH
- 0639 (ع) Arabic Letter AIN
- 063A (غ) Arabic Letter GHAIN
- 0641 (ف) Arabic Letter FEH
- 0642 (ق) Arabic Letter QAF
- 0643 (ك) Arabic Letter KAF
- 0644 (ل) Arabic Letter LAM
- 0645 (م) Arabic Letter MEEM
- 0646 (ن) Arabic Letter NOON
- 0647 (هـ) Arabic Letter HEH
- 0648 (و) Arabic Letter WAW
- 0649 (ى) Arabic Letter ALEF MAKSURA
- 064A (ي) Arabic Letter YEH
- 0660 (0) Arabic-Indic Digit Zero
- 0661 (1) Arabic-Indic Digit One
- 0662 (2) Arabic-Indic Digit Two
- 0663 (3) Arabic-Indic Digit Three
- 0664 (4) Arabic-Indic Digit Four
- 0665 (5) Arabic-Indic Digit Five
- 0666 (6) Arabic-Indic Digit Six
- 0667 (7) Arabic-Indic Digit Seven
- 0668 (8) Arabic-Indic Digit Eight
- 0669 (9) Arabic-Indic Digit Nine

What we have done so far?
Accepted Character Set Table

- Characters from Unicode Basic Latin Table (0000–007F):

- 0030 (0) Digit Zero
- 0031 (1) Digit One
- 0032 (2) Digit Two
- 0033 (3) Digit Three
- 0034 (4) Digit Four
- 0035 (5) Digit Five
- 0036 (6) Digit Six
- 0037 (7) Digit Seven
- 0038 (8) Digit Eight
- 0039 (9) Digit Nine
- 002D (-) Hyphen-Minus
- 002E (.) Full Stop (Dot)

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	0600	0610		0630	0640	0650	0660	0670	0680	0690	06A0	06B0	06C0	06D0	06E0	06F0
1	0601	0611	0621	0631	0641	0651	0661	0671	0681	0691	06A1	06B1	06C1	06D1	06E1	06F1
2	0602	0612	0622	0632	0642	0652	0662	0672	0682	0692	06A2	06B2	06C2	06D2	06E2	06F2
3	0603	0613	0623	0633	0643	0653	0663	0673	0683	0693	06A3	06B3	06C3	06D3	06E3	06F3
4		0614	0624	0634	0644	0654	0664	0674	0684	0694	06A4	06B4	06C4	06D4	06E4	06F4
5		0615	0625	0635	0645	0655	0665	0675	0685	0695	06A5	06B5	06C5	06D5	06E5	06F5
6	0606	0616	0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6
7	0607	0617	0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7
8	0608	0618	0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8
9	0609	0619	0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9
A	060A	061A	062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA
B	060B	061B	062B	063B	064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB
C	060C		062C	063C	064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC
D	060D		062D	063D	064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD
E	060E	061E	062E	063E	064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE
F	060F	061F	062F	063F	064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF

▪Accepted characters for Arabic, Persian, Urdu, Pashto, Jawi

0	0600	0610		0630	0640	0650	0660	0670	0680	0690	06A0	06B0	06C0	06D0	06E0	06F0
1	0601	0611	0621	0631	0641	0651	0661	0671	0681	0691	06A1	06B1	06C1	06D1	06E1	06F1
2	0602	0612	0622	0632	0642	0652	0662	0672	0682	0692	06A2	06B2	06C2	06D2	06E2	06F2
3	0603	0613	0623	0633	0643	0653	0663	0673	0683	0693	06A3	06B3	06C3	06D3	06E3	06F3
4		0614	0624	0634	0644	0654	0664	0674	0684	0694	06A4	06B4	06C4	06D4	06E4	06F4
5		0615	0625	0635	0645	0655	0665	0675	0685	0695	06A5	06B5	06C5	06D5	06E5	06F5
6	0606	0616	0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6
7	0607	0617	0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7
8	0608	0618	0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8
9	0609	0619	0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9
A	060A	061A	062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA
B	060B	061B	062B	063B	064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB
C	060C		062C	063C	064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC
D	060D		062D	063D	064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD
E	060E	061E	062E	063E	064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE
F	060F	061F	062F	063F	064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF

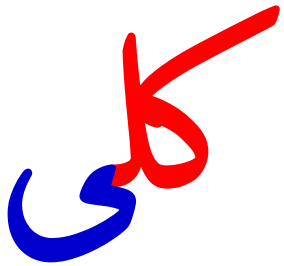
Arabic Script IDN - Major Issues

1. **Accepted/disallowed characters**
   - IDNA2008 table (Pvalid / Disallowed / ContextO)
   - Language tables
2. **Combining Marks**
3. **Non-spacing Marks (e.g. Diacritics)**
4. **Word/label separators (e.g. space, ZWNJ, ZWJ, hyphen)**
5. **Digits**
6. **Confusing of similar characters.**
7. **Bidirectional**

Confusing Similar Characters

What is the Problem ?

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



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

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

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	0600	0610		0630	0640	0650	0660	0670	0680	0690	06A0	06B0	06C0	06D0	06E0	06F0
1	0601	0611	0621	0631	0641	0651	0661	0671	0681	0691	06A1	06B1	06C1	06D1	06E1	06F1
2	0602	0612	0622	0632	0642	0652	0662	0672	0682	0692	06A2	06B2	06C2	06D2	06E2	06F2
3	0603	0613	0623	0633	0643	0653	0663	0673	0683	0693	06A3	06B3	06C3	06D3	06E3	06F3
4		0614	0624	0634	0644	0654	0664	0674	0684	0694	06A4	06B4	06C4	06D4	06E4	06F4
5		0615	0625	0635	0645	0655	0665	0675	0685	0695	06A5	06B5	06C5	06D5	06E5	06F5
6	0606	0616	0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6
7	0607	0617	0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7
8	0608	0618	0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8
9	0609	0619	0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9
A	060A	061A	062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA
B	060B	061B	062B	063B	064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB
C	060C		062C	063C	064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC
D	060D		062D	063D	064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD
E	060E	061E	062E	063E	064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE
F	060F	061F	062F	063F	064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF

