INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2001-2004

ccTLD/KL Doc xxx

English only

Original: English

Joint ICANN/ITU ccTLD Workshop

Kuala Lumpur, Malaysia, 24 July 2004

DOCUMENT FOR ccTLD/KL WORKSHOP

Source: SaudiNIC, KACST, Saudi Arabia

Title: SaudiNIC's Contributions and Experiences in Supporting Arabic Domain Names

SaudiNIC's Experiences and Contributions in Supporting Arabic Domain Names

Abdulaziz H. Al-Zoman

Associate Professor
Director of SaudiNIC
King Abdulaziz City for Science and Technology
P.O. Box 6086, Riyadh 11442
Saudi Arabia
zoman@isu.net.sa

Abstract

Domain names are very crucial part of using Internet technology. They are still written using Roman characters regardless of the worldwide spread of the Internet. Other languages are not yet fully supported to locate resources and sites on the network. Nevertheless, using Arabic domain names is essential to increase the Internet penetration in the Arab world.

Supporting the Arabic language in domain names calls for investigating and addressing a number of questions related to linguistic issues and the Arabic domain name tree structure.

SaudiNIC (Saudi Network Information Center) has been active in this area for many years. It has collected valuable experiences and presented many contributions. This document highlights and discusses these contributions and offers some recommendations regarding the accepted Arabic character set to be used in Arabic domain names, as well as some recommendations concerning the appropriate Arabic generic and country code top-level domain names (i.e., Arabic gTLDs and ccTLDs).

Attention: This is not a publication made available to the public, but **an internal ITU-T Document** intended only for use by the Member States of the ITU, by ITU-T Sector Members and Associates, and their respective staff and collaborators in their ITU related work. It shall not be made available to, and used by, any other persons or entities without the prior written consent of the ITU-T.

TABLE OF CONTENTS

1. INTRODUCTION 3

2. SaudiNIC's Contributions 4

3. LINGUISTIC ISSUES 5
   3.2. Kasheeda (Tatweel) 5
   3.3. Character folding 5
   3.4. Numbers 6
   3.5. Connecting Multiple Words 6
   3.6. AINC-ALC Recommendations 6
   3.7. Supported Character Set 7

4. ARABIC TOP-LEVEL DOMAIN NAMES 9
   4.1. Evaluations of Existing Suggested Arabic TLDs 10
      4.1.1. Existing Suggested Arabic gTLDs 10
      4.1.2. Existing Suggested Arabic ccTLDs 11
      4.1.3. Results of the Study 12
   4.2. Open Issues 13
      4.2.1. Full words vs. Abbreviations 13
      4.2.2. The Arabic Equivalent of "WWW" 13
      4.2.3. Arabic TLD Categories 13
      4.2.4. Arabic Domain Name Structure 14
   4.3. Recommended Arabic gTLDs and ccTLDs 14

5. A PILOT PROJECT: IMPLEMENTING GCC ARABIC DOMAIN NAMES 16

6. CONCLUSIONS 17

ACKNOWLEDGEMENT 18

REFERENCES 19

1. INTRODUCTION

The Internet has become a global network of most, if not all, countries of the world with hundred of millions of users. Recently, it is estimated that more than 60% of the Internet contents are in languages other than English. Also, it is estimated that by the year 2003 there will at least be 30% of web users who prefer to do their on-line activities in a language other than English, and that by 2005 only one-third of Internet businesses will use English for on-line communication [1,2].

Domain names are used widely by Internet users to locate resources on the Internet in a format that is easy to remember and understand. These names, however, are not required by the network software, but are used for human mnemonic convenience. They are used instead of the numerical addresses which are known as Internet protocol (IP) addresses, which are mainly used by machines to route data packets on the Internet. Hence, the main objective of using domain names is to ease and simplify the use of the Internet [3,4,5].

Since the Internet was originally evolved in the United Sates, it supported only 7-bit ASCII code. Domain names consist of alphanumeric strings separated by dots, e.g., www.kacst.edu.sa. They are written using Roman characters particularly letters, digits, and hyphen. To the network, however, a domain name such as “www.kacst.edu.sa” is meaningless until it is translated into a numerical IP address. Name resolution is carried out by the Internet domain name system (DNS) in that domain names are mapped to the actual corresponding IP addresses.

Regardless of the worldwide spread of the Internet, the Internet penetration in the Arab world is about 1.4% [6] which is indeed very low. One of the obstacles facing the growth of this penetration is the language barrier. The Internet domain name system has not fully supported other languages to locate resources on the Internet. Users in non-English speaking countries, such as the Arab users, are at a disadvantage. Using domain names in a language that is different from the users' native language defeats the main objective of having the domain name in characters rather than just numbers.

Hence, SaudiNIC has recognized the importance of making the Internet supporting the Arabic language not only in web contents but also in their addresses. It is required that the Arabic language should be used from the start of switching on the user's personal computer until getting information from the Internet. Thus, eliminating the need for the user to enter non-Arabic web (URL) addresses particularly if the sites are in Arabic. There are a number of reasons why Arabizing domain names is needed [7], such as:

- Making the Arabic language part of the Internet.
- There is only a small percentage of Arabs who can read and write English.
- There are many well-known Arabic names that need to be used in the Internet.
- Users in the Arab world should have privileges to use their language on the Internet like English-speaking users.
- Roman letters are not capable of representing (or substituting) Arabic letters.
- Encouraging the use of the Internet by Arabs who do not speak English. As the trend nowadays for implementing e-government and e-business then it is important to provide the information and services in the user's native language.

