

ARABIC TOP-LEVEL DOMAIN NAMES

Abdulaziz H. Al-Zoman

Associate Professor

Internet Services Unit

King Abdulaziz City for Science and Technology

P.O. Box 6086, Riyadh 11442

Saudi Arabia

zoman@isu.net.sa

Abstract

In the Internet, domain names (i.e., web site addresses) are still written using English characters regardless of the worldwide spread of the Internet. Currently, domain names do not support other languages to locate resources on the Internet. Users in non-English speaking countries, such as the Arab users, are in disadvantages.

Multilingual domain names have been the goal and activities of many non-for profit organizations that supervise and pursue their deployments. Several companies have begun to commercialize their technologies.

With respect to the Arabic language, however, current implementations are using proprietary technologies. Therefore, it is important to establish standardized Arabic solutions including selecting the Arabic generic and country code top-level domain names (Arabic gTLD and ccTLD). This paper compares a number of proposed suggestions for Arabic gTLDs and ccTLDs.

Keywords

Domain names, Arabization, Arabizing domain names, Internationalize domain names, IDNS, Internet applications.

1. Introduction

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.

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 which domain names are mapped to the actual corresponding IP addresses.

The DNS is basically a distributed database of host information that is organized in a hierarchal tree structure [1], 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 two types of TLDs [2,3]. 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 [4].

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 (DNS) [5].

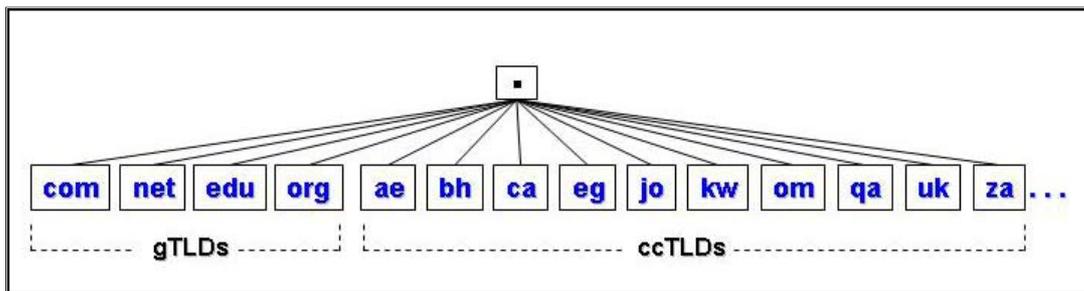


Figure 1: Domain Name Tree Structure

The Internet has become a global network of most if not all countries of the words with hundred of millions of users. Currently, 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 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 [6,7,8].

Regardless of the worldwide spread of the Internet, the Internet domain name system has not supported other languages to locate resources on the Internet. Users in non-English speaking countries, such as the Arab users, are in disadvantages. 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.

The Internet penetration in the Arab world is estimated to be 1.67 % and it is expected to be around 6.41 % by end of 2005 [9], which is indeed very low. One of the obstacles facing the growth of using Internet in the Arab world is the language barrier. Thus, many countries and nations are encouraging their people to use Internet. Therefore, it is important to make the Internet support the Arabic language not only in web contents but also in their addresses.

Multilingual domain names were first developed in Asia-Pacific countries in 1998 [7,8,10,11], which led later to the creation of a number of non-for profit organizations to supervise and pursuing the deployment of multilingual domain names. Among these organizations are: the

Multilingual Internet Names Consortium (MINC), the Arabic Internet Names Consortium (AINC), 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 3 RFCs for Internationalized DNS [12,13,14].

Several companies have begun to commercialize the technology that had been developed. However, as usual, vendors are faster than the standardizing bodies for proposing solutions for multilingual domain names which also support Arabic domain names. However, current implementations of multilingual domain names are using proprietary technologies. ICANN adopted a resolution [15] which recognizes that “it is important that the Internet evolve 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”.

Therefore, it is required that the Arabic language is being used from the start of switching on the personal computer till getting information from the Internet. Thus, eliminating the need for the users to enter non-Arabic web (URL) addresses particularly if the sites are in Arabic. This led to the need of Arabizing domain names. There are a number of reasons why Arabizing is needed, such as [16]:

- Making the Arabic language part of the Internet.
- Small percentage of Arabs who can read and write in 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 same as English-speaking users.
- English 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 using the user native language.

One of the important aspects of supporting the Arabic language in domain names is to establish standardized generic and country code top-level domain names (Arabic gTLD and ccTLD). However, a number of vendors have proposed some solutions that support multilingual domain names including the Arabic language. Part of their solutions they have proposed some suggestion for the Arabic gTLDs and Arabic ccTLDs. The objective of this paper is to study and compare these suggestions. The study is based on the collected responses (more than 230 responses) from online surveys that were put on the Internet^{*} from May to December 2003.