Site # - Microsoft Internet Explorer provided by CITC

File Edit View Favorites Tools Help

Back Address http://كلى.com/ Go

Site #

1

Arabic Label (U-Label) : كلى

Unicode (U+) : U+06CC , U+0644 , U+06A9

ASCII Label (A-Label) : xn--ghb5rwd.com

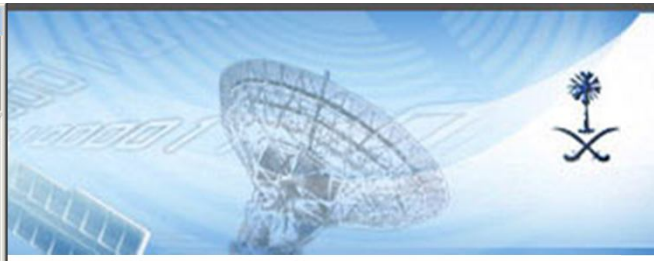
Now see where this dom

Why the domain name was registered :

1. To illustrate that an IDN script-based registration (particularly with Ara
2. To encourage the international communities in general and language co
solutions to these kind of problems.
3. To speed up the process for ccTLD IDN fast track where each IDN c
4. If new IDN-gTLDs will be open, then they should focus on languages :

For more information see this the Power Presentation that was delivered on t

Done



Site # - Microsoft Internet Explorer provided by CITC

File Edit View Favorites Tools Help

Back Address http://كلى.com/ Go

Site #

2

Arabic Label (U-Label) : كلى

Unicode (U+) : U+0643, U+0644, U+0649

ASCII Label (A-Label) : xn--fhbcp.com

Now see where this domain will go : كلى.com

Why the domain name was registered :

1. To illustrate that an IDN script-based registration (particularly with Arabic Script) is a good ground for phishing.
2. To encourage the international communities in general and language communities (that use Arabic script) in particular to work together in finding solutions to these kind of problems.

Confusing Similar Characters

What is the Problem ?

- **Security issues (stability, trust,…) e.g. phishing**
  - They should be addresses at language level first
- **Not all Arabic-script languages are ready:**
  - Not widely / commonly used
  - Language community are not ready
- **Hard to make decisions on behave of other language communities**
- **Pressure to start with ready languages**
- **… and yet has to provide a simple and transparent registration services**

Confusing Similar Characters
What is the Problem ?

- **If some one want to register "هدهد"**
  - Assuming there are 4 variants for "ه"
  - There will be 16 possible ways to write it
    - Only **4** of them may confuse the end users (25%)
  - So why we block / bundle 75% not confusing domains?

هدهد	هدهد	ہدهد	ەدهد
هدهد	هدہد	هدەد	هدهد
هدہد	هدەد	ہدهد	ہدہد
ہدەد	ەدهد	ەدہد	ەدەد

Invalid option (position)

Can be enabled

Mixed of languages

Confusing Similar Characters

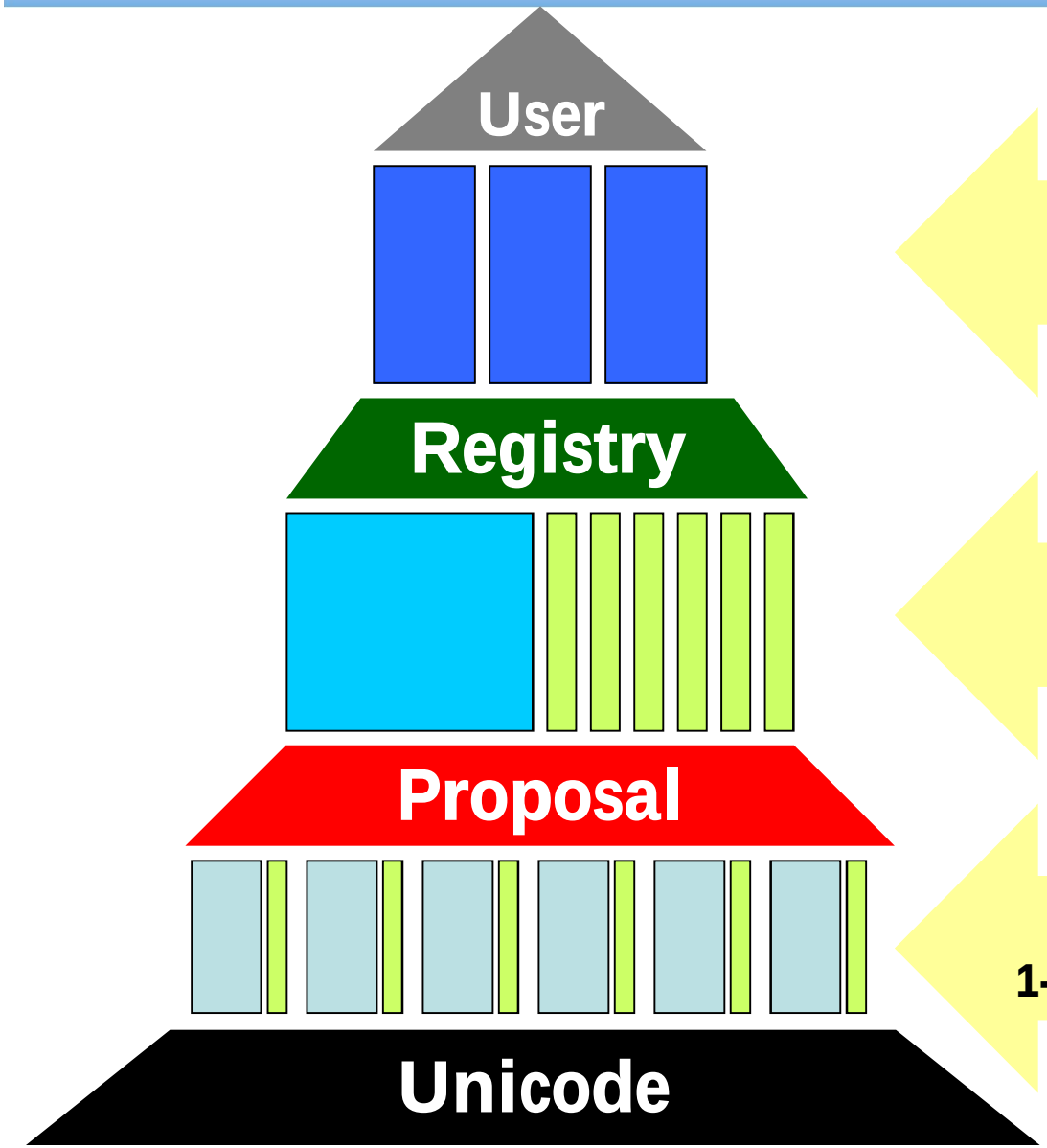
What is the Problem ?

