

Internationalized Domain
Names Registration and
Administration Guidelines
for Arabic Characters Group
of Languages

(Arabic, Persian, Urdu,...)

<[draft-bakleh-reg-adm-acg-apu-01.txt](#)>

[Target Category: Standards Track]

Expires: February, 10, 2005

Editors : Mhd. Elfatih Altijani.
Khaled Al Ahmad.
Omar Bakleh.
Reem Dannan.

Authors: Rifaah Ekrema

Fidaa Aljundi

Mohamed A. Elhamalaway, Ph.D.

Rashed Zantout, Ph.D.

Ahmad Alkassoum, Ph.D.

AbdulRahman Aljadhai Ph.D.

Yasir El Ameen

Sameh Nough

Sami Taieb Alasmaa

September, 8, 2004

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

Internationalized Domain Names Registration and
Administration Guidelines for Arabic Characters Group of
Languages (Arabic, Persian, Urdu,...)

Status of this Memo:

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#) except that the right to produce derivative works is not granted.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or to be made obsolete by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at: <http://www.ietf.org/shadow.html>.

Omar B.

Expires February 8, 2004

[Page 1]

Abstract:

This document provides guidelines for zone administrators(including but not limited to registry operators and registrars), and information for all domain names holders, on the administration of those domain names which contain characters drawn from Arabic Characters Group of Languages. Other language groups are encouraged to develop their own guidelines as needed, based on these guidelines if that is helpful.

The document gives basic guidelines for IDN registrars (as it is the case for IETF Document that talks about Japanese, Chinese and Korean domain name registration "[RFC 3490](#)"). The document provides also information for owners of IDN that contains Arabic characters on name reservation process. The document does not cover Arabic gTLD or ccTLD problems.

Comments on this document can be sent to the authors at arabic-idn-admin@ietf.org.

Table of content:

0.	Pre-Note for ASCII-version of this document.....	4
1.	Introduction.....	4
2.	Specialty of Arabic Characters Group of languages.....	4
3.	Arabic Domain Names Recommendations.....	5
4.	Basics of searching for Arabic Domain Names.....	7
5.	Ways of saving an Arabic Domain Name.....	7
6.	Administration framework of Arabic Domain names.....	8
7.	Principles underlying these guidelines.....	8
8.	Registration of IDL.....	8
9.	Versioning of the language character variant tables.....	9
10.	Technical Recommendations.....	10
11.	Full Copyright Statement.....	10
12.	Security Considerations.....	11
13.	IANA Considerations.....	11
14.	References.....	11
15.	Terms.....	11
16.	Authors.....	12

0. Pre-Note for ASCII-version of this document

In order to make meanings clear, especially in examples used to clarify some ideas, such examples will be written using their Unicode representation according to Unicode Standard 3.0.

1. Introduction:

Introducing domain names as addresses on the internet added a new vision to the Internet, it made internet addresses easy to remember, and meaningful more than using sequences of numbers that does not mean any thing to their user. Nowadays Internet users are looking forward to surf on the net using more meaningful names in their own language. Such names are called Internationalized Domain Names (IDNs). This demand opened a wide field of research and ideas as wide as the diversity of languages used by the people on the earth. Each of these languages has its own writing and reading rules. This fact threatens the integrity and the stability of Internet, unless we invent a good solution that respects and controls this mixture of rules and cultures and represents it with a unique, robust and easy-to-use way.

2. Specialty of Arabic Characters Group of Languages:

Arabic language is the official language of 22 countries; it is also used by more than 43 Islamic countries that use Arabic characters and scripts. In other words more than one billion potential users could be interested in Arabic Domain Names. The Arabic language as well as all Arabic Characters Group of Languages (Persian, Urdu, Pashto, ..) has many specialties that have to be considered when specifying any solution for Arabic Domain Names. The main specialties of these languages are summed up in the following:

- 1.2. Scripts are written from right to left.
- 2.2. The shapes of the character change in most cases according to its location in the word.
- 2.3. Characters within the word are mostly conjugated with preceding and succeeding characters.
- 2.4. Some characters does not conjugate with following character.

Omar B.

Expires February 8, 2004

[Page 4]

2.5. The structure of the sentence starts from the most general to the more specific (the opposite of the English language) (Internet draft) in Arabic is said: (draft {of} Internet).

2.6. Special characters called Tashkeel are used in order to give the right pronunciation for the word which affects the meaning. In other words tashkeel can give one word two or More different meanings. Tashkeel is not a seprate character by itself, but it modifies a normal Arabic character to give it the right pronuciation It has to be noted here that word meaning is mostly known from the context of the sentence or even the phrase in which the word is used. For Internet names these special signs have to be included but their application in the Internet Domain Names could be delayed for the time being.