2. Arabic generic top-level domains (gTLDs)

There are three main suggestions for Arabic gTLDs proposed by vendors and researchers. Table (1) lists these suggestions with the corresponding English gTLDs. These suggestions are:

1. Single Letter (SL):

This suggestion is proposed by the "i-DNS.net International"^{**} in which a single letter is used for 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".

* Two survey forms on the following URLs: <http://www.saudinic.net.sa/survey/tlds-user-survey-en.htm> , <http://www.saudinic.net.sa/survey/cctld-user-survey-en.htm>

** see <http://www.i-dns.net>

2. Word Root (WR):

This suggestion is submitted by Dr. Abu-EL-Haija to the Arabic Internet Names Consortium (AINC) [17] in which the root of the Arabic word corresponding to the 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):

This suggestion is proposed by some of the vendors such as "Native Solutions Corporation"* and "Walid Inc."⁺. This suggestion proposes to use full Arabic words that correspond to the English gTLDs. 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 1: Suggested Arabic gTLDs[^]

Suggestion	com	net	edu	gov	org	mil
SL	ش	ك	ت	ح	م	ع
WR	شرك	شبكة	علم	حكم	نظم	عسكر
FW	شركة	شبكة	تعليم	حكومة	منظمة	عسكري

These suggestions will be compared using the following criteria:

1. Length of the Arabic gTLDs.
2. Clarity of the Arabic gTLDs, .i.e., how much easy to associate an Arabic gTLD with its corresponding category.
3. Linguistic structure of the Arabic domain name.
4. Pronunciation of the Arabic gTLDs.
5. Future expansion of more Arabic gTLDs.

* see <http://www.nativesolutions.com>

+ see <http://www.walid.com>

[^] These are just examples based on some of the proposed suggestions.

A scale of 0 to 5 will be used for comparing these criteria for the suggested proposals, where 0 and 5 represent the lowest and highest mark of the scale, respectively.

2.1. Length of the Arabic gTLDs

This criterion measures the speed of typing the Arabic gTLD based on the number of characters (NC) forming it. The score will be computed using the following formula (Score = 6 – NC); if NC less or equal than 6, otherwise, score will be zero.

This gives a scale from 0 to 5 where 0 means that the Arabic gTLD is too long (i.e., 6 characters or more), while score of 5 means that the Arabic gTLD is in its shortest form, i.e., one character.

Table (2) compares the length of the proposed Arabic gTLDs which indicates that using single letter gives the best score under this criterion.

Table 2: Comparison of Lengths of the Arabic gTLDs

Suggestion	com	net	edu	gov	org	mil	Average
SL	4.43	4.50	4.43	3.86	4.57	4.57	4.39
WR	3.00	3.43	3.07	3.57	3.79	2.57	3.15
FW	2.14	2.29	2.14	2.14	2.07	1.64	2.07

2.2. Clarity of the Arabic gTLDs

This criterion measures how much an Arabic gTLD is clear for representing the category it corresponds to. In other words, how much easy is the process to associate the Arabic gTLD with the corresponding category?

Table (3) indicates that 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".

Table 3: Comparison of Clarity of the Arabic gTLDs

Suggestion	com	net	edu	gov	org	mil	Average
SL	1.43	1.64	1.43	1.86	1.50	1.36	1.54
WR	2.46	3.29	3.93	3.43	3.21	4.00	3.39
FW	4.79	4.50	4.57	4.57	4.57	4.79	4.63

2.3. Linguistic structure of the Arabic domain name

From the linguistic point of view, this factor measures how much the proposed Arabic gTLD goes with the rest of the domain name. A domain name consists of multiple words or phrases separated by dots. Therefore, the question is whether the proposed Arabic gTLD affect the overall linguistic structure of the whole domain name. It should be noted that this does not imply that a domain name must comply with the Arabic grammar but rather it should not sound awkward to the ear of the listener. Table (4) summarizes the results of comparing the proposed Arabic gTLDs with respect to the linguistic structure of an Arabic domain name.

Table 4: Comparison of Linguistic Structure of the Arabic gTLDs

Suggestion	com	net	edu	gov	org	mil	Average
SL	3.07	3.36	3.29	3.00	3.21	3.36	3.21
WR	1.93	2.29	2.21	1.86	2.07	1.71	2.01
FW	2.21	2.86	3.14	3.00	2.71	2.79	2.79

Since most of the proposed structures of the Arabic gTLDs are based on how the English gTLDs structured in which the name of the entity comes before its activity (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 present unpleasant structured of domain name

from linguistic point of view. For example, Table (5) depicts examples of some Arabic domain names in which the Arabic linguistic structure of domain names sounds awkward.