- **How to know if a domain name have an already registered variant?**
  - Assuming we have define variants based on the inputs of current 3 ready languages;
  - There are **16,384** possible variants to write the domain "هيئة-الاتصالات-وتقنية-المعلومات"
  - Options:
    - Are we going to store / bundle all of them?
      - Dose the Whois function supposed to search in all registered domains & all possible variants for each one of them?
      - Not possible: Time & Resource consuming!!!!
    - Need alternative way to handle this issue (Master-Key)

Characteristics of A Desired Solution

- **Based on standardized (or agreeable) policies and procedures**
  - documented on RFC-like or Best-Practice documents
- **Extendable to allow for adding new languages as they become ready**
- **Easy and fast to be deploy by any registry**
- **Work for both ccTLDs and gTLDs**
- **Simple and Transparent for end users**
  - Do not annoy / confuse end users with technical / special
  - Regular users should be able to register whatever he can type using his keyboard

General overview



Registry functions
(Whois + Register + Activate)

Group variant table +
Language tables

For each language:
1-Language table & 1- variant tables

Language Requirements

- **Language Table (LT)**
  - A set of codes (Base characters) that defines a certain language domain that used by Registry.
  - LT can have Alphabetical, Numbers and Separators (Hyphens, Dots)

- **Variant Table (VT)**
  - Variant Table is a table that record all relations of the LT characters with other characters across the script.
  - Each relation is defined depending on its similarity either Exact or Typo.

Language Requirements

- **Variant Table (VT) cont.**
  - Exact similarity refers to identically look between base character other character (e.g. exact match / mirror image).
  - Typo similarity refers to almost look between base character other character (e.g. typo / style match).
  - Note: A variant table will consist of a list of records, each record contains the following information:
    - Base character (from LT),
    - List of other characters that have similarity with base character (from across the script),
    - A set of positions of similarity [**B**eginning , **M**edial, **F**inal, **Is**olated],
    - Relation type (**E**xact, **T**ypo).

Language Requirements
Examples of variants

- **Example Exact Variant match**

	I	F	M	B
ف 0641 FEH	ف	ـفـ	ـفـ	ـف
ڤ 06A7 QAF WITH DOT ABOVE	ڤ	ـڤـ	ـڤـ	ـڤ

- **Example for Typo Variant match**

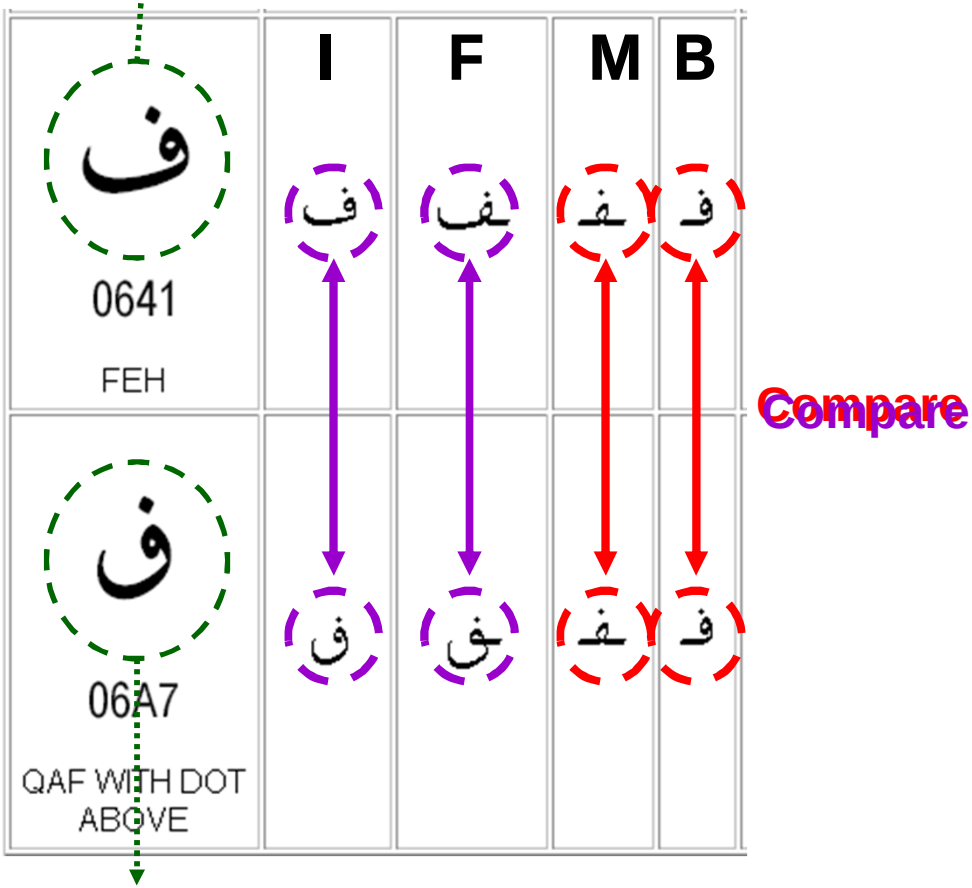
	I	F	M	B
ف 0641 FEH	ف	ـفـ	ـفـ	ـف
ڤ 06A7 QAF WITH DOT ABOVE	ڤ	ـڤـ	ـڤـ	ـڤ

Language Requirements
How to build a variant table?