2-7. Correct words are written in a unique way, but the character shaddah (U+0651); which falls under the tashkeel signs and means doubling the letter associated with it; sometimes is implicit in the word and sometimes is written. In these two cases the same word; with or without shaddah; has (often) the same meaning; but in some cases shaddah can change the meaning of the word. So to decide considering or ignoring such character an algorithm has to implemented.

2.8. Correct orthographical practices about Hamza associated with or without Alef and Yeh have to be followed. The same applies to Yeh, Heh and Teh Marbuta.

3. Arabic Domain Names Recommendations:

3.1. Arabic domain names must allow the use of (SPACE) character. This condition is vital, as Arabic characters (as cited above) can be conjugated with each other and have different shapes depending on their position in the word. If we do not allow the space, words that form the ADL will be misread and misunderstood by the users. The use of dash character (-) to separate words is NOT acceptable according to IDNA ([RFC-3490](#)) which prevents the use of characters that belong to another language when reserving a name for a certain language, besides the fact that this character is not used in Arabic, giving an odd view of the domain name.

3.2. Solving the issue that results from shaddah could be done by an algorithm that generates all possible synonyms that result from the implicit existence of shaddah. Grammatical rules that control the implicit existence of shaddah do exist. Such a procedure will prevent registering the same word that contains shaddah more than once according to the existence or the absence of this implicit shaddah by different individuals. Explicit shaddah which does not come from grammatical rules has to be treated as a character, so the registration system has to consider the existence or the absence of this shaddah as a differentiator between different domain names.

3.3. Tashkeel Embodiment [Shadda (U+0651), Fatha (U+064E), Damma (U+064F), Kasra (U+0650) and Superscript Aleph (U+0670)] must be taken in the accredited character table, as it will give the possibility to register one name more than once by using these tashkeel characters with some or all alphabetical characters of a certain name (there are 8 possible usages of tashkeel characters that can be used with every other character). In spite of the need to use these characters for correct Arabic Domain Names, we can postpone using these five characters for future use. So tashkeel characters have to be added to the accepted character set of Arabic Domain Names but they have to be ignored currently in the registration system. Meanwhile registration systems have to equate Alef Maksura (U+0649) with Yeh (U+064A) in their processing preventing registering more than one domain name based on these two characters.

3.4. Future standards for Arabic Domain Names have to abide to the specificity of the Arabic characters group of languages, which is also valid for all other language groups. A special attention is drawn to ([RFC 3490](#)) which does not allow mixing characters from any group of languages with characters belonging to another group.

The above recommendations are valid for Persian, Pashto and Urdu and other languages which use Arabic Characters.

4. Basics of searching for Arabic Domain Names:

Name variants must be considered when searching an ADL, if any name in an ADL package is the subject of a name resolving query, a positive answer is to be given if the package was reserved (except when using one of that common abbreviations). The resolving system on the client side must not ignore tashkeel for future developments, and it should abide to the rules dealing with shaddah.

5. Ways of saving an Arabic Domain Name:

5.1. Reservation of an ADL:

As it is mentioned above, similarity cases (Name variants) must be considered when reserving an Arabic domain name, all names resulting from the similarity cases must be reserved, this will prevent reserving different Arabic names that actually indicates the same content by different persons. We can use the following steps when reserving an Arabic domain name:

```
IN = IDL to be registered
if Is Valid(IN) then
Begin
  For each Name in [Names variants of (IN)] do
    If Is Valid (Name) Then
      Reserve Name
    End If
  End For
End If
```

Where Is Valid÷ as some algorithm that verifies if that the name is compliant with the ADLs standards? We will call the ADL variants as ADL package.

Registering town, city and country names as well as names bearing pure religious meanings could not be registered as Arabic Domain Names. At the same time the system should not allow registering names having meanings that contradict the culture of the people of the Arabic characters group of languages. The system must also abandon registering linguistically non-correct names.

5.2. Activation and deactivation of an ADL:

If any name of an ADL package is to be activated or deactivated ALL names within the ADL package must be activated and deactivated at the same time, the administration system must provide some mechanism to insure this process.

5.3. Deleting an ADL:

All names in an ADL package must be deleted when any name in the package is requested to be deleted.

6. Administration framework of Arabic Domain Names:

The ADL Administration framework is responsible of affording a mechanism that respects all the previous conditions of dealing with an ADL.