Table 5: 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

2.4. Pronunciation of the Arabic gTLDs

This factor measures how much easy an Arabic gTLD is pronounced (verbally exchanged) between users. Table (6) depicts the results of the comparisons between the proposed Arabic gTLDs. It is clear that pronouncing single letters or full words much easier than word roots.

Table 6: Comparison of Pronunciation of the Arabic gTLDs

Suggestion	com	net	edu	gov	org	mil	Average
SL	4.14	4.00	3.93	4.21	4.21	4.71	4.21
WR	2.29	2.93	3.14	2.86	2.79	3.23	2.87
FW	4.57	4.43	4.36	4.36	4.50	4.71	4.49

2.5. Future expansion of more Arabic gTLDs

This factor measures the abilities of the suggested Arabic gTLDs for future expansions. The first proposal, using single letters, is limited to 28 possible Arabic gTLDs because there are only 28 letters in the Arabic character set. The second proposal, 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 the third proposal, using full words, virtually has no limitation, see Table (7).

Table 7: Comparison of Future Expansion of the Arabic gTLDs

Suggestion	Score
SL	2.23
WR	3.35
FW	4.73

2.6. Comparison Summary of Arabic gTLDs

Table (8) summarizes the comparison results that were presented in the previous sections with the overall average scores. The third proposal, using full words for Arabic gTLDs, gets the highest score. This finding, particularly with respect to the use of full words for Arabic gTL, agrees with the result collected from another Internet-based survey^{*}, see Figure (2).

Table 8: Comparison Summary of Proposed Arabic gTLDs

Suggestion	Length	Clarity	Linguistic structure	Pronunciation	Future expansion	Total Score
SL	4.39	1.54	3.21	4.21	2.23	3.11
WR	3.15	3.39	2.01	2.87	3.35	2.95
FW	2.07	4.63	2.79	4.49	4.73	3.74

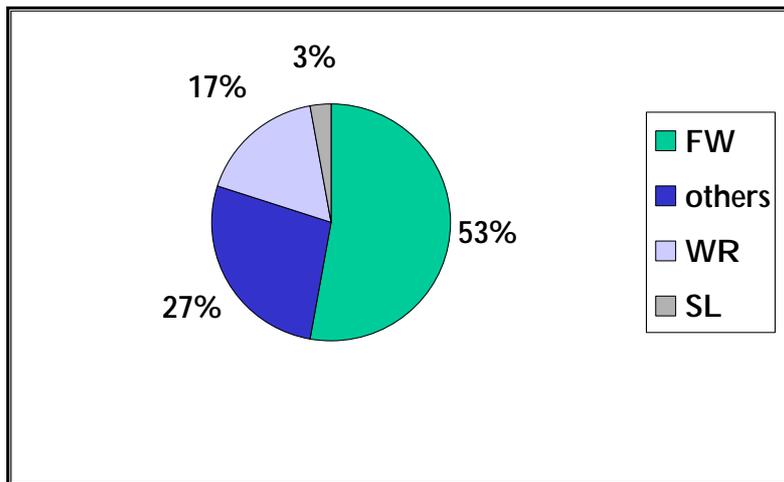


Figure 2: Survey Results for Arabic gTLDs

^{*} See http://www.saudinic.net.sa/arabicdomain/arabic_domains.htm

3. Arabic country code top-level domains (ccTLDs)

Also, there are three main suggestions for Arabic ccTLDs discussed by the Arab Internet community. Table (9) lists these suggestions for all members of the Arabic League. These suggestions are:

1. Short Form (SF):

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" [18].

2. Nationality (N):

This suggestion proposes the use of the nationality descriptive of each country.

3. Country Code (CC):

This suggestion proposes the use of the country codes based on the Arab standard specifications No. 642-1985, "Codes for names of Countries and Languages" [18], which is the equivalent to the ISO 3166 standard.

The five comparison factors used for the proposed Arabic gTLDs in the previous section will be used for the proposed Arabic ccTLDs. Additionally, two other factors also will be used, 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 ease of reaching consensus among the involved parties (e.g., Arab countries).

Table (10) depicts the comparison summary between the proposed Arabic ccTLDs. The third proposal, using country code, gets the highest score. This finding agrees with the result of another Internet-based survey^{*}, see Figure (3).

