

Arabic Domain Names Pilot Project

Implementation Experiences of

idn.idn

Submitted by the Steering Committee
of the Arabic Domain Names Pilot Project
November 2005

Abstract

Domain names are a very crucial part of applied Internet technology. To this day, they remain dependant upon Latin characters regardless of the global reach of the Internet itself. Other languages are not yet fully supported to locate resources and sites on the network. Nevertheless, using Arabic domain names is essential to increase 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. The Arabic Domain Names Pilot Project (ADNPP) has addressed these issues and are in the test-bed implementation phase.

This document highlights the experiences of ADNPP in supporting internationalized (Arabic) domain names (i.e., Arabic.Arabic) and offers some recommendations regarding the implementation of idn.idn.

1. INTRODUCTION

The Internet has become a global network of most, if not all, countries of the world with hundreds of millions of users. Recently, it is estimated that more than 60% of Internet content is in languages other than English. Also, it is estimated that by the year 2003 at least 30% of web users will prefer to perform 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.

Domain names are used widely by Internet users to locate resources on the Internet via 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 through the Internet. Hence, the main objective of using domain names is to ease and simplify the use of the Internet.

Since the Internet originally evolved in the United States, 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 Latin characters particularly letters, digits, and the hyphen character. 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.

Despite 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 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.

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, which is indeed very low. One of the obstacles facing the growth of Internet use in the Arab world is the language barrier. Many countries and nations are encouraging their people to use Internet, therefore it is important to ensure that the Internet supports the Arabic language, not only in web content but also in it's addresses.

Internationalized domain names were first developed in Asia-Pacific countries in 1998, which led later to the creation of a number of nonprofit 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), Arabic Team for Domain Names, Arabic Domain Names Pilot Project , 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 dedicated to exploring the possibility of supporting a truly international Internet. The IDN group of IETF has issued 3 RFCs for Internationalized DNS.

It is required that the Arabic language be used from the start of switching on the user's personal computer till the required data is retrieved from the Internet. This entails the elimination of the need for entering non-Arabic web (URI) addresses particularly if the sites

are in Arabic. There are a number of reasons why Arabizing domain names is needed, such as:

- 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.
- English letters are not capable of representing (or substituting for) Arabic letters.
- Encouraging the use of the Internet by Arabs who do not speak English. As the trend nowadays is towards the implementation of e-government and e-business, therefore it is important to provide the relevant information and services in the user's native language.

Arab countries have recognized the importance of assuring Internet supporting the Arabic language not only in web content but also in it's addresses. Thus, an Arabic Team for Domain Names was created under the auspices of the Arab League in 2004 to coordinate these efforts and in the Arab region. In their 2nd meeting, held in Cairo on the 7th and 9th of May 2005, it was recommended that the GCC Pilot Project for Arabic Domain Names be extended to include all members of the Arab League. Hence, the project was renamed "Arabic Domain Names Pilot Project," and it was determined that henceforth all related efforts will be under the auspices of League of Arab States (www.arableagueonline.org).

Several companies have begun to commercialize the technologies that have been developed to support multilingual domain names. These developments lack a standardized reference. This is because, as usual, vendors are faster than the standards bodies in proposing solutions. Therefore, current implementations of multilingual domain names implement 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". Hence, adopting proprietary solutions may lead to:

- Non-recognition by international bodies such as ICANN and IETF.
- Incompatible technical and linguistic solutions.
- Multiple registrations for the same category.
- Disjoint networks each with its own an Arabic domain name space.

Therefore, it is urgently required from the local and international Internet community to produce a set of standards that are acceptable by the Internet community at 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.
These 2 points have been addresses by the "Guidelines for an Arabic Domain Name System" that was an Internet draft and now is with the Arab League for final approval.
3. Technical solutions to Arabize the domain name system.
This is partially addressed by the IETF RFCs.
4. The administrative and organizational issues of Arabic root servers.
This is ICANN territory.

2. ARABIC DOMAIN NAMES PILOT PROJECT (ADNPP)

Since there is no indication that ICANN is going to support full IDN (i.e., idn.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 for Arabic domain names.

The success of the pilot project led the Arabic Team for Domain Names, in their 2nd meeting that was held in Cairo on the 7th and 9th of May 2005, to recommend the expansion of the GCC Pilot Project for Arabic Domain Names to include all members of the Arab League (22 countries). Hence, the project was renamed as follows: "Arabic Domain Names Pilot Project" and it falls under the auspices of the Arab League.

Two committees have been created for the management and operation of the project: A Steering Committee and a Technical Committee. The Steering Committee's tasks include: General supervision of the project, management supervision of the Arabic root servers, and setting policies and procedures which include participation policies and use terms and conditions. While the Technical Committee's tasks include: providing technical support for participants and users, technical coordination between participants, technical supervision of the Arabic root servers, and enhancing and improving the project from the technical standpoint.

The mission of the project is:

"Implementing a test bed for Arabic domain names (ADN) in the Arab world. This will allow for the early experience the use of Arabic domain names by all Arab countries, the identification of their needs, the agreement upon uniform standards, the identification of possible problems, and the development of required tools and policies."

The project is expected to contribute to the following strategic objectives:

1. To establish and implement Arabic domain names.
2. To increase Internet use in the Arab world by addressing linguistic barriers facing Arabic-speaking users.
3. To promote the use of Arabic language and to increase the Arabic content on the Internet.
4. To promote Arab cultural identity on the Internet.