This need is not only for the Arabic language but also for other languages. Multilingual domain names were first developed in Asia-Pacific countries in 1998 [1,2,8,9], which led later to the creation of a number of non-for profit organizations to supervise and pursue the deployment of multilingual domain names. Among these organizations are: the Multilingual Internet Names Consortium (MINC), the Arabic Internet Names Consortium (AINC), the Arabic Domain Name Task Force (ADNTF) under the supervision of UN-ESCWA, the Chinese Domain Name Consortium (CDNC), the International Forum for IT in Tamil (INFITT), and the Japanese Domain Names

Association (JDNA). Also, the Internet Corporation for Assigned Names and Numbers (ICANN) established an internal Internationalized Domain Name (IDN) Working Group, and the Internet Engineering Task Force (IETF) created an internationalized DNS group that have been dedicated for exploring the possibility of supporting internationalize Internet. The IDN group of IETF has issued so far three Request For Comments (RFC) for Internationalized DNS [10,11,12].

Several companies have begun to commercialize the technologies that have been developed to support multilingual domain names. These developments lack the standardized references. This is because, as usual, vendors are faster than the standardized bodies for proposing solutions. Therefore, current implementations of multilingual domain names are using proprietary technologies. ICANN adopted a resolution which recognizes that “it is important that the Internet evolves to be more accessible to those who do not use the ASCII-character set”, and stresses that “the internationalization of the Internet domain name system must be accomplished through standards that are open, non-proprietary, and fully compatible with the Internet’s existing end-to-end model and that preserve the globally unique naming in a universally resolvable public space” [13]. Hence, adopting proprietary solutions may lead to:

- Unrecognition by the international bodied such as ICANN and IETF.
- Incompatible solutions from technical and linguistic point of view.
- Multiple registrations for the same category.
- Disjoint networks each with its own an Arabic domain name space.

Therefore, it is urgently required from the Arabic Internet community to produce a set of standards that are acceptable by the Internet community in large. These standards should cover several aspects of supporting Arabic domain names at different levels, such as:

1. Linguistic issues and the accepted Arabic character set.
2. The Arabic domain name tree structure, i.e., Arabic gTLDs and ccTLDs.
3. Technical solutions to Arabize the domain name system
4. The administrative and organizational issues of Arabic root servers.

This document will address only the first two issues. The 3rd point is partially addresses by the IETF RFCs [10,11,12].

2. SaudiNIC's Contributions

The contributions of SaudiNIC in the area of supporting the Arabic language in domain names include the following:

- Publishing scientific papers [7,18, 20, 21, 22, 23].
- Establishing and/or participating in international and regional organizations such as the Multilingual Internet Names Consortium (MINC), the Arabic Internet Names Consortium (AINC), the Arabic Domain Name Task Force (ADNTF) under the supervision of UN-ESCWA, the Gulf Cooperation Council (GCC) ccTLDs group.
- Chairing the linguistic Committee of AINC and publishing its reports [14].
- Coauthoring an Internet Draft, "ADN Task Force Guidelines for Arabic DNS", that is still in the drafting mode (i.e., has not been published yet)
- Participating in conferences and meetings and presenting the importance of supporting Arabic domain names.
- Participating in a pilot project among the GCC countries testing the use of Arabic language in domain names.
- Establishing a web site for promoting the use of Arabic domain names:
  http://www.saudinic.net.sa/arabicdomain/arabic_domains.htm

The rest of this document will highlight the collective experiences and contributions of SaudiNIC with respect to Arabic domain names.

3. LINGUISTIC ISSUES

There are a number of linguistic issues that have to be discussed and agreed upon with respect to the usage of the Arabic language in domain names. This section will highlight some of them. For more detailed discussion see [7, 14].

3.1. Al-Tashkeel (Diacritics)

Al-Tashkeel (diacritic) is a small sing that is usually put on top or under an Arabic letter for the purpose of correct pronunciation which may leads to a different meaning. Al-tashkeel is not a letter by itself but it is a mean to correctly pronounce a letter. It is not widely used except incase of the possibility of mispronouncing words that have the same letters but with different pronunciations, and hence having different meanings.

Recommendation: With respect to domain names, al-tashkeel can be supported only in the user interface but should not be stored in the zone file. Therefore, it can be striped off at the preparation of internationalized strings ("stringprep") phase.

3.2. Kasheeda (Tatweel)

Kasheeda is not a letter. It is a horizontal line (like dash) used to lengthen the connection line between letters. It is used sometimes to enhance the display of Arabic words on screens or printouts.

Recommendation: Kasheeda should not be used in Arabic domain names.

3.3. Character folding

A 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 Alif Maksura and Yeh at the end of a word.
- Folding Waw with Hamzah and Waw.

Character folding is unacceptable in the Arabic language 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 (word) out many other forms of words that are made by all the combination of folded characters. Hence, the other forms will be masked by the common form.

Hand writing mixes between different characters (e.g., Heh and Teh-Marbuta) and this is due to laziness or weakness in spelling. 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 principles and ethics of the language.

Recommendation: Character folding should not be allowed.

3.4. Numbers

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 (al-maghrib al-arabi).
- Set II: (٠، ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩),
  Mostly used in the eastern part of the Arab world (al-mashriq al-arabi).

There have been some suggestions to use Set I because it is thought that there is similarity (or confusion) between the Arabic zero (0) and the dot (.). But the differences appear clearly in publications. The zero is larger and is printed higher than the dot. Also, With respect to a domain name, it is quite easy to distinguish between the zero and the dot based on the context of the domain name. And since the two sets are used they should be both supported.

Recommendation: Both sets should be supported in the user interface and both are folded to one set (Set I) at the preparation of internationalized strings (e.g., "stringprep") phase.

3.5. Connecting Multiple Words

In the Arab language words are separated by spaces. Connecting words without spaces is usually not acceptable. Therefore, a single space is the best word separator in an Arabic domain name with multiple words.