Table 9: Proposed Arabic ccTLDs

Country Official Names	Country Code		Nationality		Country Short Name	
	Arabic	English	Name	Length	Name	length
Hashemite Kingdom of Jordan	ار	jo	أردني	5	الأردن	6
United Arab Emirates	ام	ae	إماراتي	7	الإمارات	8
Kingdom of Bahrain	بح	bh	بحريني	6	البحرين	7
Republic of Tunisia	تو	tn	تونسي	5	تونس	4
People's Democratic Republic of Algeria	جز	dz	جزائري	6	الجزائر	7
Federal and Islamic Republic of Comoros	قم	km	قمري	4	جزر القمر	9
Republic of Djibouti	حي	dj	جيبوتي	6	جيبوتي	6
Kingdom of Saudi Arabia	سع	sa	سعودي	5	السعودية	8
Democratic Republic of Sudan	سد	sd	سوداني	6	السودان	7
Syria Arab Republic	سر	sy	سوري	4	سورية	5
Somalia Democratic Republic	صو	so	صومالي	6	الصومال	7
Republic of Iraq	عر	iq	عراقي	5	العراق	6
Sultanate of Oman	عم	om	عماني	5	عمان	4
Palestine	فل	ps	فلسطيني	7	فلسطين	6
State of Qatar	قط	qa	قطري	4	قطر	3
Stat of Kuwait	كو	kw	كويتي	5	الكويت	6
Lebanese Republic	لب	lb	لبناني	6	لبنان	5
Socialist People's Libyan Arab Jamahiriya	لي	ly	ليبي	4	ليبيا	5
Arab Republic of Egypt	مص	eg	مصري	4	مصر	3
Kingdom of Morocco	مغ	ma	مغربي	5	المغرب	6
Islamic Republic of Mauritania	مو	mr	موريتاني	8	موريتانيا	9
Yemen Arab Republic	يم	ye	يمني	4	اليمن	5
Average length	2			5.32		6

^{*} See http://www.saudinic.net.sa/arabicdomain/arabic_domains.htm

Table 10: Comparison Summary of Proposed Arabic ccTLDs

Suggestion	Length	Clarity	Linguistic structure	Pronunciation	Future expansion	Undesirable code	Reaching Consensus	Total Score
SF	0.86	3.86	3.43	3.86	3.29	4.29	2.00	3.08
N	1.86	4.43	4.00	4.71	2.86	3.57	2.00	3.35
CC	4.86	3.57	3.14	3.86	3.71	3.14	4.43	3.82

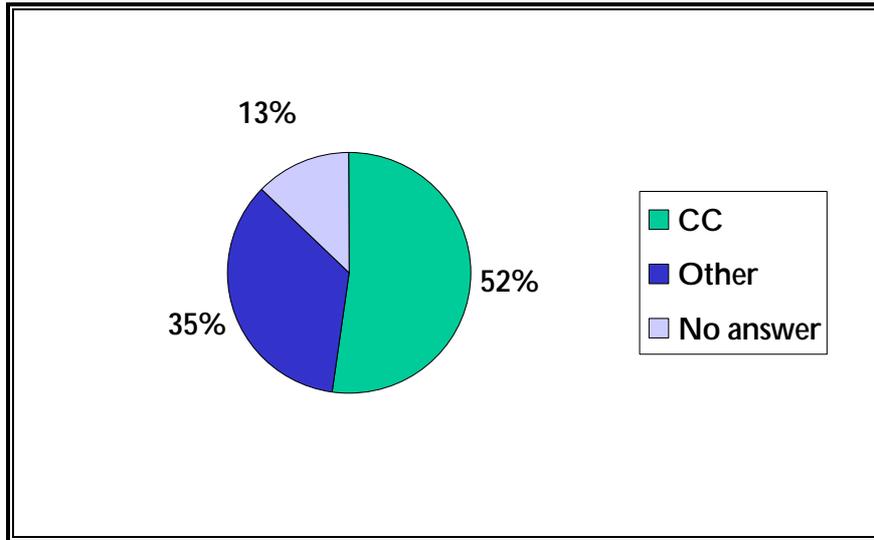


Figure 3: Survey Results for Arabic ccTLD

4. Conclusions

This paper studied and compared a number of proposed 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 proposed 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., *منظمة، شركة، ...*) 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 activities such as "تعليمي", "تجاري", and "معلوماتي". These recommendations are based on the believe 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 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.

Figure (4) show the suggested Arabic domain name tree structure whether using country codes or nationalities for the Arabic ccTLDs.