- **Steps (done for each base character in LT):**
  1. List all possible shapes for the basic character
  2. Search for all its variants from the rest of the Arabic script
  3. Then compare the basic character with its variants in all possible positions.
  4. Find all similarity position(s).
  5. Record the similarity (type & position)

Language Requirements
Example: Position of similarity

A base character from LT



A variant character from script

Exact

Find similarity positions

I F M B

ف = ف (BM :E)

0641 06A7

Typo

Find similarity position

I F M B

ف = ف (FI :T)

0641 06A7

Language Requirements
Example: Variant Table

```
0636;
0637;
0638;
0639;
063A;
0641; 06A7(FI:T), 06A7(BM:E)
0642;
0643; 06A9(FI:T), 06A9(BM:E), 06AA(BMFI:T)
0644;
0645;
0646; 06BA(BM:E)
0647; 06BE(M:E), 06BE(BFI:T), 06C1(I:E), 06C1(MF:T), 06D5(FI:E)
0648;
0649; 06CD(FI:T), 06D2(FI:T)
064A; 067B(BMFI:T), 06D0(BMFI:T)
0660; 0030(BMFI:T)
```

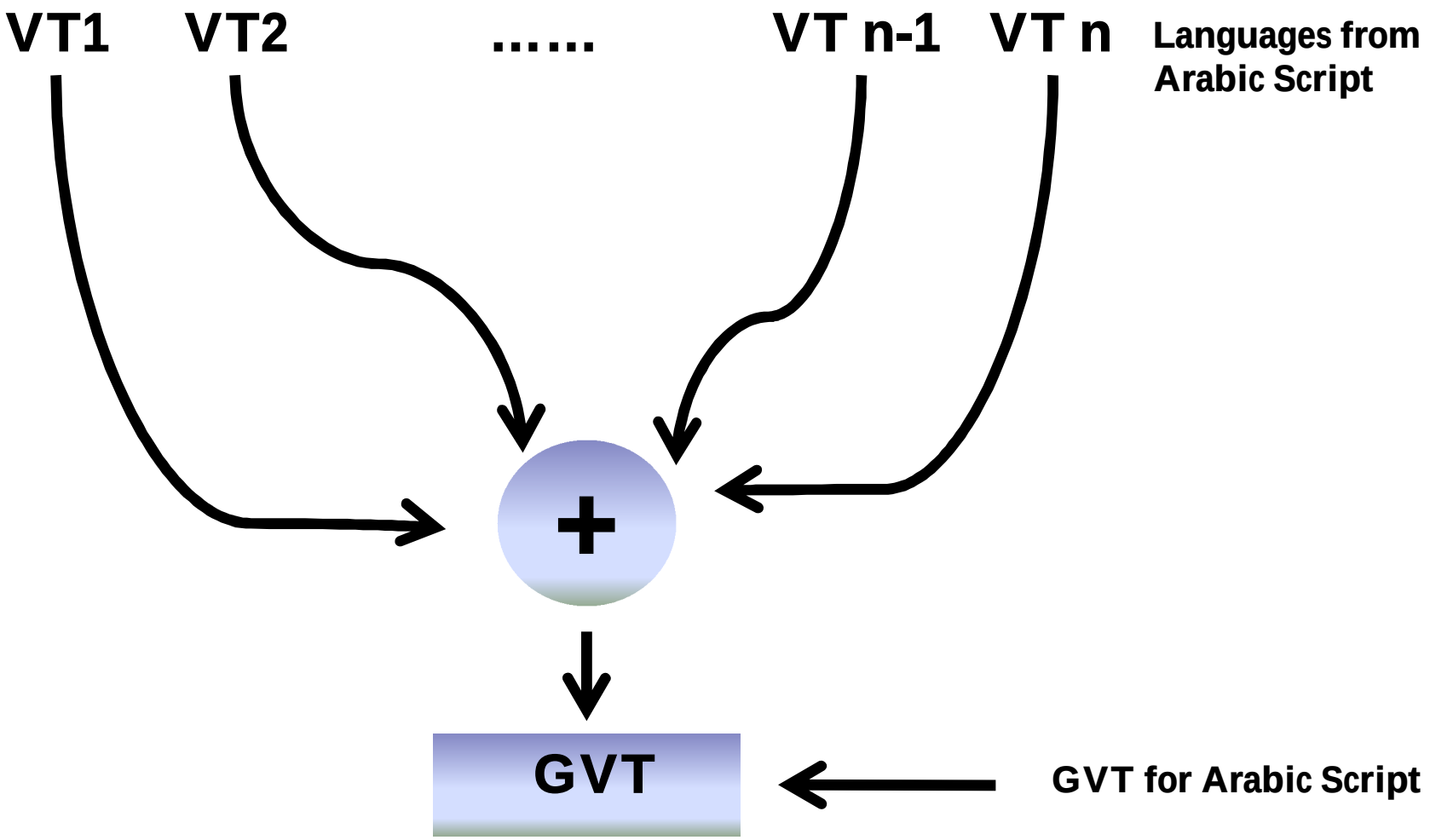
```
0641; 06A7(FI:T), 06A7(BM:E)
```

Registry Requirements

- **Language Table (one for every supported language)**
  - Users can only register domains using base characters from only one language table.