Recommendation: Space should be used to separate words if it is technically visible. Otherwise, it is recommended that multiple words are separated by the character "-" dash.

If the space is used as a word separator in Arabic domain names then it should be only a single space and it should not be used at the beginning or at the end of words.

3.6. AINC-ALC Recommendations

There are some non-for-profit international organizations that are trying to promote the multilingual issues and standards on the Internet, such as AINC (Arabic Internet Names Consortium). One of the strategic tasks of AINC is to setup some linguistic guidelines to be used by technology providers. Therefore, an Arabic Linguistic Committee was established during the formation of AINC in April 2000 to carry this task. One of the main goals of the committee is to define the accepted Arabic character set to be used for writing domain names in Arabic.

Table (I) lists Arabic Linguistic Committee recommendations [14] regarding some linguistic issues. More detailed information can be found in the following URL:
http://www.saudinic.net.sa/arabicdomain/arabic_domains.htm:

Table I: AINC-ALC Recommendations

Issue	*Recommendations*
Tashkeel (Diacritics)	**Tashkeel should not be allowed. However, if there is a need to allowed users to entered Tashkeel as part of a domain name then it should be stripped off by nameprep**
Kasheeda	**Kasheeda should be disallowed**
Folding Teh Marbuta + Heh **Folding different forms of Hamzah** **Folding Alif Maqsura+Ya**	**Folding should not be allowed**
Numbers **Arabic Zero**	**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 **Spaces**	**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.**
Accepted Character Set	**It is recommended to use UNICODE 3.1. The following Unicode characters are accepted in Arabic domain names:** **U0621(hamza) until U063A (gheen)** **U0641 (feh) until U064A (yeh)** **(U0660 until U0669)٩-٠Arabic numbers:** **Latin numbers: 0-9 (U0030 – U0039)** **Hyphen (U002D)** **Dot (U002E)** **Other than these characters are not allowed**

3.7. Supported Character Set

It is recommended to use only the following Unicode characters. These are based on the study in [7] and the report from the Arabic linguistic committee of AINC [14]. The following codes are based on Unicode version 3.1 [15].

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	(ه)	ARABIC LETTER HEH
0648	(و)	ARABIC LETTER WAW
0649	(ى)	ARABIC LETTER ALEF MAKSURA
064A	(ي)	ARABIC LETTER YEH
0660	(٠)	ARABIC-INDIC DIGIT ZERO
0661	(١)	ARABIC-INDIC DIGIT ONE
0662	(٢)	ARABIC-INDIC DIGIT TWO
0663	(٣)	ARABIC-INDIC DIGIT THREE
0664	(٤)	ARABIC-INDIC DIGIT FOUR
0665	(٥)	ARABIC-INDIC DIGIT FIVE
0666	(٦)	ARABIC-INDIC DIGIT SIX
0667	(٧)	ARABIC-INDIC DIGIT SEVEN
0668	(٨)	ARABIC-INDIC DIGIT EIGHT
0669	(٩)	ARABIC-INDIC DIGIT NINE

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)

4. ARABIC TOP-LEVEL DOMAIN NAMES

The domain name system (DNS) is a distributed database of host information that is organized in a hierarchal tree structure [3], see Figure (1). Theoretically, there is a "root domain" at the top of the domain name tree which is usually left unnamed. Immediately underneath the root come the top-level domains (TLDs). Basically, there are tow types of TLDs [3]. One is the generic TLDs (gTLDs), such as .com, .org, .net, and .edu. The second one is the country code TLDs (ccTLDs), such as .ae (United Arab Emirates), .bh (Bahrain), .ca (Canada), .de (Germany), .eg (Egypt), .jo (Jordan), .kw (Kuwait), .om (Oman), .qa (Qatar), .sa (Saudi Arabia), and .uk (United Kingdom). There are more than 240 ccTLDs following the two-letter country codes defined in the ISO standard number 3166 [16].

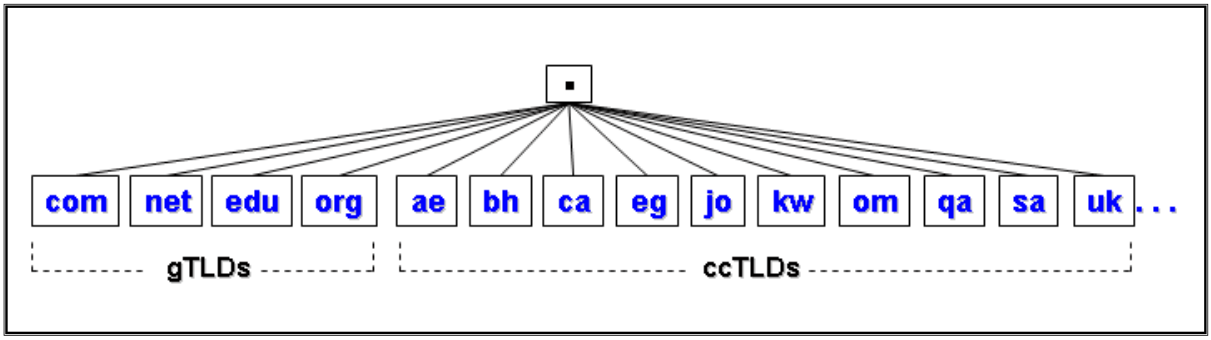


Figure 1: Domain Name Tree Structure

A domain name, whether under a gTLD or ccTLD offers a global presence which makes sure that the corresponding web site is accessible through the Internet from anywhere. More than 170 millions of such names are estimated to be already stored in the Internet domain name system [17].

Part of supporting the Arabic language in domain names is defining the Arabic domain name tree structure. This means that the Arabic Internet community should produce a set of agreed upon Arabic gTLDs and ccTLDs.

