Internet-Draft
draft-farah-adntf-adns-guidelines-02.txt
Category: Informational
Expires: June 2007

W. Abdel-Ati, i-DNS.net
A. Al-Haija, JUST
A. Al-Zoman, SaudiNIC
A. El-Sherbiny, ESCWA
M. Farah, ESCWA
K. Fattal, MINC
A. Hashim, Etisalat
C. Sha'ban, AGIP
December 2006

Guidelines for an Arabic Domain Name System (ADNS)

Status of this Memo

This memo provides information for the Internet community. Distribution of this memo is unlimited. Suggestions for improvements, amendments or additions are welcome. It does not specify an Internet standard of any kind.

This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

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.

Except for informational purposes, the use or duplication of the Arabic language related issues and solutions contained in this document is strictly prohibited in any way, shape and form without prior consent of the Arabic Domain Name Task Force (adn-tf@un.org) and MINC (sec04@minc.org).

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.

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 obsoleted 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://ietf.org/1id-abstracts.html

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

This Internet Draft is deemed to have met Multilingual Internet Consortium (MINC)'s minimum requirements on technical and linguistic matters; henceforth it has been issued the MINC RFC number MINCRFC AR0101, where "AR" is the designation of the Arabic language.

Copyright Notice

Abstract

There have been several attempts aimed at developing an Arabic Domain Name System (ADNS) using Arabic characters in an Arabic-language coherent fashion. In the beginning of the second quarter of 2003, an Arabic Domain Name Task Force (ADNTF) was formed under the auspices of United Nations Economic and Social Commission for Western Asia (ESCWA), and the guidance of Multilingual Internet Names Consortium (MINC); one of its main objectives was to help define standards for ADNS through a Request For Comments (RFC) document. This document resolves many technical and linguistic issues, including the adoption of the client-side DNS-based approach to name resolution; syntax of the proposed Arabic Domain Names together with the character set and many Arabic language-specific issues were clearly resolved. This Internet-Draft proposes guidelines that are compatible with the Internet Consortium for Assigned Names and Numbers (ICANN) and the Internet Engineering Task Force (IETF) as far as Domain Names System (DNS) and Internationalized Domain Names (IDN) standards are concerned. Technical, management, operational, and language-specific issues are discussed and recommendations are made.

Table of Contents

1. INTRODUCTION

The Arab region suffers from a digital divide that is mostly manifested in the form of the lowest regional Internet usage rate in the world. Language is identified to be one of the main barriers to widespread Internet usage. Along with the attempts to increase the volume of Arabic content on the Internet, there have been also several attempts aiming at the Arabization of Domain Names themselves. These attempts, when completely successful, will create the thrust for a second wave of Internet spread across the Arab region. The future of Arabic Internet names is imminent; there is substantial market and user demand for Arabic Domain Names. To satisfy this demand, the entire environment will need to be developed to take into account technology standardization, policy and administrative arrangements, as well as new applications. The significance of these efforts should not be underestimated, as it is part of a far nobler goal: the ongoing internationalization of the Internet.

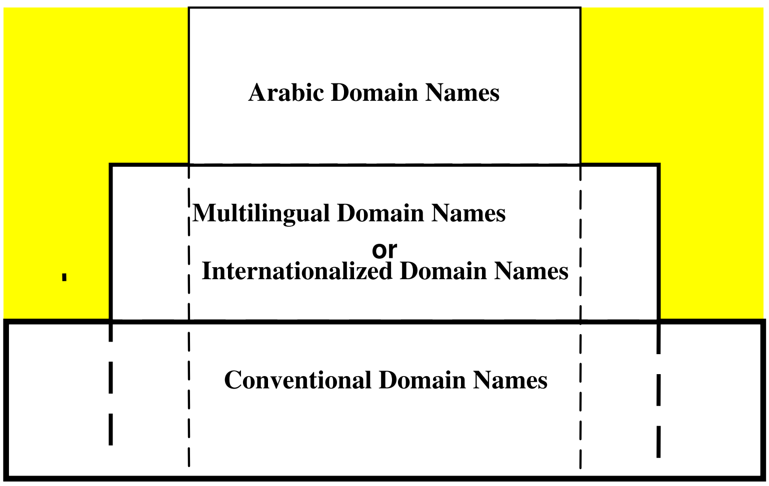
The IDN Standards issued by the IETF solve the generic domain name access issue for scripts beyond the limitation of the existing ASCII character set. Localized implementations are to be drawn from this set of standards. This draft provides specific guidelines for the use of Arabic language and provides a foundation for other documents encompassing languages that use similar scripts (e.g. Urdu, Farsi). The ADNTF will cooperate with experts from the Urdu and Farsi speaking Internet community in order to cover these languages and address other organizational and policy issues in an interoperable manner.

2. EVOLUTION OF ARABIC DOMAIN NAMES

The efforts exerted so far to define an Arabic Domain Name System (ADNS) were not done in isolation of the world; they were carried out within the context of the global movement towards Internationalized Domain Names (IDNs) and Multilingual Domain Names (MLDNs). Most of these IDNs or MLDNs were also developed within a wider framework of the Domain Name System (DNS).

In the conventional DNS, one has to differentiate between three types of players: (a) organizations, (b) technology providers, and (c) service providers: namely Registries/Registrars. Each of those three types of players is responsible for a different set of goals and normally undertakes a special set of activities.