- **Group Variant Table (GVT):**
  - Generated from variant tables.
    - It combines all VTs into one table that group all base characters with all relations across script.
    - Each variant list will be assigned to a unique group key (master key) that identify that group and will be used for generating the Master Key.

Registry Requirements
Example: Group variant table



Registry Requirements
Example: Group variant table

GVT keys

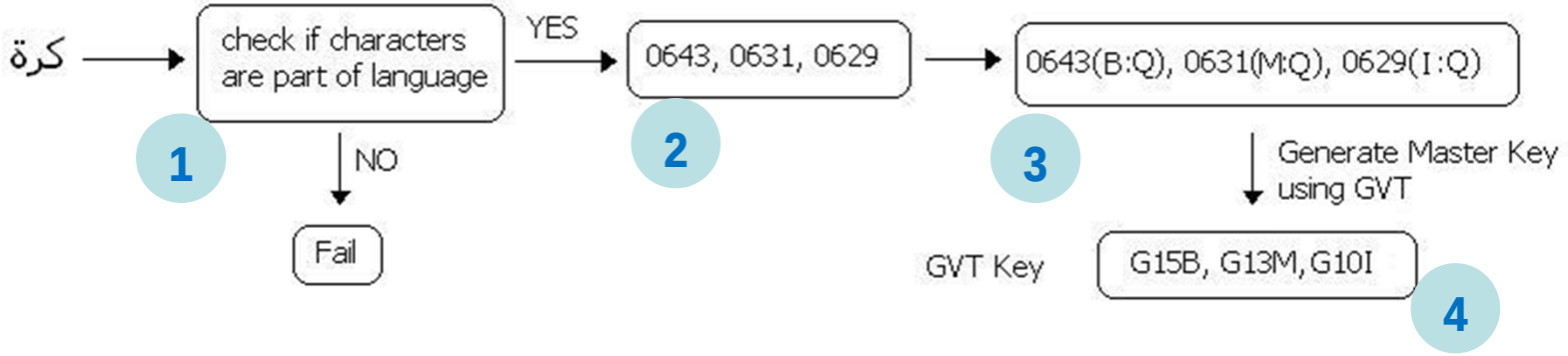
Relations for variant characters

```
G41M; 063A(Q)
G41F; 063A(Q)
G41I; 063A(Q)
G42B; 0641(Q), 06A7(Q), 06A7&0641(E)
G42M; 0641(Q), 06A7(Q), 06A7&0641(E)
G42F; 0641(Q), 06A7(Q), 06A7&0641(T)
G42I; 0641(Q), 06A7(Q), 06A7&0641(T)
G43B; 0642(Q)
G43M; 0642(Q)
```

Keys are used for Querying GVT

Registry Requirements
Generating Master Key

- **Generating Master Key:**
  - 1) Check if input string follows certain language (using LT).
  - 2) Generate UNICODE code for that input.
  - 3) Find the position for each character depending on language properties (UNICODE Standard).
  - 4) Query (generate) Master key by taking every code from 3) and do simple lookup in GVT.



Registry Requirements
Finding Exact Strings

- **Find all Exact strings using Master Key for activation purpose.**

G15B, G13M, G13I

Use Master key to reversal lookup and find all possible characters (Query relation)

كرة	كرة	كرة
Arabic	**Persian**	**Urdu**
U+0643, U+0631, U+0629	U+06A9, U+0631, U+0629	U+06A9, U+0631, U+06C3

Registry Requirements
Registry-Registrant interface

- **Lookup process (whois)**
  - Check domain syntax under any supported language using LTs.
  - Check if the same domain is available or not.
  - If it is found return the unavailable / whois-information; otherwise continue
  - Get the master- key for the domain (based on GVT)
  - Check if the master- key was registered before or not
  - If master- key is found return unavailable / whois-information; otherwise return domain is available

- **Registration process:**
  - Registrant should select one of these languages and a domain (U-Label)
  - Registry should accept inputs based on the selected language table
  - If domain name can be registered (available based on Lookup process) then register the domain

- **Activation process (enable exact variants)**
  - Original Registrant can activate any exact variant from the registered domain's Master Key.
  - List possible Exact variants that can be typed using one of the LT without intermingling between them
  - Activate one / many of Exact variants (if not activated before)

Registry Requirements
Support new languages

- **Adding new Language (GVT) to existence GVT done in three steps:**
  - 1) Scan GVT keys in new GVT and check if keys with Q exist in any key in old GVT if so take variant list of that key (from new GVT) and add it with variant list of old GVT.
  - 2) Add the rest of key of new GVT at the end of old GVT keys.
  - 3) check new GVT if the keys with Q appear in different GVT keys or not.

GVT(Old) + GVT(New languages)

Keys appear in different GVT keys → **Merging fails!**

Keys don't appear in different GVT keys → **Merging successes!**

Cure: regenerate old GVT using existence VTs with new VT.

Then regenerate all old Master keys using new GVT!

Done!

GVT as Open Source Product
GVT in real life

- **So far GVT algorithm became as Open source product.**
- **The product demonstrates the algorithm as piece of code.**
- **What can be done with it:**
  - Use it as resource to understand the algorithm.
  - Enhance it (if needed) and use it as tool.
  - A live demo for the algorithm!

Conclusion

- **We tried to have a prototype that fulfill the concepts of script based registry that is:**
  - Optimized, Simple, Transparent, Automated

- **Next steps:**
  - Finalize the Language tables & variant tables for the Arabic Language.
  - RFC or best-practice document.
  - ICANN should delegate variants at TLD level
    - E.g. Arabic => كويت => U+0643 U+0648 U+064A U+062A

      Persian => كويت => U+06A9 U+0648 U+06CC U+062A

Thank you !