4.1. Evaluations of Existing Suggested Arabic TLDs

There have been some suggestions with respect to the Arabic gTLDs and ccTLDs. They were studied and compared in [18]. The following subsections will summarize what has been found in this study.

4.1.1. Existing Suggested Arabic gTLDs

There are three main suggestions for Arabic gTLDs proposed by vendors and researchers. Table (II) lists these suggestions with the corresponding English gTLDs. These suggestions are:

1. Single Letter (SL):
   A single letter is used as an Arabic gTLD. For example, the Arabic letter "ش" is used for the Arabic gTLD corresponding to the English gTLD ".com" and the letter "م" is used for the gTLD corresponding to the English gTLD ".org".

2. Word Root (WR):
   The root of the Arabic word corresponding to an English gTLD is used for the Arabic gTLD. For example, the root ("شرك") of the Arabic word "شركة" is corresponding to the English gTLD ".com" and the root ("نظم") of the Arabic word "منظمة" is corresponding to the English gTLD ".org".

3. Full Word (FW):
   A full Arabic word that corresponds to an English gTLD is used. For example, the Arabic word "شركة" is used for the Arabic gTLD corresponding to the English gTLD ".com" and the Arabic word "منظمة" is used for the Arabic gTLD corresponding to the English gTLD ".org".

Table II: Proposed Arabic gTLDs

English gTLDs	SL	FW	WR
com	ش	شركة	شرك
net	ك	شبكة	شبك
edu	ت	تعليمي	علم
gov	ح	حكومي	حكم
org	م	منظمة	نظم
mil	ع	عسكري	عسكر
int	د	دولي	دول
aero	ط	طيران	طير
coop	ن	تعاون	عون
name	س	شخص	وسم
biz	ل	أعمال	عمل
info	خ	إعلام	خبر أو طلع
Museum	ف	متحف	تحف

4.1.2. Existing Suggested Arabic ccTLDs

There are four main suggestions for Arabic ccTLDs discussed by the Arab Internet community. Table (III) lists these suggestions for all members of the Arabic League. These suggestions are:

1. **Short Form:**
   This suggestion proposes the use of the short forms of country names based on the Arab standard specifications No. 642-1985, "Codes for names of Countries and Languages" [19].
2. **Short Form without (ال):**
   This is similar to suggestion (1) but without "ال" for the purpose to shorten the length of the Arabic ccTLDs.
3. **Nationality:**
   This suggestion proposes the use of the nationality descriptive of each country.
4. **Country Code:**
   This suggestion proposes the use of the 2-letter country codes based on the Arab standard specifications No. 642-1985, "Codes for names of Countries and Languages" [19], which is the equivalent to the ISO 3166 standard [16].

Table III: Existing Arabic ccTLDs

Country Official Names	Country Code		Nationality	Short Name	Short Name without (ال)
	English	**Arabic**			
Hashemite Kingdom of Jordan	**jo**	ار	أردني	الأردن	أردن
United Arab Emirates	**ae**	ام	إماراتي	الإمارات	إمارات
Kingdom of Bahrain	**bh**	بح	بحريني	البحرين	بحرين
Republic of Tunisia	**tn**	تو	تونسي	تونس	تونس
People's Democratic Republic of Algeria	**dz**	جز	جزائري	الجزائر	جزائر
Federal and Islamic Republic of Comoros	**km**	قم	قمري	جزر القمر	جزر القمر
Republic of Djibouti	**dj**	جي	جيبوتي	جيبوتي	جيبوتي
Kingdom of Saudi Arabia	**sa**	سع	سعودي	السعودية	سعودية
Democratic Republic of Sudan	**sd**	سد	سوداني	السودان	سودان
Syria Arab Republic	**sy**	سر	سوري	سورية	سورية
Somalia Democratic Republic	**so**	صو	صومالي	الصومال	صومال
Republic of Iraq	**iq**	عر	عراقي	العراق	عراق
Sultanate of Oman	**om**	عم	عماني	عمان	عمان
Palestine	**ps**	فل	فلسطيني	فلسطين	فلسطين
State of Qatar	**qa**	قط	قطري	قطر	قطر
Stat of Kuwait	**kw**	كو	كويتي	الكويت	كويت
Lebanese Republic	**lb**	لب	لبناني	لبنان	لبنان
Socialist People's Libyan Arab Jamahiriya	**ly**	لي	ليبي	ليبيا	ليبيا
Arab Republic of Egypt	**eg**	مص	مصري	مصر	مصر
Kingdom of Morocco	**ma**	مغ	مغربي	المغرب	مغرب
Islamic Republic of Mauritania	**mr**	مو	موريتاني	موريتانيا	موريتانيا
Yemen Arab Republic	**ye**	يم	يمني	اليمن	يمن

4.1.3. Results of the Study

These suggested Arabic gTLDs and ccTLDs were studied and compared using the following criteria [18]:

1. Length of the Arabic TLDs.
2. Coherence and Clarity of the Arabic TLDs, .i.e., how much easy to associate an Arabic TLD with its corresponding category.
3. Linguistic structure of the Arabic domain name, i.e., consistency with the Arabic language.
4. Easiness of pronunciation of the Arabic TLDs.
5. Future expansion of more Arabic TLDs (i.e., extendibility).

Additional two factors were used, for comparing Arabic ccTLDs, namely:

6. Undesirable code for Arabic ccTLDs. This factor indicates whether the proposed Arabic ccTLDs introduce undesirable Arabic words. This is particularly if arbitrary combination of Arabic characters is used to produce an Arabic ccTLD.
7. The easy of reaching consensus among the involved parties (e.g., Arab countries).