The context of Arabic Domain Names



2.1 Dynamics of the Previous Phase

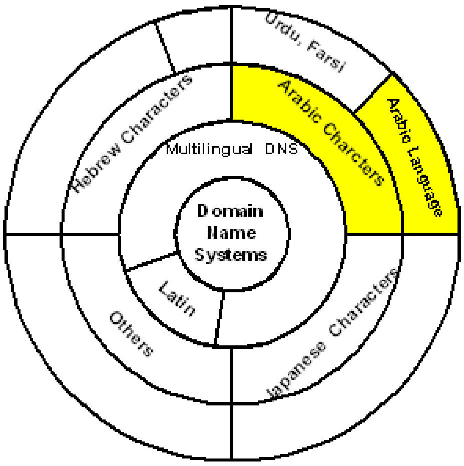
2.1.1. Global Evolution

During the previous five years, the evolution of MLDNs was not easy. While the Internet Consortium for Assigned Names and Numbers (ICANN) was evolving, it was naturally preoccupied with reorganization issues related to the entities responsible for coordinating the development of the conventional Internet. It was this preoccupation that left room for uncoordinated efforts and the emergence of competing standards for the ADNS creating a state of uncertainty.

MLDN activities and efforts were started in Eastern Asia, by Korean, Chinese and Japanese languages much earlier than the Arabic language; where a multitude of technology providers, registries, and registrars emerged.

Technologies differ amongst different providers mainly in terms of the manner in which they use the client-server relationship, in addition to the differences in the character-set and the language script itself.

Domain Name Systems and Languages



2.1.2. Regional Evolution

During this early period, implementations of the ADNS varied enormously amongst technology providers and their respective registries. Those technology providers competed feverishly in order to impose standards upon the community, and to create a status quo that they could use to reinforce their position and also to gain profits to sustain their innovation cycle.

The battle created a chaotic situation and standardization was not achieved; registries were technology-centric, and took the risk of adhering to standards and/or technologies that may become obsolete in a very short term. This consequently risked the sustainability of the Domain Names of their end-users. Further to this, the uniqueness of an Arabic Domain Name on the Internet is currently not guaranteed; so two entities/persons can register the same name on two different registries!

On the other side, many registries refused to implement any solution before it is adopted by an independent authority; consequently, most of the technology providers couldn't secure enough clients. Some of them went out of business due to the chaotic situation.

It was hoped that the Arab Internet Names Consortium (AINC) would assume the role of the coordinating body. Unfortunately, it couldn't due to internal conflicts. The absence of a strong regional coordinating body prevented development in this area. At the start of the year 2003, the situation could be summarized as follows:

- Professionals and consumers lacked awareness of the viability and importance of Arabic Domain Names in general;
- Time and effort was wasted on competing technologies and standards put a drain on the resources of emerging ADNS companies;
- The absence of a coordinating body reduced overall effectiveness and hampered efforts to move forward towards a regulated environment.

It remains to say that the past period resulted in an accumulation of experience amongst the involved players. This experience will be an asset that will facilitate the next phase of the ADNS evolution.

2.2. Milestones

The Internet Engineering Task Force (IETF) issued in March 2003 a set of RFCs for Internationalized Domain Names [N1, N2, N3] and (Appendix 2) - that are supposed to become the de facto standard for all languages. From then on, the battle for standards had been, to a large extent, resolved. New and emerging technology providers will no longer need to compete on the basic standards but rather on efficiency levels and the cost of the technology. All registries and registrars will be compatible and most importantly, the domain names themselves will be unique as they should be.

In the beginning of April 2003, the United Nations Economic and Social Commission for Western Asia (ESCWA) became involved in the revitalization of Arab regional efforts by calling for an Expert Group Meeting to be held in early June. On 11 June 2003, the Multilingual Internet Naming Consortium (MINC) announced its policies on linguistic and cultural relevance [I8]. In April 2004, the Joint Engineering Team (JET) produced an RFC 3743 [I7] on IDN registration and administration for Chinese, Japanese and Korean languages.

2.3. Revitalization of the Arab Regional Efforts

The aforementioned Expert Group Meeting at UN House in Beirut from 3 to 5 June, 2003 intended to establish a new roadmap for development of the Arabic Domain Name industry and discussed activities required to establish consensus on the ADNS. Considering the potential and impact of the ADNS, this meeting was intended to identify obstacles and set objectives and initiatives for the promotion of the ADNS in a coordinated fashion.

Upon the recommendations of the participants, an Arabic Domain Name Task Force (ADNTF) was formed under the auspices of ESCWA, which also acted as its secretariat. The following objectives were agreed upon:

- Raising awareness among stakeholders about the importance of the Arabic Domain Names System (ADNS);
- Defining standards for ADNS through a Request For Comments (RFC) document;
- Promoting the adoption of standards in a coordinated fashion;
- Obtaining global recognition for the adopted standards;
- Facilitating the deployment of these standards by the various stakeholders.

In preparing this Internet-Draft, three members of the ADNTF (namely Mr. Abdel-Ati, Mr. Al-Zoman, and Mr. El-Sherbiny) were given responsibility for drafting the document. Mr. El-Sherbiny acted as focal point/coordinator, discussing various issues with ADNTF members, compiling contributions and structuring the document.

The current phase is concerned with defining a set of agreeable and consistent standards for ADNS, which are compatible with existing domain naming standards. Producing this set of standards is a necessity in the process of streamlining the efforts of the region in the same direction.

On the other hand, the League of Arab States (LAS) has established the Arab Working Group on Arabic Domain Names, to decide and agree upon various issues related to establishing of an ADNS, among which are those topics studied in this document in addition to other organizational, technical and logistical issues. This Arab Working Group on AND organized its first meeting in Damascus on January-February of 2005 and issued a report whose recommendations are fully reflected in this version of the document.

3. ARABIC LANGUAGE-SPECIFIC ISSUES

The main objective of the creation of an ADNS is to have a vehicle to increase Internet use amongst all strata of the Arabic-speaking communities. If the structure or hierarchy of the ADNS does not meet certain core criteria, then the intended wide-scale dissemination of the Internet would be hampered.

Furthermore, a non-user friendly ADNS would further add to the ambiguity and the eccentricity of the Internet to the Arabic-speaking communities, thus contributing negatively to the spread of the Internet and leading to further isolation of these communities at the global level.

Hence, there have been intensive efforts especially those spearheaded by Dr. Al-Zoman and recently contributed to by ESCWA to reach some consensus on a multitude of linguistic issues with the following goals:

- To define the accepted Arabic character set to be used for writing domain names in Arabic;
- To define the top-level domains of the Arabic domain name tree structure (i.e., Arabic gTLDs and ccTLDs).