- **Developing Team**
  - Abdulaziz Al-Zoman
  - Raed Al-Fayez
  - AbdulRahman Al-Ghadir

Thank you

شكرا

U+0634 U+06A9 U+0631 U+0627

شكرا

U+0634 U+0643 U+0631 U+0627

G35B G44M G32F G22I

- **Demo**

- **Backup & old slides**

Registry Requirements
Example: Group variant table

- **How to build Group variant table (GVT):**

1. Merge all variant tables (TVT, EVT) for all languages together in one Combined Variant Table (CVT)
   - Use "OR" operation in conflicts on position of similarity
   - Set a flag to show if it is an Exact or Type variant
2. For each code point in each LT build 4 groups based on its possible shapes (B,M,E,S)
3. Use the CVT to find variant code points that share the similarity in the same position and combine their initial groups together

Registry Requirements
Example: Group variant table

- **What dose each row in GVT store?**
  - Group –ID (example: G12B)
    - Each codepoint from any supported language can be found in up to 4 groups (based on it possible shapes B,M,E,S)
  - Members of the group
    - List code point that have any similarity between them in the same position (B,M,E,S)

GVT

...

G27B; 0641(E),06A7(E)

G27M; 0641(E), 06A7(E)

G27E; 0641(E), 06A7(T)

G27S; 0641(E), 06A7(T)

...

CVT

0641; 06A7(3)

0641; 06A7(12)

Introduction
Arabic Script IDN Major Issues

- **Acceptable/disallowed characters**
  - IDNA200x table (Pvalid / Disallowed / ContextO)
  - Language tables
- **Non-spacing Marks**
  - Subtending Marks (U+0600 – U+0603)
  - Honorifics (U+0610 – U+0614)
  - Koranic annotation signs (U+0615 -
  - Points (U+064B – U+0652, U+0670)
  - Combining Maddaa and Ha
  - Other combining Marks (U+06
- **Confusing similar characters (e.g.**
- **World/label separators (space, Z**
- **Bidirectional**
- **They are addressed at different levels**
  - IDNA protocol level
  - Registry level
  - Application level

استبدال
بالجدول

Introduction
About Arabic script

- **The 2nd most widely used alphabetic writing system in the world (used by more than 43 countries)**
  - more than one billion potential users could be conce ed in using Arabic script domain name
- **Used by many languages su s: A c, ian, rdu, Turkish, Kurdish, Pashto, Swa**
  - that may add or change ch o not appear in Arabic phonology.
  - A new character usually create ne of an existing Arabic chara , for e e dots.
  - These additions have meaning to th ut o the original language.
  - Therefore, many characters would easily confus d with some other characters from other languages

حذف

Introduction
Unicode Arabic Block (5.1)

- Subtending marks
- Radix symbols
- Letterlike symbols
- Punctuation
- Currency Signs
- Poetic marks
- Honorifics
- Koranic annotation signs
- Based on ISO 8859-6
- Addition for early Persian and Azerbaijani
- Points from ISO 8859-6
- Combining Madah and Hamza
- Other combining marks
- Arabic-Indic digits
- Archaic letters
- Point
- Exte l A letter
- s for
- nts
- s for rwa
- Addi ns for urushaski
- Addi ons for early Persian

حذف

298 Code points

Introduction
Who is ready?

- **Some language communities are somehow ready (alphabetical order):**
  - Arabic
  - Jawi
  - Pashto
  - Persian
  - Urdu
  - . . .

حذف

Why we need this?
Transparent for the End users

- تجميع لشرح مميزات الحل

- **Do not annoy/confuse end users with technical/special terms:**
  - Type & Exact Variants
  - Register & Bundle & Activate

- **Regular users should be able to register what he can type using his keyboard**
  - Advance users may seek for other options / solutions to fulfill their needs.

Why we need this?

- تجميع لشرح مميزات الحل

- **Optimized solution**
  - Only block domains that are **really** confusing to the end users
- **Simplify Registry operations**
  - Fast and accurate Whois service
  - Simple registration & activation services
- **Transparent to end users (registrant and navigator)**
  - Keep it simple & similar to what they used to
- **Automate expandability**
  - No need to meet and discuss the same issues again when a new language is ready!