Here are some major points from the study:

- With respect to the length of the proposed Arabic gTLDs and ccTLDs it is obvious that using single letter for a gTLD and 2-letter country code for ccTLD gives the best score.
- Arabic gTLDs that are based on full words are much clearer than Arabic gTLDs that are based only on single letters. For example, the full Arabic word "شبكة" is much clearer than just a single letter "ك" for representing the category "network".
- Most of the proposed Arabic gTLDs are basically based on one-to-one translation of the English gTLDs either using full words or single letters, which might not suite the Arabic language. This means that the name of the entity comes before its category (i.e., gTLD). For example, the domain name of "the food and agriculture organization (fao)" is "fao.org"; the actual name and the domain name both have the entity type (organization/org) at the end. However, for the proposed Arabic gTLDs particularly the use of full words presents unpleasant structure of a domain name from linguistic point of view. For example, Table (IV) depicts examples of some Arabic domain names in which the Arabic linguistic structure of domain names is sound awkward.

Table IV: Example of some Arabic Domain Names

Entity name	***Type***	***Arabic domain name***
اتحاد السباحة العربي	org (منظمة)	اتحاد-السباحة-العربي.منظمة
شركة الأمانة	com (شركة)	الأمانة.شركة
الشبكة العربية للمعلومات	net (شبكة)	عربية.شبكة

- Pronouncing single letters or full words is much easier than word roots.
- Using single letters is limited to 28 possible Arabic gTLDs because there are only 28 letters in the Arabic character set. Using word roots has little limitation in which multiple words may give the same root, e.g., the following words: "تعليم" and "إعلام" have the same root "علم". While using full words virtually has no limitation.
- It has been found that using Arabic full words for gTLDs is more suitable for the Arabic language. This is because they are straightforward to relate to the corresponding categories and easy to pronounce.
- With respect to the proposed Arabic ccTLDs, it has been found that using the Arabic standard for country codes would be the easiest and fastest way to reach agreements among the

concerned parties. Also, it is quite comprehensive which include most of the world's countries. This is said despite its shortcoming of having some unpleasant codes. Enhancement to the standard to overcome this problem can be done for certain country codes to become three letters rather than two letters. For example, "مصر" and "قطر". However, from pure linguistic point of view, the use of nationality (the second proposal) would be recommended.

4.2. Open Issues

Most of the proposed Arabic gTLDs are based on one-to-one translation of the English gTLDs either using full words or single letters. They are based on the activities of the entities (i.e., commercial, government, education) which might not suite the Arabic language. This is due to the fact that the entity category in Arabic (i.e., شركة، منظمة ...) comes in front of the entity name and not at the end. Hence, this leads to creating awkward Arabic domain names.

Therefore, this section will list three major issues related to the selection of the proper Arabic TLDs.

4.2.1. Full words vs. Abbreviations

In general, abbreviations are not widely used in Arabic language. Even if they are used they are pronounced in full wording. For example, the Arabic abbreviation (ص.ب) is used but its pronounced (صندوق بريد), also the Arabic abbreviation (أهـ) is used but its pronounced (انتهى كلامه). Thus, the use of full words in domain names will be more acceptable than abbreviations. This applicable to both Arabic gTLDs and ccTLDs.

4.2.2. The Arabic Equivalent of "WWW"

Do we really need an Arabic equivalent to the English prefix "www" in domain names? Some suggested using the Arabic word "موقع". However, knowing that the inventor of the "www" has regretted the introduction of this prefix, put us at an advantage not to make the same mistake. Therefore, there should be no need for a specific prefix to identify web locations in the Internet. Domain names without prefixes will be sufficient to locate these sites.

4.2.3. Arabic TLD Categories

The activity type (e.g., commercial, government, network, …) is used in defining English gTLDs (e.g., .com, .gov,. .net, …). This classification might not be the right choice for the Arabic gTLDs. This is due to the fact that activity type in Arabic (i.e., شركة، منظمة ...) comes in front of the entity name and not at the end. Hence, this leads to creating awkward Arabic domain names.

Geographical classification is widely used in Arabic language for indenting people, entities, and products. The geographical descriptive words usually come at the end of the phrase. Table (V) lists some examples.

Table V: Examples of Arabic Names with Geographical Descriptive Words

أبو إسماعيل البخاري	جمعية الحاسبات السورية	حمص حلبي
قناة السويس	حزب العمل اللبناني	مشمش شامي
اتحاد كرة الطائرة السعودي	نادي الأهلي المصري	شاه ي مغربي
فرس عربي	دينار كويتي	جامعة الدول العربية

4.2.4. Arabic Domain Name Structure

A domain name consists of multiple words (codes) that are separated by dots. The following could be the proposed Arabic domain name structure based on the assumption that the geographical classification is adopted:

<A-TLD>.<entity-name>

Where, <entity-name> represents the Arabic name of the entity, and <A-TLD> represents an Arabic TLD. For example,

شركة-الزومان.سعودية
شركة-أرامكو.سعودية
المركز-التجاري.سورية
اتحاد-كرة-الطائرة.عربي
جامعة-الخرطوم.سودان

4.3. Recommended Arabic gTLDs and ccTLDs

Based on the discussions in the previous sections, the current suggested Arabic gTLDs which use the entity type for the classification are not suitable for the Arabic language. Therefore, it is suggested to use the geographical classification instead.

With respect to Arabic gTLDs, it is recommended to use geographical descriptive words such as "دولي" and "عربي". In later phase (if needed) other Arabic gTLDs can be added which represents activities such as "تعليمي" "تجاري", and "معلوماتي".

With respect to Arabic ccTLDs, it is recommended to use the standard short country names (e.g., السعودية) or code (e.g., سع) as listed in [19]. See table VI.