As indicated in the studies carried out by Dr. Al-Zoman [I4] and [I5], there are many valid criteria to evaluate the proposed Arabic generic top-level domains gTLDs, or the Arabic country code top-level domains ccTLDs namely:

(a) Length of the generic Top-Level Domain (gTLD) or the Country Code Top-Level Domain (ccTLD);
(b) Coherence and clarity;
(c) Consistency with the Arabic language;
(d) Ease of pronunciation;
(e) Extendibility.
(f) The name as a whole (x.y.z) can be easily guessable, by being as close as possible to the real-world name;
(g) The name as a whole (x.y.z) is acceptable to the native Arabic-speaker's ear, i.e. user friendly.

The last two items are necessary in order to achieve wide-scale dissemination. They are of utmost importance in the deployment and take-up of ADNS.

The first meeting of the Arab Working Group on ADN, held in Damascus January-February 2005, gave special attention to the above criteria and stressed the following:
(a) Simplification of the domain names, whenever possible, to facilitate the interaction of the Arabic user with the Internet.

(b) Adoption of solutions that do not lead to confusion either in reading or in writing, provided that this does not compromise the linguistic correctness of used words.
(c)Mixing Arabic and non-Arabic letters in the domain name is not acceptable.

3.1. Linguistic Issues

There are a number of linguistic issues that have been proposed with respect to the usage of the Arabic language in domain names. This section will highlight some of them. This section is extracted from the paper of Dr Al-Zoman [I4] and the report of the first meeting of the Arab Working Group on ADNS [N5]. For details the reader is encouraged review the references.

3.1.1. Tashkeel (Diacritics) and Shadda

In the start-up phase of ADNS, both Tashkeel and Shadda should not be supported in the zone file, yet they can be supported only in the user interface, and stripped off at the preparation of internationalized strings (stringprep) phase.

Later on, this guideline concerning the use of Tashkeel or Shadda can be revisited after adequate research and field studies.

3.1.2. Kasheeda or Tatweel (Horizontal Character Size Extension)

Kasheeda (Tatweel) should not be used in Arabic domain names.

3.1.3. Character folding

Character folding is the process where multiple letters (that may have some similarity with respect to their shapes) are folded into one shape. This includes:

- Folding Teh Marbuta and Heh at the end of a word;
- Folding different forms of Hamzah;
- Folding Alef Maksura and Yeh at the end of a word;
- Folding Waw with Hamzah and Waw.

With respect to the Arabic language, character folding is not acceptable because it changes the meaning of the words and it is against the simplest spelling rules. Replacing a character with another character, which may have the same shape but different pronunciation, will give a different meaning. This will lead to have only one form for a word out many other forms of words that are made by all the combinations of folded characters. Hence, the other forms will be masked by the common form. [I4]

"It is often that because of laziness or weakness in spelling, handwriting mixes between different characters (e.g., Heh and
Teh-Marbuta). However, this is not the case in published and printed materials. One of the motivations to support the Arabic language in domain names is to preserve the language particularly with the spread of the globalization movement. Hence, character folding is working against this motivation since it is going to have a negative effect on the principle and ethics of the language. Therefore, we should let the technology work for the language and not the other way. Character folding should not be allowed. "[I4]

3.2. Supported Character Set

It is recommended to use only the following UNICODE characters. These are based on the study and the report from the Arabic linguistic committee of AINC based on UNICODE version 3.1

TABLE 1: CHARACTERS FROM UNICODE ARABIC TABLE (0600—06FF)

Unicode Character Name

0621 ARABIC LETTER HAMZA
0622 ARABIC LETTER ALEF WITH MADDA ABOVE
0623 ARABIC LETTER ALEF WITH HAMZA ABOVE
0624 ARABIC LETTER WAW WITH HAMZA ABOVE
0625 ARABIC LETTER ALEF WITH HAMZA BELOW
0626 ARABIC LETTER YEH WITH HAMZA ABOVE
0627 ARABIC LETTER ALEF
0628 ARABIC LETTER BEH
0629 ARABIC LETTER TEH MARBUTA
062A ARABIC LETTER TEH
062B ARABIC LETTER THEH
062C ARABIC LETTER JEEM
062D ARABIC LETTER HAH
062E ARABIC LETTER KHAH
062F ARABIC LETTER DAL
0630 ARABIC LETTER THAL
0631 ARABIC LETTER REH
0632 ARABIC LETTER ZAIN
0633 ARABIC LETTER SEEN
0634 ARABIC LETTER SHEEN
0635 ARABIC LETTER SAD
0636 ARABIC LETTER DAD
0637 ARABIC LETTER TAH
0638 ARABIC LETTER ZAH
0639 ARABIC LETTER AIN
063A ARABIC LETTER GHAIN
0641 ARABIC LETTER FEH
0642 ARABIC LETTER QAF
0643 ARABIC LETTER KAF
0644 ARABIC LETTER LAM
0645 ARABIC LETTER MEEM
0646 ARABIC LETTER NOON
0647 ARABIC LETTER HEH
0648 ARABIC LETTER WAW
0649 ARABIC LETTER ALEF MAKSURA
064A ARABIC LETTER YEH
0660 ARABIC-INDIC DIGIT ZERO
0661 ARABIC-INDIC DIGIT ONE
0662 ARABIC-INDIC DIGIT TWO
0663 ARABIC-INDIC DIGIT THREE
0664 ARABIC-INDIC DIGIT FOUR
0665 ARABIC-INDIC DIGIT FIVE
0666 ARABIC-INDIC DIGIT SIX
0667 ARABIC-INDIC DIGIT SEVEN
0668 ARABIC-INDIC DIGIT EIGHT
0669 ARABIC-INDIC DIGIT NINE

Source: A. Al-Zoman, "Supporting the Arabic Language in Domain Names", October 2003