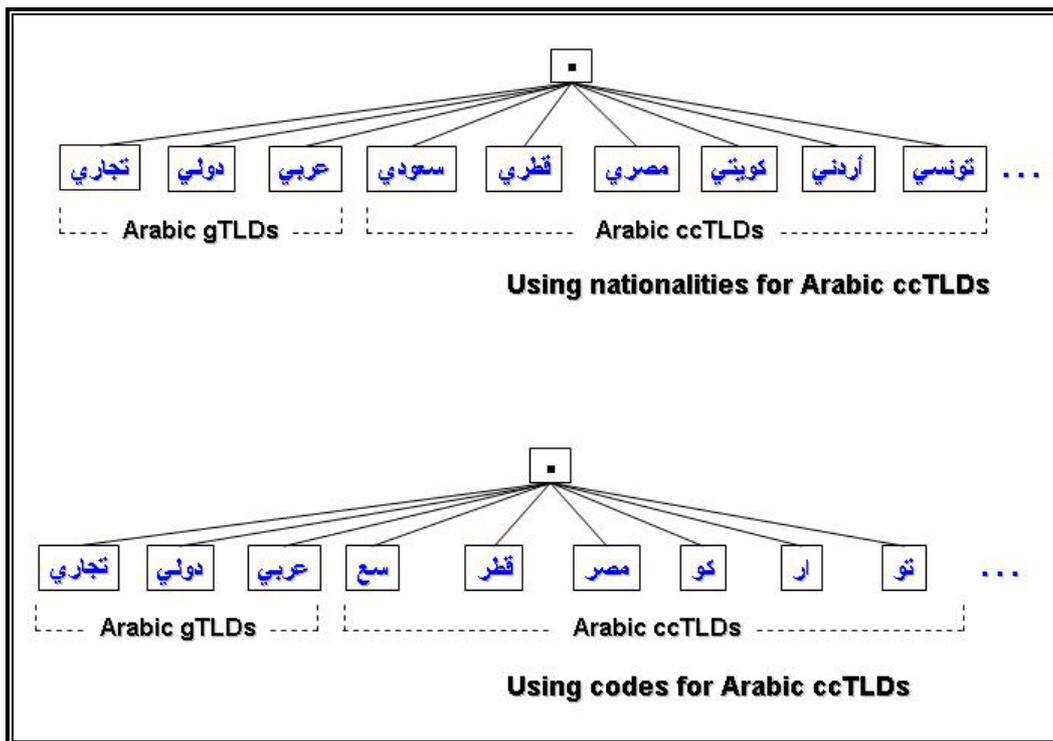


Figure 4: Arabic Domain Name Tree Structure

5. References

- [1] Paul Albitz and Cricket Liu, "DNS and BIND", 4th Ed., O'Reilly & Associates, Inc., USA, 2001.
- [2] J. Postel, "Domain Name System Structure and Delegation", RFC 1591, USC/Information Sciences Institute, March 1994.
- [3] ICANN, "ICP-1: Internet Domain Name System Structure and Delegation (ccTLD Administration and Delegation)", May 1999, URL: <http://www.icann.org/icp/icp-1.htm>
- [4] 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>
- [5] Internet Software Consortium, "Internet Domain Survey", Jan 2003, URL: <http://www.isc.org/ds/>
- [6] Global Reach, "Global Internet Statistics", 30 Sept 2002, URL: <http://www.glreach.com/globstats>

- [7] WIPO, "Internationalized Domain Names – Intellectual Property Considerations", Multilingual domain names: Joint ITU/WIPO Symposium, Geneva, Dec. 6-7, 2001.
- [8] ITU, "Technology and Policy Aspects", Multilingual domain names: Joint ITU/WIPO Symposium, Geneva, Dec. 6-7, 2001.
- [9] Madar Research Journal, "Number of Saudi Internet users to edge towards 4.5 million by end of 2005 ", 20 October, 2002, URL:
<http://www.madarresearch.com/news/newsdetail.aspx?nwsId=1>
- [10] Lee D. and others, "Extensions to DNS for Supporting Internationalized Domain Names", LECTURE NOTES IN COMPUTER SCIENCE, Issue 2344, 2002, Pages 791-801.
- [11] 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.
- [12] Faltstrom, P., Hoffman, P. and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.
- [13] Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", RFC 3491, March 2003.
- [14] A. Costello, "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", RFC 3492, March 2003.
- [15] ICANN, "Internationalized Domain Names (IDN) Committee", 8 Nov. 2002, URL:
<http://www.icann.org/committees/idn/>
- [16] 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].
- [17] Ahmad Abu-EL-Haija, "TOP LEVEL DOMAINS (TLD's), internal report submitted to AINC, 2001.
- [18] ASMO, "Arab Standard Specifications, No. 642-1985: Codes for names of Countries and Languages", Arab League, Arab Org. for Standardization and Metrology, 1985.