Table VI: Proposed Arabic ccTLDs

Country Official Names	Country Code		Short Name
	English	**Arabic**	
Hashemite Kingdom of Jordan	**jo**	ار	الأردن
United Arab Emirates	**ae**	ام	الإمارات
Kingdom of Bahrain	**bh**	بح	البحرين
Republic of Tunisia	**tn**	تو	تونس
People's Democratic Republic of Algeria	**dz**	جز	الجزائر
Federal and Islamic Republic of Comoros	**km**	قم	جزر القمر
Republic of Djibouti	**dj**	جي	جيبوتي
Kingdom of Saudi Arabia	**sa**	سع	السعودية
Democratic Republic of Sudan	**sd**	سد	السودان
Syria Arab Republic	**sy**	سر	سورية
Somalia Democratic Republic	**so**	صو	الصومال
Republic of Iraq	**iq**	عر	العراق
Sultanate of Oman	**om**	عم	عمان

Palestine	**ps**	فل	فلسطين
State of Qatar	**qa**	قط	قطر
Stat of Kuwait	**kw**	كو	الكويت
Lebanese Republic	**lb**	لب	لبنان
Socialist People's Libyan Arab Jamahiriya	**ly**	لي	ليبيا
Arab Republic of Egypt	**eg**	مص	مصر
Kingdom of Morocco	**ma**	مغ	المغرب
Islamic Republic of Mauritania	**mr**	مو	موريتاني
Yemen Arab Republic	**ye**	يم	اليمن

Figure (2) show the suggested Arabic domain name tree structure whether using country codes or nationalities for the Arabic ccTLDs.

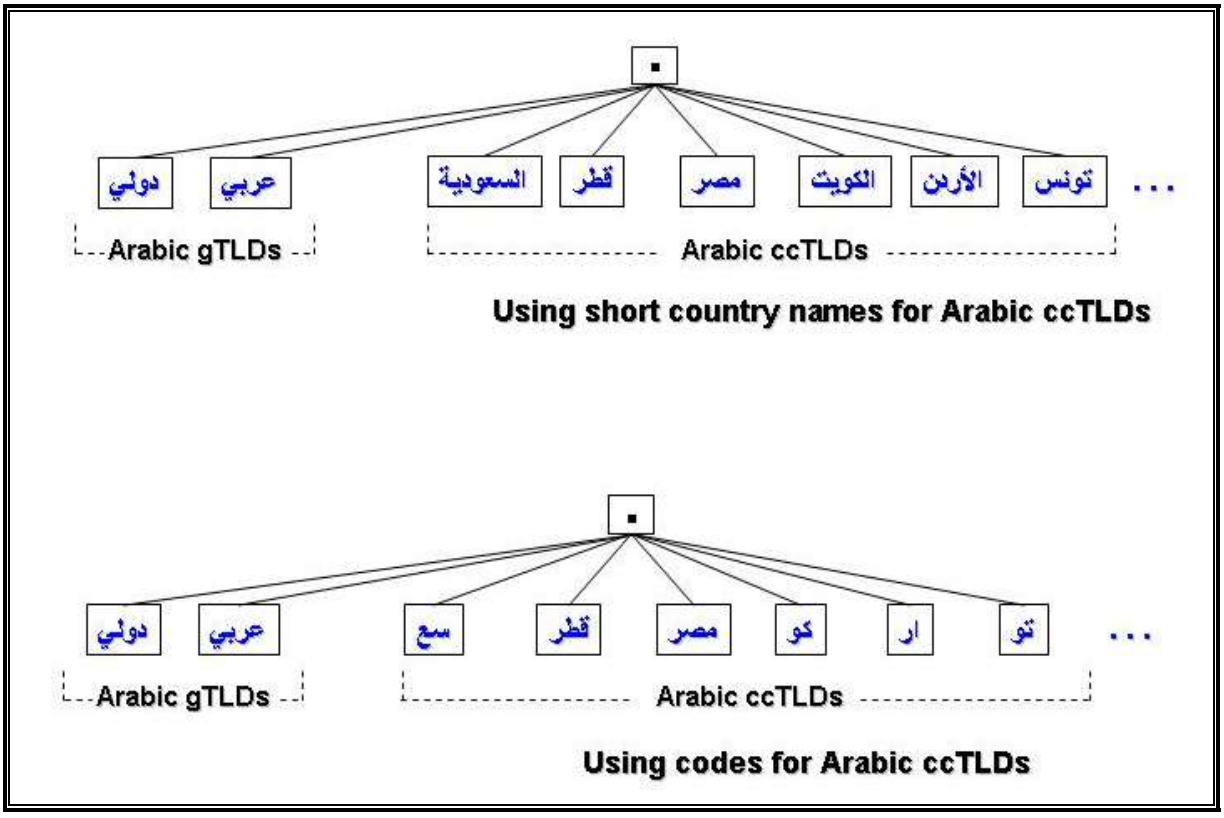


Figure 2: Arabic Domain Name Tree Structure

5. A PILOT PROJECT: IMPLEMENTING GCC ARABIC DOMAIN NAMES

Since there is no indication that ICANN is going to support full IDN in the near future, the managers of the GCC (Gulf Cooperation Council) ccTLDs (i.e., ae, bh, kw, om, qa, sa) in their meeting on 7th of March 2004 agreed to initiate a pilot project whose mission is:

> "To implement a test bed for Arabic Domain names in the GCC countries. This will allow all GCC countries to early experience the use of Arabic domain names, identify our needs, locate possible problems, and develop some tools."

The objectives of the pilot project are:

- To gain experience and knowledge of the Arabic Domain names and share it with Arab countries.
- Test the implantations of Arabic Domain names (see Figure ()).
- Build the local awareness about Arabic Domain names.
- Establish joint work with other entities (i.e., ISPs, universities …).
- Possibly develop some tools related to Arabic domain names and DNS.