TABLE 2: CHARACTERS FROM UNICODE BASIC LATIN TABLE (0000-007F):

Unicode	Digit Name
0030	DIGIT ZERO
0031	DIGIT ONE
0032	DIGIT TWO
0033	DIGIT THREE
0034	DIGIT FOUR
0035	DIGIT FIVE
0036	DIGIT SIX
0037	DIGIT SEVEN
0038	DIGIT EIGHT
0039	DIGIT NINE
002D	HYPHEN-MINUS
002E	FULL STOP (Dot)

Source: A. Al-Zoman, "Supporting the Arabic Language in Domain Names", October 2003

3.3. Arabic Domain Name Structure

A domain name consists of multiple words (codes) that are separated by dots (u+002E). Based on research, rationale, and reference to [I4], after considering and weighing a multitude of alternatives and combinations and after eliminating of many possible combinations, the following structure is proposed for an Arabic Domain Name based on the conclusion that the geographical classification is adopted and there is no more activity classification corresponding to (.com), (.org), etc.

The proposed structure has the following syntax (to be read from right to left)

<A-TLD>.<entity-name>

Where, <entity-name> represents the Arabic name of the entity and
<A-TLD> represents an Arabic TLD. UNICODE values in hexadecimal form are written below from left to right representing Arabic characters originally typed from right to left.

Example 1:

شركة-الزومان.سعودية

u+0634 u+0631 u+0643 u+0629 u+02D u+0627 u+0644
u+0632 u+0648 u+0645 u+0627 u+0646 u+002E u+0633
u+0639 u+0648 u+062F u+064A u+0629

Example 2:

شركة-أرامكو.سعودية

u+0634 u+0631 u+0643 u+0629 u+02D u+0623 u+0631 u+0627
u+0645 u+0643 u+0648 u+002E u+0633 u+0639 u+0648 u+062F
u+064A u+0629

Example 3:

المركز-التجاري.سورية

u+0627 u+0644 u+0645 u+0631 u+0643 u+0632 u+02D u+0627
u+0644 u+062A u+062C u+0627 u+0631 u+064A u+002E u+0633
u+0648 u+0631 u+064A u+0629

Example 4:

اتحاد-كرة-الطائرة.عربي

u+0627 u+062A u+062D u+0627 u+062F u+02D u+0643 u+0631
u+0629 u+02D u+0627 u+0644 u+0637 u+0627 u+0626 u+0631
u+0629 u+002E u+0639 u+0631 u+0628 u+064A

Example 5:

جامعة-الخرطوم.سودان

u+062C u+0627 u+0645 u+0639 u+0629 u+02D u+0627 u+0644
u+062E u+0631 u+0637 u+0648 u+0645 u+002E u+0633 u+0648
u+062F u+0627 u+0646

One of the features of this structure is switching the order of reading and writing the category identifier to be at the beginning and to be part of the name. The rationale behind the sequence is that in the Arabic language, it is more proper not to use the company.com structure, but rather use com-company instead.

3.4. Recommended Arabic gTLDs and ccTLDs

Based on [I4], precisely, suggested Arabic gTLDs which use the entity type for the classification are not suitable for the Arabic language.

Therefore, with respect to Arabic TLDs, it is suggested to use the geographical classification as a start up for both Arabic gTLDs and Arabic ccTLDs.

For Arabic gTLDs, it is suggested to use geographical descriptive words such as (دولي)(u+062F u+0648 u+0644 u+064A) and (عربي)(u+0639 u+0631 u+0628 u+064A), which can be later expanded to include other activities such as educational or commercial.

As for ccTLDs, previous efforts have gone a great way towards establishing and implementing country-specific Arabic ccTLD names. Several alternatives underwent a long discussion process.

There were two choices for this representation [I6]. The first was based on a full word representation, while the second was based on a two-character coded abbreviation table [N4]. The full word option also involves the use, or lack thereof, of the Arabic noun identification letter (Al-Altareef)(ال)(u+0627 u+0644) depending on the country.

Although short names represent a high degree of practicality, some of the two letter abbreviations carry inappropriate meanings. Full word names, on the other hand, can be used within advertising material for clearer name representation.

Based on [N5], the Arab Working Group on ADNS recommended the referral to the League of Arab States' Arab Standardization Organization's specification no.: 642-1985, regarding the short names for Arab countries, with the adoption of the short names, and not the symbolic two-character coded abbreviation, as a ccTLD [N4] and [N5].

The appendix of [N5] indicates the that the standard short name of Arabic countries is to be used except when there is more than one word in this short name of a given country; in such a case, only one indicative word should be adopted like the single Arabic work "Alimaarat" instead of the 3-words name "Alimaarat Alarabyia Almotahhida", same for "Libya" instead of the "Libya Algamahiryia Alarabyia" and "AlKamar" instead of "Jozor AlKamar".

The following table below shows the recommended ccTLD codes for the Arab countries in the recommended single-word format. This table is adopted from the report on the first meeting of the Arab Working Group [N5].

