

# Arabic Script IDN Working Group (ASIWG)

ICANN Paris  
IDN Workshop  
26 Jun08

# Spot The Difference...



≠



**input[0] = U+0627**  
**input[1] = U+0654**  
**input[2] = U+066e**  
**input[3] = U+06ec**  
**input[4] = U+0627**

**input[0] = U+0623**  
**input[1] = U+0646**  
**input[2] = U+0627**

They Look The Same To Us ... But  
Not To A Computer

ı ≠ İ + ı  
U+0623 U+0654 U+0627

# When 1 is not 1...

Arabic-Indic VS. Eastern Arabic-Indic digits

— ٩ ٨ ٧ ٦ ٥ ٤ ٣ ٢ ١ .

— ٩ ٨ ٧ ٤ ٥ ٣ ٢ ١ .

١ ٢ ٣ ٧ ٨ ٩ .

**input[0] = U+06f1**  
**input[1] = U+06f2**  
**input[2] = U+06f3**  
**input[3] = U+06f7**  
**input[4] = U+06f8**  
**input[5] = U+06f9**  
**input[6] = U+06f0**

≠

١ ٢ ٣ ٧ ٨ ٩ .

input[0] = U+0661  
input[1] = U+0662  
input[2] = U+0663  
input[3] = U+0667  
input[4] = U+0668  
input[5] = U+0669  
input[6] = U+0660

# The Arabic Language is only a part of the Arabic Script table

0600		Arabic														06FF	
	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			0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6	
7			0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7	
8			0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8	
9			0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9	
A			062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA	
B	060B	061B	062B		064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB	
C	060C		062C		064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC	
D	060D		062D		064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD	
E	060E	061E	062E		064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE	
F	060F	061F	062F		064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF	

# ASIWG

- Stands for **Arabic Script IDN Working Group**
- **ASIWG** is a self-organizing group that consists of interested parties in the implementation of Arabic script in Internationalized Domain Names

# Initial Goals

## Initial Goals

- Establish a framework for implementation of IDNs in Arabic Script which:
  - Requires development of language tables & related variants (confusingly similar characters)
- Consider technical, linguistic & policy matters in the development of a Unified Arabic script table:
  - Highlight problems & come up with recommendations or guidelines where feasible
  - May serve as a best practice for other scripts to follow

# Current Representation

- Current Participation:
  - Egypt – Iran – Kuwait – Pakistan – Saudi Arabia – Syria – UAE - Malaysia
  - Afiliias – PIR – UN ESCWA - APTLD
  - ICANN – ISOC Africa
  - Experts (Michael Everson & John Klensin)
- Languages represented:
  - Arabic – Persian – Urdu – Sindhi – Pashto - Jawi



# Guiding Principles

- **Standardized** solution conforming to IETF IDNA Protocol and IDN Guidelines to ensure security and stability of the Internet
- **Extendable** solution attempting to be inclusive with respect to all languages that implement the Arabic script worldwide
- **Simple & Transparent** solution guided by end-user requirements (registrant and navigator)
- **Fast & Easy** solution that is practical for registries to deploy
- **ccTLDs & gTLDs** solution, that works for both

# Structure

- Open Working Group
  - Decisions based on rough consensus
  - Members expected to provide substantive input into the WG, and contribute experts to the group
    - Members fund own participation, to the extent possible
  - Desire participation from:
    - Registries implementing Arabic script
    - Linguistic community experts
    - DNS/Standards experts

# Outcomes

- Reviewed the Arabic Code Block, agreed on a set of recommendations and sent them to the IETF (IDNA Protocol Group)
- Definition of:
  - Variants at the **script level** are:  
*Characters that are identical to each other in any of the four positions in the Arabic script*
  - Variants at the **registry level** are:  
*Two strings that appear to be confusingly similar to each other in more than one language implementing the same script*

# Accepted characters for Arabic, Persian, Urdu, and Pashto

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0																
1																
2																
3																
4																
5																
6																
7																
8																
9																
A																
B																
C																
D																
E																
F																

# Recommendations, so far

- Variants should be registered/reserved to the same registrant
- Restrict the 3 Arabic sets of numerals from mixing at the protocol level
- Dealing with <space> is more of an application layer issue, to be resolved later and potentially by a different group
- At this stage, implement IDNs without Harakat/Tashkeel (Diacritics)

# Other Ongoing Work

- Complete the identification of visually confusable characters
- Agreeing on Pvalid Characters / Code points that are not to be included in the Unified Arabic Script Table, at this stage at least
- Coming up with guidelines on the registration process for languages implementing Arabic Script
- Outreaching to other language communities
- Studying funding options and sustainability

# Channels to Submit Work

- Submit the group's recommendations as either Informational RFCs or Best Current Practice (BCP) RFCs to the IETF
- Submit the group's recommendations to ICANN

# Online Resources

- **Wiki Page:**

<http://www.asiwg.org>

- **Mailing List:**

[Idna-arabicscript@lists.irnic.ir](mailto:Idna-arabicscript@lists.irnic.ir)

- **Archives:**

<http://lists.irnic.ir/mailman/listinfo/idna-arabicscript>



# Thank You

<http://www.asiwg.org>

Next Meeting – Nov 2008 | Next Call – Jul 2008

Mailing List: [ldna-arabicscript@lists.irnic.ir](mailto:ldna-arabicscript@lists.irnic.ir)

# Appendix

# Meeting Plans

- Face-to-face meetings, in the region:
  - First meeting: in Apr08 in Dubai back-to-back with ICANN regional meeting
  - Second meeting: in May08 in Dubai
  - Next face-to-face in Nov08 in Cairo, after the ICANN meeting
- Meetings organized and supported by Afiliias, PIR & UN ESCWA, TRA (UAE). Support promised also by APTLD
- Attempt to coincide with other Internet meetings, to the extent feasible
- Conference calls arranged to keep the working going inter-sessionally, planned: Jul08 – Aug08 – Oct08

# Approach

Level 4	Application / User Level
Level 3	Registry / Language / Policy Level
Level 2	Unified Arabic Script Table
Level 1	IDNA Protocol Level

- Consideration of a layered model to address issues with regards to scripts
- Attempt to solve as much issues as possible on the lower levels

# Level 1: IDNA Protocol Level

Level 4	Application / User Level
Level 3	Registry / Language / Policy Level
Level 2	Unified Arabic Script Table
Level 1	IDNA Protocol Level

Level 1: deals with decisions that should be made at the IDNA protocol level (Pvalid, Disallowed, ContextO)

# Level 2: Unified Arabic Script Level

Level 4	Application / User Level
Level 3	Registry / Language / Policy Level
Level 2	Unified Arabic Script Table
Level 1	IDNA Protocol Level

Level 2: deals with decisions that should be made across all of the Arabic Script, with an intent to be inclusive of all languages that uses the script

# Level 3: Registry/Lang/Policy Level

Level 4	Application / User Level
Level 3	Registry / Language / Policy Level
Level 2	Unified Arabic Script Table
Level 1	IDNA Protocol Level

Level 3: deals with decisions that should be made on a language implementation basis at the registry level

# Level 4: Application / User Level

Level 4	Application / User Level
Level 3	Registry / Language / Policy Level
Level 2	Unified Arabic Script Table
Level 1	IDNA Protocol Level

Level 4: deals with decisions that should be made at the software application or user level