The project will be done in three phases:

Phase 1: Testing ADNS:

- Setting up Arabic GCC root servers (3 Weeks).
- Resolving Arabic GCC domain names (1 Week).
- Testing other DNS software and browsers (2 Weeks).
- Writing technical documents about the gained experience (2 Weeks):
  - Setting up Arabic GCC Root server.
  - Reaching Arabic GCC ccTLD Domain names.

Phase 2: Developing policies and regulations:

- Studying the current available polices for domain registration from ICANN & WIPO.
- Defining our special needs.
- Writing policies and regulations for registering Arabic domains.

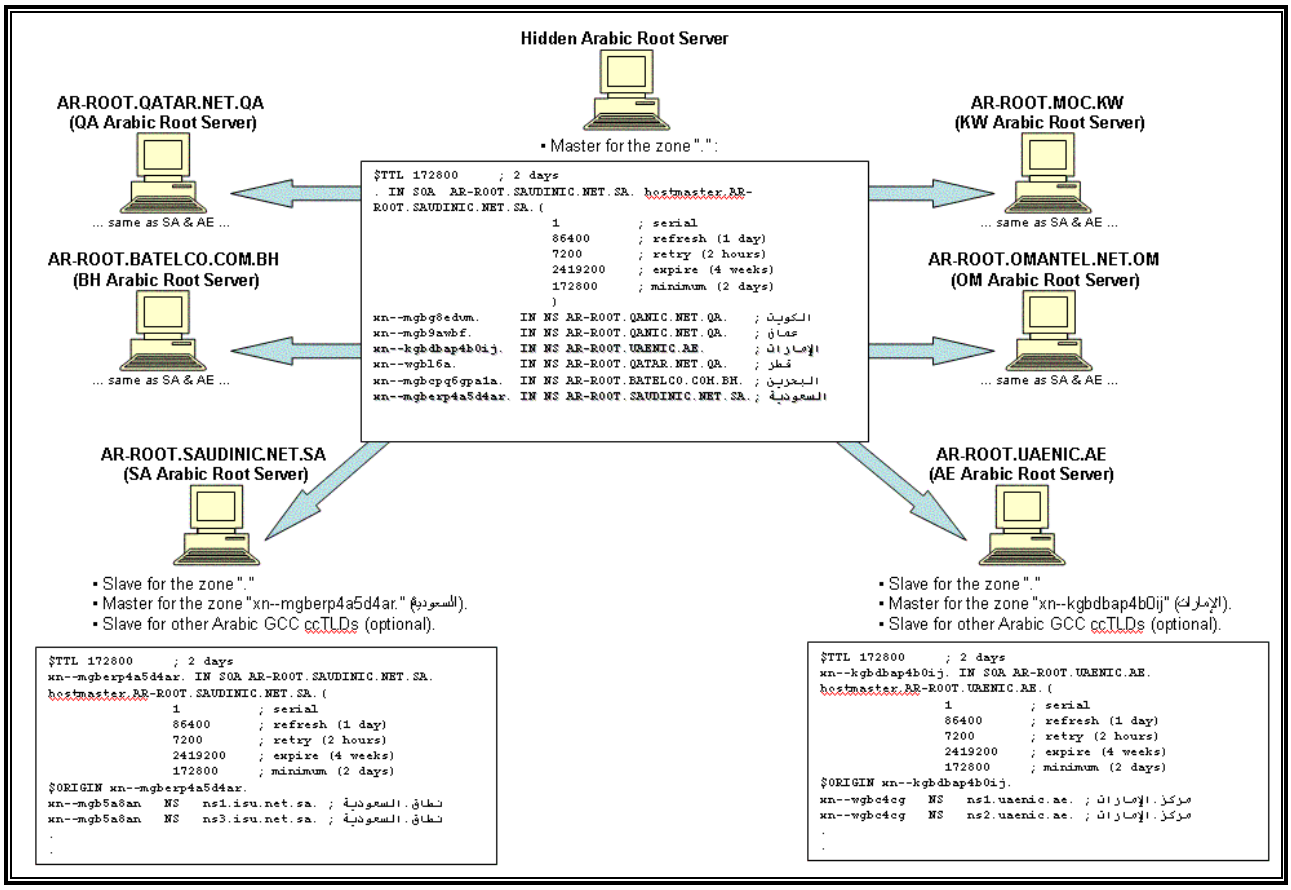
Phase 3: Public awareness:

- Build a website for the project and publish some tools and useful documents in it.
- Encourage other Arab countries and entities to participate in this project.
- Registering some test Arabic Domain names.

The project is expect to deliver the following:

- Prepare the GCC root servers.
- Test the GCC Arabic Domain names.
- Draft Technical guidelines.
- Define policies and regulations for registering Arabic domain names.
- Prepare a website for the project.
- Public Seminar of the Implementation Project.

For more information, visit the following web site: http://www.arabic-domains.org.sa



Source: Raed Al-Fayez, Pilot Project - Implementing GCC Arabic Domain Names (14-Mar-2004)

Figure 3: GCC Arabic Domain Name Pilot Project

6. CONCLUSIONS

This report highlighted the experiences of SaudiNIC in the field of Arabizing domain names. It listed and discussed most of the Arabic linguistic issues related to Arabic domain names, and it gave the recommended Arabic character set based on Unicode version 3.1.

The report also studied and compared a number of existing suggestions for both Arabic gTLDs and Arabic ccTLDs. With respect to the Arabic gTLDs, there are three proposals: single letters, word roots, and full words. It has been found that using Arabic full words for gTLDs is more suitable for the Arabic language. This is because they are easy to relate to the corresponding categories and to pronounce. However, they are too long.