Official State Names	Official State Names (Arabic)	Recommended name (single-word format) With or without Al-Altareef
Hashemite Kingdom of Jordan	الأردن	u+0627 u+0644 u+0623 u+0631 u+062F u+0646
United Arab Emirates	الإمارات	u+0627 u+0644 u+0625 u+0645 u+0627 u+0631 u+0627 u+062A
Kingdom of Bahrain	البحرين	u+0627 u+0644 u+0628 u+062D u+0631 u+064A u+0646
Republic of Tunisia	تونس	u+062A u+0648 u+0646 u+0633
People's Democratic Republic of Algeria	الجزائر	u+0627 u+0644 u+062C u+0632 u+0627 u+0626 u+0631
Federal and Islamic Republic of Comoros	القمر	u+0627 u+0644 u+0642 u+0645 u+0631
Republic of Djibouti	جيبوتي	u+062C u+064A u+0628 u+0648 u+062A u+064A
Kingdom of Saudi Arabia	السعودية	u+0627 u+0644 u+0633 u+0639 u+0648 u+062F u+064A u+0629
Democratic Republic of Sudan	السودان	u+0627 u+0644 u+0633 u+0648 u+062F u+0627 u+0646
Syrian Arab Republic	سورية	u+0633 u+0648 u+0631 u+064A u+0629
Somalia Democratic Republic	الصومال	u+0627 u+0644 u+0635 u+0648 u+0645 u+0627 u+0644
Republic of Iraq	العراق	u+0627 u+0644 u+0639 u+0631 u+0627 u+0642
Sultanate of Oman	عمان	u+0639 u+0645 u+0627 u+0646
Palestine	فلسطين	u+0641 u+0644 u+0633 u+0637 u+064A u+0646
State of Qatar	قطر	u+0642 u+0637 u+0631
State of Kuwait	الكويت	u+0627 u+0644 u+0643 u+0648 u+064A u+062A
Lebanese Republic	لبنان	u+0644 u+0628 u+0646 u+0627 u+0646
Socialist People's Libyan Arab Jamahiriya	ليبيا	u+0644 u+064 u+0628 u+064A u+0627

Arab Republic of Egypt	مصر	u+0645 u+0635 u+0631
Kingdom of Morocco	المغرب	u+0627 u+0644 u+0645 u+063A u+0631 u+0628
Islamic Republic of Mauritania	موريتانيا	u+0645 u+0648 u+0631 u+064 u+062A u+0627 u+0646 u+064A u+0627
Yemen Arab Republic	اليمن	u+0627 u+0644 u+064A u+0645 u+0646

Source: ESCWA ICT Division, May 2005.

3.5. Arabic Linguistic Issues Affected By Technical Constraints

In this section the technical aspect of some linguistic issues as well as TLD mapping is discussed

3.5.1. Numerals

According to Dr. Al-Zoman [I4], in the Arab world, there are two sets of numerical digits used

- Set I: (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) mostly used in the western part of the Arab world.

- Set II: (٠، ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨،٩) (u+0660, u+0661, u+0662, u+0663, u+0664, u+0665, u+0666, u+0667, u+0668, u+0669) mostly used in the eastern part of the Arab world.

Although visual differentiation between the Arabic zero (٠) (u+0660)
and the dot (.)(u+002E) in printed material is possible (the zero is larger in size and is printed higher than the dot), using it in domain names may lead to confusion. Folding set II to set I will eliminate the problem of the zero, in specific, and that of numerals in general.

According to [N5], the recommendation is that both sets may be supported in the user interface and that both are folded to one set (Set I) at the preparation of internationalized strings (e.g., "stringprep") phase; i.e. storage of numerals in the zone file is done in ASCII format.

3.5.2. The Space Character

The space character is strictly not allowed in domain names, as it is a control character. Instead, the hyphen (Al-sharta) (i.e.u+02D) is proposed as a separator between Arabic words: confusion can take place if the words are typed without a separator, unlike in ASCII. According to [N5], it is acceptable to use the hyphen to separate between words within the same domain name label; however, it is recommended to find technical solutions that can enable the use of the space character for this purpose.

4. THE SOLUTION CONCEPT

4.1. DNS-Based Solution

Historically, there have been different approaches to the ADNS problem. Solutions fell under one of two categories, namely DNS solutions and Keyword solutions. "Keywords" are not domain names. Rather, they exist as an additional layer above the DNS. Therefore, whilst DNS-based solutions only require the use of the Internet's DNS resolution infrastructure, keyword-based solutions also require a "URL Forwarding" technique to map simple references/names/phrases to domain names or IP addresses.

As a pre-requisite to using "keywords", each resolvable domain name is registered in a keyword-based directory in addition to the DNS registry. The keyword directory is searched during the "look up" process, and matches in the keyword registry are used to locate a particular URL or a list of matching sites under that particular keyword. On the other hand, DNS-based solutions are IETF compliant, they do preserve the language integrity and they also allow hyperlinks.

Keyword approaches are viable only as a supplemental scheme over and above a robust DNS based solution, but they do not replace them. It is therefore recommended not to use a keyword-based solution, but rather to employ a DNS-based solution to preserve the integrity of the Arabic language, to eliminate any confusion and to become fully interoperable with existing DNS schemes [I1, I2, I3]. The realm of RFC 3743 from the JET adopts a similar solution for Chinese, Japanese, and Korean languages. [I7]

4.2. Client-Side Approach

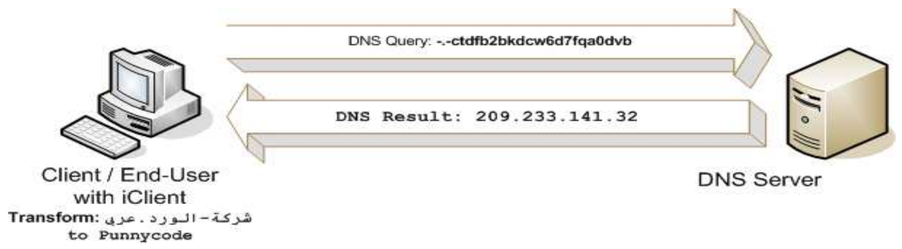
Generally, there are two schemes for resolving MDNs: server-based and client-based.

The proposed architecture for an ADNS is in accordance with the IETF standard for Internationalized Domain Names (IDN) [N1, N2, N3] which recommends that a client-side resolution scheme accommodate non-Latin languages like Arabic. This is a layer above the current Internet structure.

Different approaches for resolution of Multilingual Domain Names

Server side approach

Client side approach



To ensure a smooth and stable operational environment, further research is currently ongoing by ESCWA, on both the root server management and the client IDN standards implementation.

5. FUTURE OPERATIONAL CONSIDERATIONS

It is important to describe the operational aspects of ADNS in order to provide ccTLD owners with a set of guidelines and policies for operation. JET has made a similar effort in RFC 3743 published in April 2004 [I7].