7. Principles underlying these guidelines:

The previous guidelines must be considered with all Arabic characters group of languages. Registration systems for each language could be separated so that every language has its own registration systems. At the same time coordination between these systems is required to prevent reserving some words that have the same writing in more than one language but may have different meaning in another language.

This must not affect the integrity of the Internet, as users of a certain national domain name must be able to use domain names of an other nationality. This can be accomplished by giving each nationality of those who use the same characters a special string as a Nationality Identifier (or NID). This NID will help the ADL resolving system to determine the parameters of the " ToAscii " function ([RFC 3490](#)).

8. Registration of Arabic ADL:

ALL the names contained in an ADL package must be reserved automatically by the reservation system. This will prevent registering ADLs that have the same meaning but written differently by more than one person.

9. Versioning of the language character variant tables:

It is recommended to use the last version of the UNICODE. Only the following Unicode characters are accepted in Arabic domain names (according to Arabic Language standards).

U+0020	SPACE
U+0621	ARABIC LETTER HAMZA
U+0622	ARABIC LETTER ALEF WITH MADDA
U+0623	ARABIC LETTER ALEF WITH HAMZA
U+0624	ARABIC LETTER WAW WITH HAMZA
U+0625	ARABIC LETTER ALEF WITH HAMZA BELOW
U+0626	ARABIC LETTER YEH WITH HAMZA ABOVE
U+0627	ARABIC LETTER ALEF
U+0628	ARABIC LETTER BEH
U+0629	ARABIC LETTER THE MARBUTA
U+062A	ARABIC LETTER TEH
U+062B	ARABIC LETTER THEH
U+062C	ARABIC LETTER JEEM
U+062D	ARABIC LETTER HAH
U+062E	ARABIC LETTER KHAH
U+062F	ARABIC LETTER DAL
U+0630	ARABIC LETTER THAL
U+0631	ARABIC LETTER REH
U+0632	ARABIC LETTER ZAIN
U+0633	ARABIC LETTER SEEN
U+0634	ARABIC LETTER SHEEN
U+0635	ARABIC LETTER SAD
U+0636	ARABIC LETTER DAD
U+0637	ARABIC LETTER TAH
U+0638	ARABIC LETTER ZAH
U+0639	ARABIC LETTER AIN
U+063A	ARABIC LETTER GHAIN
U+0641	ARABIC LETTER FEH
U+0642	ARABIC LETTER QAF
U+0643	ARABIC LETTER KAF
U+0644	ARABIC LETTER LAM
U+0645	ARABIC LETTER MEEM
U+0646	ARABIC LETTER NOON
U+0647	ARABIC LETTER HEH
U+0648	ARABIC LETTER WAW
U+0649	ARABIC LETTER ALEF MAKSURA
U+064A	ARABIC LETTER YEH
U+064E	ARABIC FATHA
U+064F	ARABIC DAMMA
U+0650	ARABIC KASRA
U+0651	ARABIC SHADDA

Omar B.

Expires February 8, 2004

[Page 9]

U+0660	ARABIC-INDIC DIGIT ZERO
U+0661	ARABIC-INDIC DIGIT ONE
U+0662	ARABIC-INDIC DIGIT TWO
U+0663	ARABIC-INDIC DIGIT THREE
U+0664	ARABIC-INDIC DIGIT FOUR
U+0665	ARABIC-INDIC DIGIT FIVE
U+0666	ARABIC-INDIC DIGIT SIX
U+0667	ARABIC-INDIC DIGIT SEVEN
U+0668	ARABIC-INDIC DIGIT EIGHT
U+0669	ARABIC-INDIC DIGIT NINE
U+0670	ARABIC LETTER SUPERScript ALEF

10. Technical Recommendations:

10.1. The ADL solutions is like any other IDN subject of approved RFCs that speaks about the technical details of the realization of IDNs. A solution must be developed using the IDNA ([RFC 3490](#)). This in our opinion is the proper way to keep the integrity of the Internet. The solution has also to take the nameprep standard in consideration. We have to notice that the nameprep standard denies the use of the space in a IDNs, and have to note that such denial is not convenient for the Arabic Languages. As mentioned above using the space as a separator between words is not a negotiable matter (from a lingual point of view), so any development of an ADL solution must provide a reasonable answer that enables ADLs to contain spaces otherwise the solution will not be compliant with the Arabic language organizations recommendations in this field.

10.2. The solution must handle the problem of variants that results from the existence of space as a separator, this solution can be achieved by ignoring the absence of space between words where a word ends with a non joint letter. The treatment of those variants must be considered when designing the registration system.