However, most of the existing Arabic gTLDs are basically based on one-to-one translation to the English gTLDs either using full words or single letters, which might not suite the Arabic language. This is due to the fact that entity category in Arabic (i.e., شركة، منظمة ...) comes in front of the entity name and not at the end. Hence, this leads to creating awkward Arabic domain names.

Therefore, it is recommended to use geographical descriptive words such as "دولي" and "عربي", and in later stage other descriptive words such as activities, e.g., "تعليمي" "تجاري", and "معلوماتي". These

recommendations are based on the belief that technologies have to serve the language and not the other way.

With respect to the proposed Arabic ccTLDs, it has been found that using the Arabic standard for country short names or codes (ASMO, Arab Standard Specifications, No. 642-1985) would be the easiest and fastest way to reach agreements among the concerned parties. The standard is quite comprehensive which include most of the world's countries but is old and inundated. This is said despite its shortcoming of having some unpleasant codes. Enhancement to the standard to overcome this problem can be done for certain country codes to become three letters rather than two letters. For example, "مصر" and "قطر". However, from pure linguistic point of view, the use of country short names would be recommended.

Now we need more support from ICANN/ITU to get the Arabic domain names part of the Standard domain name system. We do not want alternative Arabic root servers. We want the issues of administrating and managing the "Arabic root servers" be resolved in an acceptable time frame.

ACKNOWLEDGEMENT

It is hard to acknowledge every individual who has participated (directly or indirectly) in the final results of this document. However, I would try to acknowledge them through the groups or organizations that they are representing or participating in. Therefore, I would thanks King Abdulaziz City for Science and Technology, all GCC ccTLDs (AE-NIC, BH-NIC, KW-NIC, OM-NIC, QA-NIC, SA-NIC), AINC, MINC, ADNTF of UN-ESCWA, and Arabic Linguistic Committee of AINC. I hope I have not forgotten anyone. If this is the case, please forgive me!

REFERENCES

[1] WIPO, "Internationalized Domain Names – Intellectual Property Considerations", Multilingual domain names: Joint ITU/WIPO Symposium, Geneva, Dec. 6-7, 2001.

[2] ITU, "Technology and Policy Aspects", Multilingual domain names: Joint ITU/WIPO Symposium, Geneva, Dec. 6-7, 2001.

[3] Paul Albitz and Cricket Liu, "DNS and BIND", 4th Ed., O'Reilly & Associates, Inc., USA, 2001.

[4] Paul Mockapetris, "Domain names, Concepts and facilities", STD 13, RFC 1034, November 1987.

[5] Paul Mockapetris, "Domain Names - Implementation and Specification", STD 13, RFC 1035, November 1987.

[6] Ajeeb.com, "More than 3.5 million Internet users in the Arab world", 25 March, 2001 [in Arabic], , URL: http://it.ajeeb.com/viewarticle.asp?article=1662&category=34

[7] A. Al-Zoman, "Using Arabic Language in writing domain names", Arab journal of library and information science, Vol 22, No. 3, July 2002, pp. 21-38 [in Arabic].

[8] Lee D. and others, "Extensions to DNS for Supporting Internationalized Domain Names", LECTURE NOTES IN COMPUTER SCIENCE, Issue 2344, 2002, Pages 791-801.

[9] Shi H.; Sato I.; Goto S., "Extension of DNS to the internationalized domain names", IEICE Transactions on Information and Systems, Volume E84-D, Issue 5, 2001, Pages 588-595.

[10] Faltstrom, P., Hoffman, P. and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.

[11] Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", RFC 3491, March 2003.

[12] A. Costello, "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003.

[13] ICANN, "Internationalized Domain Names (IDN) Committee", 8 Nov. 2002, URL: http://www.icann.org/committees/idn/

[14] A. Al-Zoman, "AINC-ALC Chairman Report: April 2001-April 2002", Technical Report, http://www.saudinic.net.sa/ar/arabicdomain/status-report-april2002.doc [in Arabic].

[15] The Unicode Consortium, "The Unicode Standard", http://www.unicode.org/charts/

[16] ISO, "ISO 3166 code lists – English country names and code elements", URL: http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html

[17] Internet Software Consortium, "Internet Domain Survey", Jan 2003, URL: http://www.isc.org/ds/

[18] A. Al-Zoman, "Arabic Top-Level Domains", Paper presented in Expert Group Meeting on Promotion of Digital Arabic Content, the United Nations, Economic and Social Commission for Western Asia, Beirut, 3-5 June 2003.

[19] ASMO, "Arab Standard Specifications, No. 642-1985: Codes for names of Countries and Languages", Arab League, Arab Org. for Standardization and Metrology, 1985.

[20] A. Al-Zoman, "Arabic Top-Level Domain Names", International Journal of Computer Processing of Oriental Languages, Volume 17 Number 3 September 2004, To Appear.

[21] A. Al-Zoman, "Linguistic Issues in Arabic Domain Names", In Proceedings of the 17th National Conference on Computers, King Abdulaziz University, Al-Madina Almunawarah, Saudi Arabia, 5-8 April, 2004, pp 235-250 [in Arabic]

[22] A. Al-Zoman, "Arabic Top-Level Domain Names", In Proceedings of the 17th National Conference on Computers, King Abdulaziz University, Al-Madina Almunawarah, Saudi Arabia, 5-8 April, 2004, pp 281-296 [in Arabic]

[23] A. Al-Zoman, " Using Arabic Language in writing domain names ", In Proceedings of the International Arab Conference on Information Technology 2001, Jordan University for Science and Technology, Irbid, Jordan, 13-15 Nov., 2001, pp 264-272[in Arabic]