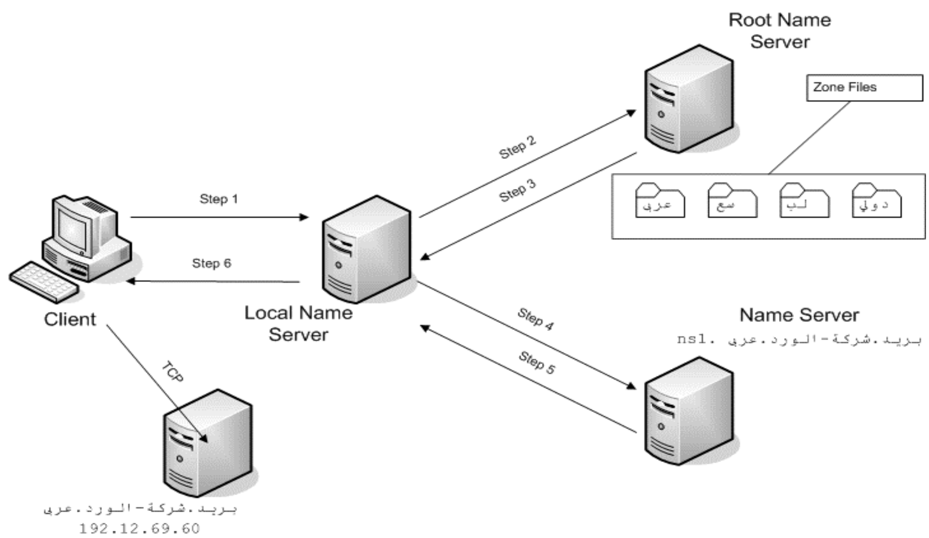
These issues cannot be discussed in detail within the scope of this document. Further efforts should be directed to enrich the operational aspects listed below:

5.1 Registrar-Related Aspects

Existing IETF documents describe registry-management methods, and registrars often develop applications to build DNS records based on data collected from domain owners within the guidelines of the adopted policies.
A regional regulatory authority should appoint either (a) a single entity or multiple entities working in coordination to maintain a registry of Arabic Domain Names. On the country-specific level, ccTLDs will be managed independently in each country by the country-appointed Network Administrator. In as far as ccTLDs are concerned; each country will run its own Arabic ccTLD along with the standard ASCII ccTLD. The regional regulatory authority will also be in charge of approving the Arabic representation of ccTLD names of non-Arab countries requesting such representation.

A good commercial model would be to follow the ICANN model where there are accredited registrars that can appoint resellers at a premium. That way the strong technically-qualified companies would act as registrars and a wide reseller network can be established.

Steps of ADN Resolutions



5.2 The Network Structure and Related Components

The proposed architecture for ADNS is based on a client-side resolution scheme, which is a layer above the current Internet structure. On the client side, workstations will be running some DNS resolution agent service at system level. So when the local agent receives a DNS resolution request from upper-level applications, it will take over the duty to talk to DNS servers configured for the workstation. When the agent receives responses from DNS server, it will pass back the results to upper-level applications.

Regarding IDN resolution, a software client intercepts the resolution request before it reaches local resolution agent, and replaces the multilingual query with ASCII Compatible Encoding (ACE) formatted value, in this case PUNNYCODE. So for local resolution agent, it just follows the normal DNS resolution process just as it does for ASCII formatted queries.

As an example, in order to resolve the domain name (بريد.شركة-الورد.عربي) (u+0628 u+0631 u+064A u+062F u+002E u+0634 u+0631 u+0643 u+0629 u+02D u+0627 u+0644 u+0648 u+0631 u+062F u+002E u+0639 u+0631 u+0628 u+064A) the process would be as follows:

Step 1: the client converts the domain name to PUNNYCODE and sends a query containing the domain name to the local name server.

Step 2: the local name server may not have the information about the domain name, so it sends the query to one of the root servers.

Step 3: the root server cannot match the entire name, so it returns the best match, i.e. the NS (name resolution) record for (شركة-الورد.عربى) (u+0634 u+0631 u+0643 u+0629 u+02D u+0627 u+0644 u+0648 u+0631 u+062F u+002E u+0639 u+0631 u+0628 u+0649). It also returns all records that are related to this record.

Step 4: the local name server sends the same query to the authoritative name server for the mail zone
(ns1.بريد.شركة-الورد.عربى) u+06E u+073 u+0031 u+002E u+0628 u+0631
u+064A u+062F u+002E u+0634 u+0631 u+0643 u+0629 u+02D u+0627 u+0644 u+0648 u+0631 u+062F u+002E u+0639 u+0631 u+0628 u+0649).

Step 5: the server has information about the domain and returns the answer: IP address = 192.12.69.60

Step 6: the local name server then responds to the client with the IP value, the client can then establish a TCP connection to the destination.

6. CONCLUSIONS, RECOMMENDATIONS, AND OPEN ISSUES

6.1. Conclusions

The proposed guidelines are in full accordance with the IETF IDN standards and take into account some Arabic language-specific issues as recommended by ICANN and by Dr. Al-Zoman's research. This is to ensure that an Arabic-language Internet, where access to Digital Arabic content is limited to some isolated portion of cyberspace, is never created.

As for linguistic issues, it is a compromise between grammatical rules of the Arabic language and the ease of use of the language on the Internet.

The proposed ADNS system is fully compatible with ICANN and IETF recommendations. It is a client-side solution to transform the Arabic UNICODE characters into an ASCII string that can operate in full compatibility with the existing Internet protocols and structure. In this way, the creation of an isolated Arabic 'intranet' is avoided. Arabic Domain Names will be transformed to PUNNYCODE representation at the client machine using a plug-in, the client would then communicate with the local name server using ASCII strings (which is the current standard of operation).

It is also proposed to go for TLD mapping for ccTLDs and gTLDs. It is also recommended to use the hyphen as a word separator and to use both Arabic numbers and Indian numbers to solve the zero and dot confusion problem.