10.3. Tashkeel must be ignored only by the registration system, and the layer that provides the ToAscii function. This procedure gives the user the ability to use Tashkeel without affecting the functionality of the IDNA.

10.4. The variants resulting from the misuse of hamza with or without Alef and Yeh and incorrect use of Yeh, Heh and Teh Marbuta have to be treated as separate cases.

11. Full Copyright Statement:

Copyright (C) The Internet Society (2005).

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

12. Security Considerations:

This memo relates to IETF process, not any particular technology. There are security considerations when adopting any technology, but there are no known issues of security with IETF Contribution rights policies.

13. IANA Considerations

IANA is expected to create and maintain a registry of algorithm names to be used as "Algorithm Names" as defined in [Section 2.3](#). The initial value should be "HMAC-MD5.SIG-ALG.REG.INT". Algorithm names are text strings encoded using the syntax of a domain name. There is no structure required other than names for different algorithms must be unique when compared as DNS names, i.e., comparison is case insensitive. Note that the initial value mentioned above is not a domain name, and therefore need not be a registered name within the DNS. New algorithms are assigned using the IETF Consensus policy defined in [RFC 2434](#). The algorithm name HMAC-MD5.SIG-ALG.REG.INT looks like a FQDN for historical reasons; future algorithm names are expected to be simple (i.e., single-component) names.

14. References:

[RFC3492]
Punycode: A Bootstring encoding of Unicode for
Internationalized Domain Names in Applications (IDNA)
A. Costello
Univ. of California, Berkeley
Category: Standards Track, March 2003

[RFC3491]
Nameprep: A Stringprep Profile for Internationalized
Domain Names (IDN)
P. Hoffman, IMC & VPNC
M. Blanchet, Viagenie
Category: Standards Track, March 2003

[RFC3490]
Internationalizing Domain Names in Applications (IDNA)
P. Faltstrom, Cisco, Category: Standards Track
P. Hoffman, IMC & VPNC, A. Costello, UC Berkeley, March 2003

15. Terms:

IDN: Internationalized Domain Name.

IDNA: Internationalized Domain Name in Application. ([RFC 3490](#))

ADL: Arabic Domain Label.

Variant: A name that have the same meaning but there exist small differences in the way they are written.

ADL package: Group of names resulting from the generation from the variants of an ADL.

Nameprep: A Stringprep Profile for Internationalized Domain Names; [draft-ietf-idn-nameprep](#), Feb 2002, Paul Hoffman, Marc Blanchet, work in progress. Punycode: An encoding of Unicode for use with IDNA, [draft-ietf-idn-punycode](#), Feb 2002, Adam M. Costello, work in progress.

16. Authors:

Rifaah Ekrema: rifaah@aietf.org

Founder and Chairman of Arab Information Engineers Task Force.

AIETF Organization

P.O: 30775

Damascus, Syria

Phone: +963 93 611087

Fax: +963 11 2238490

rifaah@aietf.org

Mohamed A. Elhamalaway:

Computers Engineering Professor in Al-Azhar University, Egypt.

Al-Azhar University, Systems & Computers Engineering dept.

Cairo, Egypt.

Phone: +20 6321465

Fax: +20 6377446

mhamalwy@yahoo.com

Fidaa Jondi:

Director of Bunyan Co. For Projects Managements.

Nozum Alhausabah Company,

Dubai, UAE.

Phone: +971 50 6507282

fida@hausabah.com

Ph.d. Rached Zantout:

Coordinator Department of Electrical Engineering,

Faculty of Engineering,

Hariri Canadian Academy.

Beirut, Lebanon

Phone: +961 3 842076

rached@cyberia.net.lb

Ph.d. Ahmed Guessoum:

Sharjah University, Sharjah, UAE

Phone: +973 50 5190558

guessoum@sharjah.ac.ae

Ph.D. AbdulRahman Aljadhai:

Technical Faculty, Riyadh, KSA

phone: +966 1 4529307

Mobile: +966 50 4420530

asj@linux.org.sa

Yasser Hassan Al Ameen:

President (Sudan Internet Society).

Sudan Internet Society, Khartoum, Sudan

Phone: +249 12 391157

yassir@isoc.sd

Sameh Nauh:

Head of Networks department in Damascus University,

Networks Department, Damascus University

Damascus, Syria

Phone: +963 93 469634

sameh@damasuniv.shern.net

Sami Taieb Alasmaa:

Arabic Language Teacher in Africa University.

Arabic Department, Khartoum University,

Khartoum, Sudan

Phone: +249 11 764892

Fax: +249 11 764893

dartaiebalasma@hotmail.com