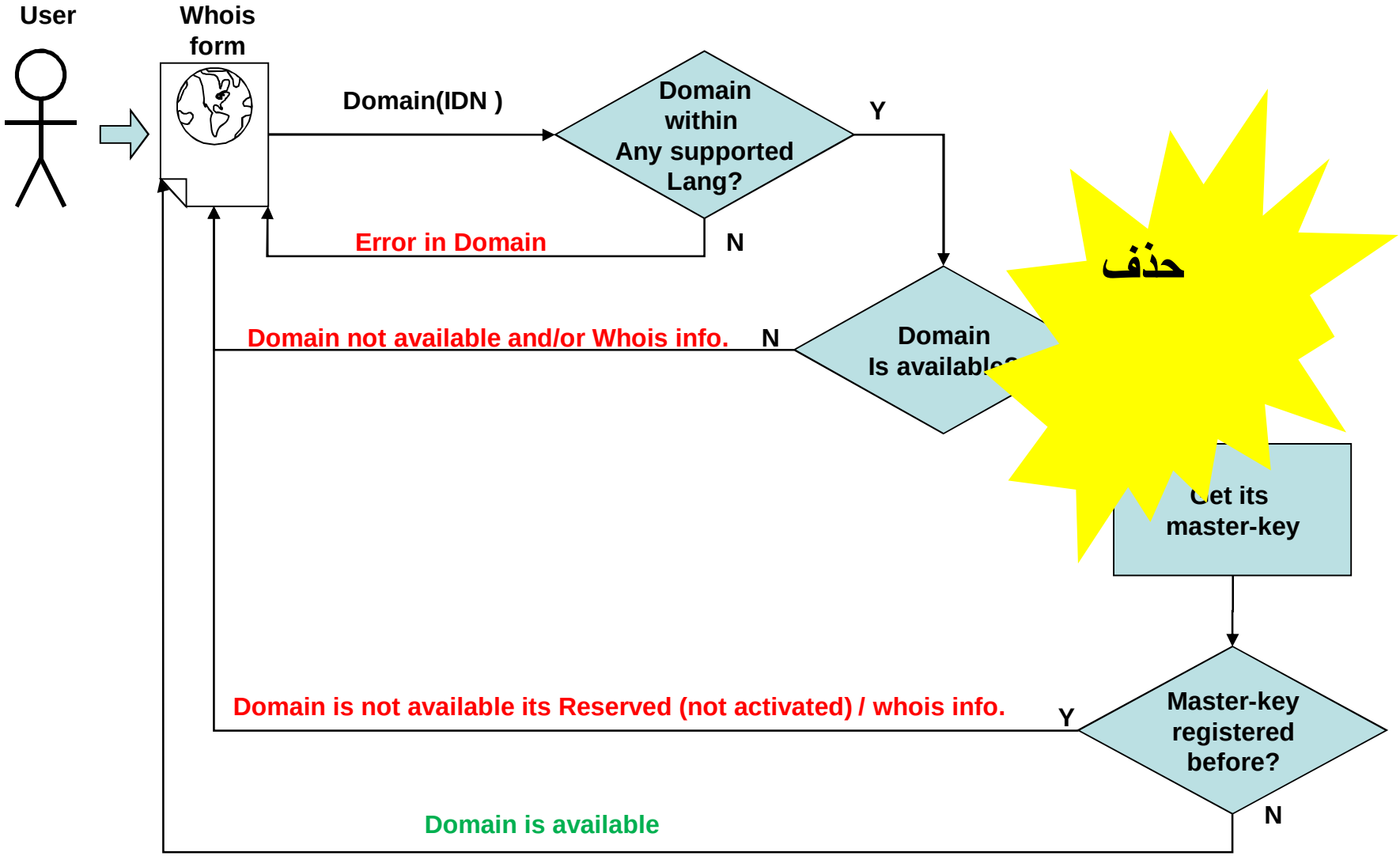
Why we need this?
Automate expandability

- تجميع لشرح مميزات الحل

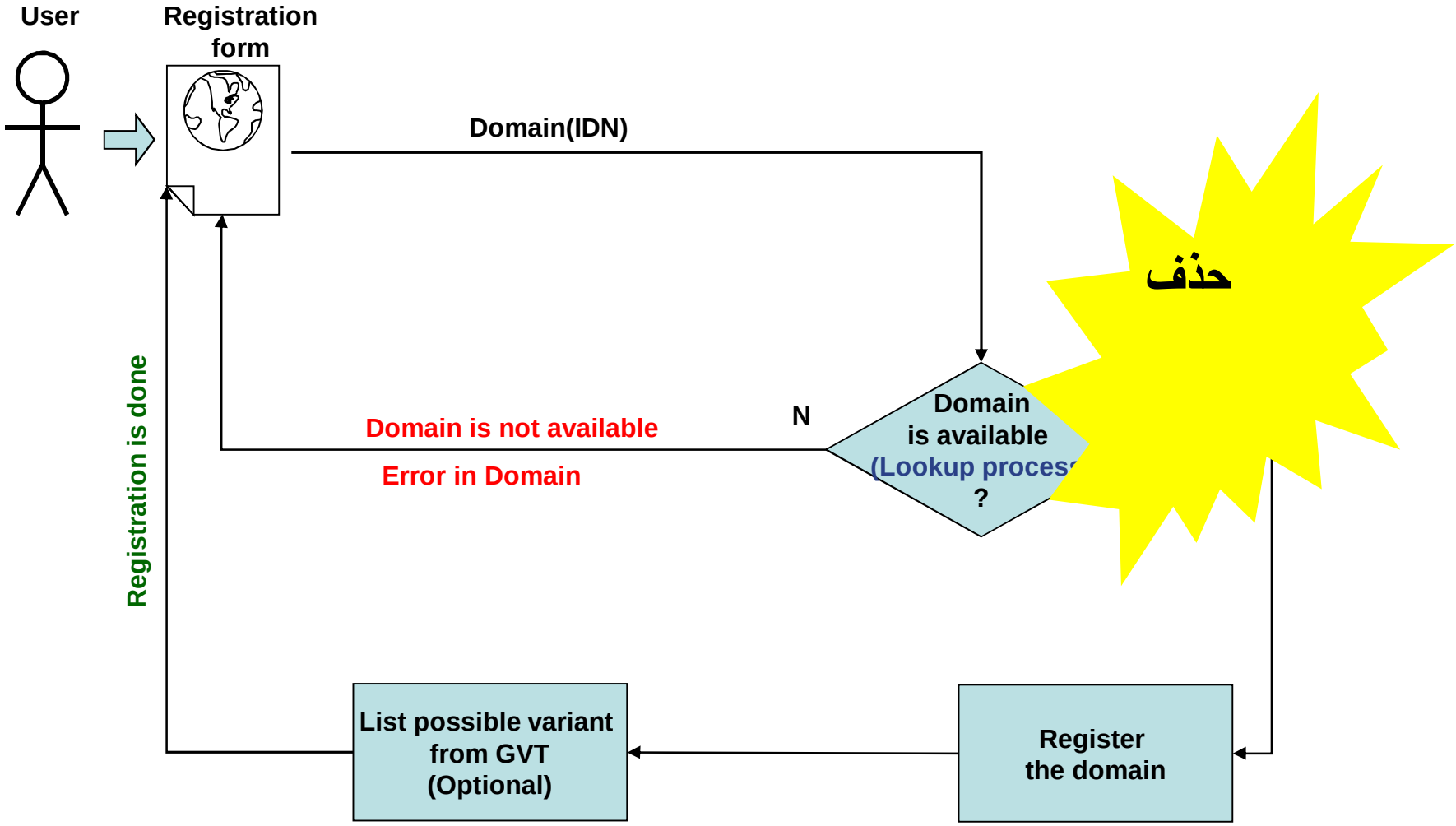
- **Extendibility against**
  - Adding new languages
  - Updating existing language tables & variants lists
  - Handling Unicode updates (e.g. Unicode 5.2)