6.2. Recommendations

It is time for the Arabic language to be widely disseminated on the Internet. With the number of Arabic Internet users in excess of 5 million and increasing exponentially, it is estimated that the quick implementation of the recommendations of this document, will open a market in excess of 500,000 domain names making this estimate based on the fact that the number of domains in existence in the Western world is 10% of the number of Internet users. Domain names and email addresses are key catalysts to the start of the Arabic Internet industry. Added to this, all the industries that will emerge like Web hosting, search engines and e-commerce will benefit from the development of an ADNS.

6.3. Open Issues

Stakeholders need to coordinate their efforts and collectively form a recognized regional regulatory authority. This authority will be entrusted to appoint a single entity, or alternatively set sufficient policy guidelines for multiple entities, to operate a worldwide registry of approved Arabic gTLDs. Such a scheme would be in line with the recommendations of RFC 3743 pertaining to zone administration.

The issue of trademarks and registration policies should also be identified by the regional regulatory authority along with legal experts. A list of forbidden or banned domain names must be identified to protect political or religious names; these alongside other political and ethical considerations will have to be examined further.
Migration issues still need to be considered and must only be carried out after a thoughtful and coordinated linguistic and technical strategy for seamless migration has been agreed upon by all stakeholders. Such a strategy could in principle include elements of TLD mapping solutions.

A broader collaboration is needed before all languages using the same script (e.g. Urdu, Farsi) can fully adopt a unified approach to domain name resolution. This work has to be done within the framework of existing standards in order to produce a coherent solution that serves all the languages, maintaining their individuality while providing a vehicle for better integration with the connected world.

7. ACKNOWLEDGEMENTS

ESCWA ICT Division provided support and partial funding for the development of this document with the objective of reaching a standard for a comprehensive Arabic Domain Names System (ADNS). Thanks are due to Mr. A. Farahat, ICT Division chief, for his guidance and supervision, and to Mr. S. Ghazzi for his efforts in re-formatting the document in adherence with IETF standards.

Thanks are due to MINC for persistent efforts in the field of Multilingual Domain Names, especially in promoting Arabic on the Internet and for its policies on linguistic and cultural relevance. Thanks are also due to SaudiNIC of King Abdulaziz City for Science and Technology (KACST) for its continuous efforts in supporting the development of Arabic Domain Names.

8. NORMATIVE REFERENCES

[N1] Faltstrom, P., Hoffman, P. and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.
[N2] Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", RFC 3491, March 2003.
[N3] Costello, "Punnycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003.

[N4] ASMO, "Arab Standard Specifications, No. 642-1985: Codes for Names of Countries and Languages", Arab League, Arab Standardization and Metrology Organization, 1985.
[N5] League of Arab Stated, report of the first meeting of the Arab working group on ADNS, Damascus, February 2005.

9. INFORMATIVE REFERENCES

[I1] http://www.auri.net/dns/How_Does_It_Work.html
[I2] http://www.verisign.com
[I3] http://www.icann.org/riodejaneiro/idn-topic.htm#5
[I4] A. Al-Zoman, "Supporting the Arabic Language in Domain Names", October 2003
[I5] A. Al-Zoman, "Arabic Top-Level Domains", paper presented in EGM on promotion of Digital Arabic Content, the United Nations, ESCWA, Beirut, June-2003.
[I6] W. Nasr Abdel-Ati "Tld Mapping for the Arabic domain Name system", Nov 2003.
[I7] Konishi, K., Huang, K., Qian, H., Ko, Y. "Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean", RFC3743, April 2004.
[I8] http://www.minc.org.

10. CONTACT INFORMATION

ADNTF Secretariat, ESCWA, ICT Division,UN House, Beirut, Lebanon,
adn-tf@un.org

Wael Nasr Abdel-Ati, i-DNS.net, 104 Elm Street,Menlo Park, CA 94025, USA, Wael@i-DNS.net

Ahmad Abu Al-Haija, Jordan University of Science and Technology(JUST), P.O. Box 840210, Amman, Jordan, haija@just.edu.jo

Abdulaziz H. Al-Zoman, PhD, Director of SaudiNIC, Internet Services Unit, KACST, Riyadh, Saudi Arabia, zoman@isu.net.sa

Ayman El-Sherbiny, Information and Communication Technology Division, ESCWA, UN-House, P.O. Box 11-8575, Beirut, Lebanon, El-sherbiny@un.org

Mansour Farah, Information and Communication Technology Division, ESCWA, UN-House, P.O. Box 11-8575, Beirut, Lebanon, farah14@un.org

Khaled Fattal, Multilingual Internet Names Consortium(MINC), 63 Robinson Road, 06-07, Afro Asia Building, Singapore 068894, khaledfattal@hotmail.com

Abdulla Hashim, EIM and UAE NIC, Etisalat, Dubai, United Arab Emirates, hashim@emirates.net.ae

Charles Sha'ban, Abu-Ghazaleh Intellectual Property (AGIP), TAGI CAMPUS Building No. 1, Queen Noor Street, P.O. Box 921100 Amman 11192 Jordan, cshaban@tagi.com

APPENDICES

Appendix 1: Abbreviations and Acronyms

ADNS: Arabic Domain Name System

TLD: Top-Level Domain
gTLD: Generic TLD
ccTLD: Country Code Top-Level Domain
DNS : Domain Name System
ICANN: Internet Corporation for Assigned Names and Numbers
ICT: Information and Communication Technology
IDN: Internationalized Domain Names
IETF : Internet Engineering Task Force
JET: Joint Engineering Team
KACST: King Abdulaziz City for Science and Technology
LAS: League of Arab States
MINC: Multilingual Internet Names Consortium
MLDN: Multilingual DNS technology
RFC: Request for Comments
SaudiNIC: Saudi Network Information Center

Appendix 2: IDN Standards
Source: http://www.verisign.com

IDN Standards Update

* IDN-related Requests for Comment (RFCs) published.