- **Define set of rules and algorithms that Registries can use to operate TLD using Arabic Script**
  - Rules for storing some structures for backward compatibly

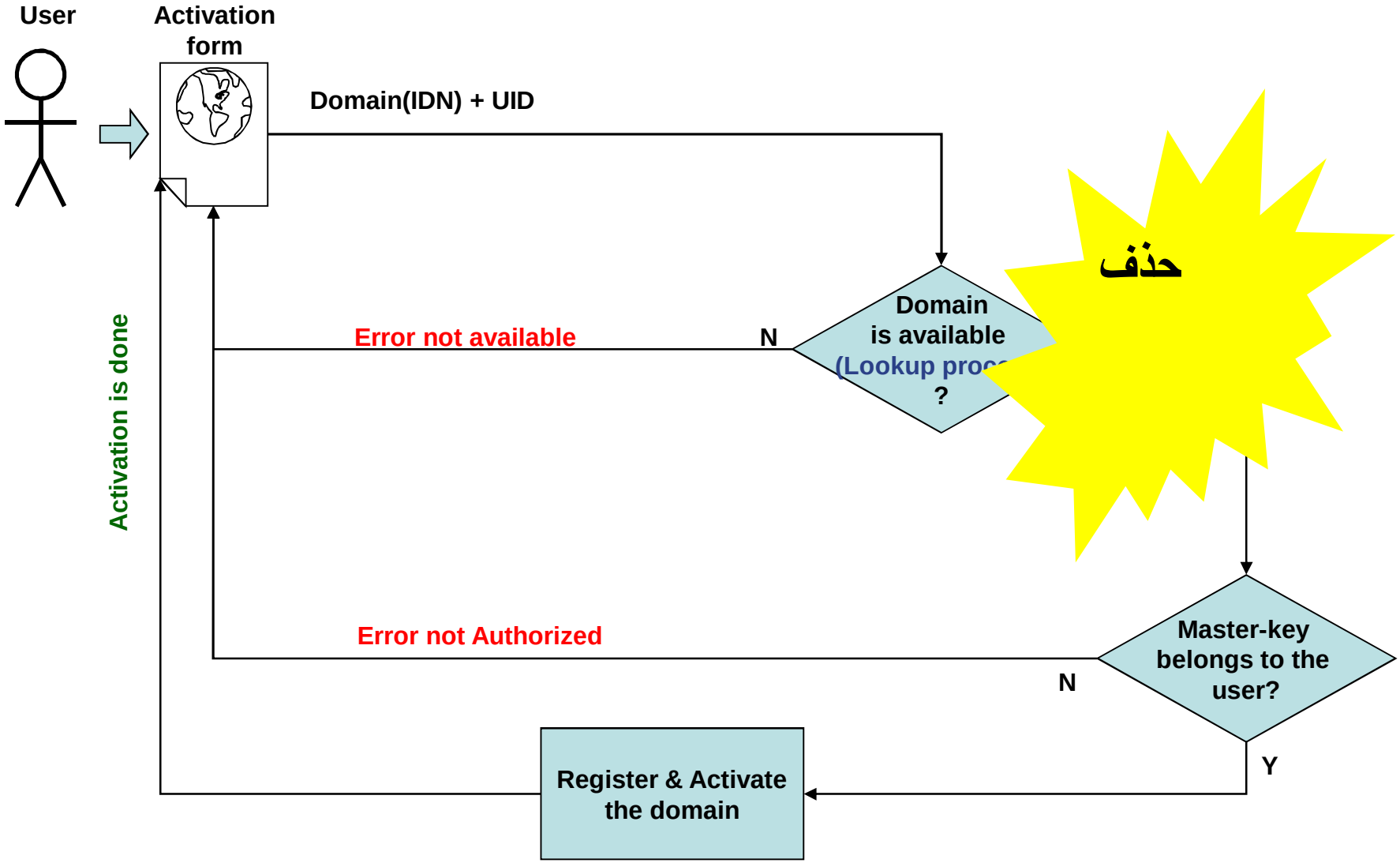
Registry Requirements
1. Lookup process (whois)



Registry Requirements
2. Registration process



Registry Requirements
3. Activation process



Steps to support a new language

- **Need process owners !!!!**
- **Language community should prepare**
  - Language Table
  - Exact Variant Table (EVT)
  - Typo Variant Table (TVT)
- **Support the new language:**
  - Add it to the list of supported languages
  - Add their Language table
  - Regenerate GVT
- **Communicate it with others (ICANN, Registries ..)**

حذف

Registry Requirements
Registry-Registrant interface

- **Lookup process (whois)**
  - Check domain syntax
    - Check if the domain within any supported language
  - Check if the same domain is available or not
    - Check if the request domain exists
    - If it is found return the unavailable / whois-information; otherwise
  - Check if the domain is variant for a registered domain n
    - Get the master- key for the string (based on GVT)
    - Check if the master- key was registered before or not
    - If master- key is found return unavailable / whois-information; o main is available
- **Registration process:**
  - Registrant should select one of these languages and t ain (U-Label)
  - Registry should accept inputs based on the selected lang age table
  - If domain name is register-able (available based on Lookup process)
    - Register the domain name along with its original image
    - List of allowable exact variants (to be activated if needed)
- **Activation process (enable exact variants)**
  - Original registrant can activate exact variants from his registered domain
    - List possible **Exact** variants that can be typed using one of the supported languages without intermingling between languages tables and taking care of position similarity (suggestion)
    - Activate one / many of the previous variants (if not activated before)

تبسيط