The Domain Name System (DNS) only recognizes ASCII characters A-Z, 0-9 and '-'. This limits the number of characters that can be utilized to build domain names to 37 of the more than 40,000 characters identified within Unicode. To create domain names from the wider range of Unicode characters, a character-encoding scheme that uniquely maps Unicode code points to an ASCII representation must be used and standardized.

The Internet Engineering Task Force (IETF) has led the effort in standardizing the way that non-ASCII characters are to be represented within and handled by DNS. The IETF published three standards related to Internationalized Domain Names (IDN):

* Encoding scheme for IDNs
* Name preparation
* IDNs in applications

Encoding Scheme

The encoding scheme for IDNs will be an ASCII Compatible Encoding (ACE) that will encode the local language characters of an IDN into ASCII characters such that DNS can accurately answer a request for an address record. There are several types of ACE. In order to select an ACE as the standard, IETF must consider the difficult balance between compression and implementation. The preferred ACE will allow the greatest number of characters (code points) to be represented and will not be difficult to deploy. The VeriSign IDN Test bed leverages an ACE known as Row-based ASCII Compatible Encoding (RACE). At the time of the opening of the Test bed, RACE was a leading candidate to become the standard. Today, another ACE known as Punycode is the leading candidate. Now that the standard has been published, Test bed is migrating to that standard.

Name Preparation

The name preparation standard will provide the rules that will ensure uniqueness in registering Unicode code points. The rules outline the criteria through which a set of non-ASCII characters will be refined to ensure that there is no ambiguity within the registrations of a specific name space. These rules are Mapping, Normalization and Prohibition.

* Mapping: Characters may be mapped to nothing, a single character or multiple characters based upon their usefulness in text only or case.
An example of usefulness: the soft hyphen (u+00AD) is discretionary and only has use within text and is invisible or ignored. The more common example is the mapping of a capital letter to a small letter such as 'B' (u+0042) to 'b' (u+0062). This is to ensure that a registration such as ibm.com does not have a conflict with other registration such as IBM.com or iBm.com.

* There are cases where a single character will map to multiple characters. The small letter sharp s or ' ' (u+00DF) has an upper case representation of 'SS' (u+0053, u+0053). This is also the same upper case representation for 'ss' (u+0073, u+0073). Therefore, ' ' maps to 'ss'.

* Normalization: Once a set of characters has been mapped, the set is normalized. Some input method editors (IME) enter characters that look exactly like another character, but have different code points. For example, 1 is a fullwidth digit one (u+FF11) and will normalize into a digit one (1) (u+0031). Normalization also ensures predictable results through ordering where characters have a number of combining diacritics.

* Prohibition: After normalization, the mapped and normalized set of characters is checked against a table of prohibited characters. These characters are prohibited for a variety of reasons but the most common are spaces that could lead to confusion and control characters that cannot be displayed.

IDNs in Applications

The IDN in applications standard focuses on the location where the Unicode to ASCII mapping will take place. IETF's approach makes the applications that send and receive traffic from DNS (browsers, e-mail clients, etc.) encode and decode the Unicode characters.

The Bottom Line

All of these issues are currently outlined in the IETF Internet draft entitled Preparation of Internationalized Host Names. The VeriSign IDN Testbed is following this draft and will change as this draft is updated.
In summary, enhancing the current DNS to include more than just English characters is not a simple undertaking. There are quite a few open issues surrounding the deployment and use of IDNs that need to be resolved by the IETF.

Character variants

The majority of domain name registrants register domain names that have meaning for them in their language - the domain name may be a name, word or phrase. These words or phrases have meaning in the registrant's language. Yet, the domain name may have different meanings in the context of other languages or cultures.

The domain name registration process was designed without consideration of language context. Technically speaking, the registrant registers a domain name using a set of characters within a script. Since scripts may be used by more than one language, the domain name is not registered in a specific language - it is registered in a specific script or combination of scripts. For example, the Latin script is used by many languages including English, French and German. A domain name registered using the Latin script could have meaning for several languages.

The overlap between scripts and languages define the variant issue. The Internationalized Domain Name (IDN) in Applications (IDNA) protocol enables the translation of all Unicode code points into unique ASCII strings. This broader range of characters has the potential to cause end-user confusion due to characters with similar appearances or interpretations, also known as variants. To reduce confusion and improve the end-user experience, it is necessary to address the variant issue.

While there are different types of variants, character variants are not covered by the recently released IDN-related Requests for Comment (RFCs) as local scripts and languages drive them. Communities throughout the world, especially in Asia-Pacific, have asked Top-Level Domain (TLD) registries to address character variant issues in their domain spaces to ensure a positive end-user experience. Implementing its character variant solution helps improve the end-user experience.

Appendix 3: IETF and ICANN Recommendations
Source: http://www.icann.org/riodejaneiro/idn-topic.htm#5

IETF is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual.

Actual technical work is done in Working Group organized by topics in several areas.

As mentioned above, a standard has come out of the IETF and recommended by ICANN.

Standards for ICANN Authorization of Internationalized Domain Name Registrations in Registries with Agreements http://www.icann.org/riodejaneiro/idn-topic.htm#5:

At the same time, the premise of this paper is that it would be a mistake for ICANN to pursue a burdensome and/or intrusive approach to IDN implementation" for example, by putting ICANN in the position of approving a character-equivalence table for each language, and of maintaining such tables. The deployment of IDNA within existing
top-level domain registries is fundamentally a registry responsibility, and the registries will be in the best position to make appropriate implementation decisions themselves, and should have the freedom to make adjustments as experience dictates. Just as DNS registries embrace a wide diversity in registration policies and administrative procedures, reflecting the diversity of local Internet communities, it seems apparent that the vast diversity of human character sets and the languages from which they come compels a
language-by-language, registry-led approach to the development of detailed registration policies and administrative procedures.

FULL COPYRIGHT STATEMENT

The document expires on June 2007.