While the main objectives of the project are:

1. To make the Internet easier to use for native Arabic speakers.
2. To gain experience and knowledge in the use of Arabic domain names and share it with the Internet community.
3. To test the implantations of Arabic domain names based upon the guidelines drafted by the "Arabic Team for Domain Names".
4. To build local awareness related to Arabic domain names.
5. Possibly, to develop necessary tools required for Arabic domain names and DNS.
6. To develop required policies and guidelines that help achieve the above objectives.

The major achievements of the project so far are:

1. Draft a number of policy documents:
 - a. Project Initiation Document
 - b. Participation Policy for Arabic ccTLD managers
 - c. Terms and Conditions
 - d. Guidelines for Writing Arabic Domain Names
2. Technical work:
 - a. Building a website and mailing list for the project.
 - b. Adding some tools and forms.
 - c. Preparing some technical documents.
 - d. Adding new participants to the project (ccTLD).
3. There are at the current time 6 participants:
 - a. Saudi Arabia
 - b. United Arab Emirates
 - c. State of Qatar
 - d. Egypt
 - e. Tunisia
 - f. Palestine
 - g. Syria (soon to join)

3. LINGUISTIC ISSUES

An Internet draft was prepared by a task force operating under the auspices of the UN Economic Commission for Western Asia (ESCWA). Then the draft was reviewed by the Arabic Team for Domain Names and submitted to the Arab League for final approval.

The main parts of the document are the accepted Arabic character set and the Arabic Top-Level domains. The following tables summarize the main recommendations of the document. Also the document covered some management issues. The ADNPP has agreed to follow this document throughout the whole project, leaving sufficient space for each ccTLD manager to draft their own guidelines for registering Arabic domain names which should not result in any conflicts with the drafted document.

Table 1: Characters from Unicode Arabic Table (0600—06ff)

Unicode	Character Name	Unicode	Character Name
0621	Arabic letter hamza	0638	Arabic letter zah
0622	Arabic letter alef with madda above	0639	Arabic letter ain
0623	Arabic letter alef with hamza above	063A	Arabic letter ghain
0624	Arabic letter waw with hamza above	0641	Arabic letter feh
0625	Arabic letter alef with hamza below	0642	Arabic letter qaf
0626	Arabic letter yeh with hamza above	0643	Arabic letter kaf
0627	Arabic letter alef	0644	Arabic letter lam
0628	Arabic letter beh	0645	Arabic letter meem
0629	Arabic letter teh marbuta	0646	Arabic letter noon
062A	Arabic letter teh	0647	Arabic letter heh
062B	Arabic letter theh	0648	Arabic letter waw
062C	Arabic letter jeem	0649	Arabic letter alef maksura

062D	Arabic letter hah	064A	Arabic letter yeh
062E	Arabic letter khah	0660	Arabic-indic digit zero
062F	Arabic letter dal	0661	Arabic-indic digit one
0630	Arabic letter thal	0662	Arabic-indic digit two
0631	Arabic letter reh	0663	Arabic-indic digit three
0632	Arabic letter zain	0664	Arabic-indic digit four
0633	Arabic letter seen	0665	Arabic-indic digit five
0634	Arabic letter sheen	0666	Arabic-indic digit six
0635	Arabic letter sad	0667	Arabic-indic digit seven
0636	Arabic letter dad	0668	Arabic-indic digit eight
0637	Arabic letter tah	0669	Arabic-indic digit nine

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)

Table 2: Arabic ccTLD for Members of Arab League

Country Official Names	Short Name (Arabic)	Unicode	Puny-Code
Hashemite Kingdom of Jordan	الأردن	u+0627 u+0644 u+0623 u+0631 u+062F u+0646	xn--igbhzh7gpa
United Arab Emirates	الإمارات	u+0627 u+0644 u+0625 u+0645 u+0627 u+0631 u+0627 u+062A	xn--kgdbap4b0ij
Kingdom of Bahrain	البحرين	u+0627 u+0644 u+0628 u+062D u+0631 u+064A u+0646	xn--mgbcpq6gpa1a
Republic of Tunisia	تونس	u+062A u+0648 u+0646 u+0633	xn--pgbs0dh
People's Democratic Republic of Algeria	الجزائر	u+0627 u+0644 u+062C u+0632 u+0627 u+0626 u+0631	xn--lgbbat1ad8j
Federal and Islamic Republic of Comoros	القمر	u+0627 u+0644 u+0642 u+0645 u+0631	xn--mgbu4chg
Republic of Djibouti	جيبوتي	u+062C u+064A u+0628 u+0648 u+062A u+064A	xn--ngbee7iid
Kingdom of Saudi Arabia	السعودية	u+0627 u+0644 u+0633 u+0639 u+0648 u+062F u+064A u+0629	xn--mgberp4a5d4ar
Democratic Republic of Sudan	السودان	u+0627 u+0644 u+0633 u+0648 u+062F u+0627 u+0646	xn--mgbaxp8fpl
Syria Arab Republic	سورية	u+0633 u+0648 u+0631 u+064A u+0629	xn--ogbpf8fl
Somalia Democratic Republic	الصومال	u+0627 u+0644 u+0635 u+0648 u+0645 u+0627 u+0644	xn--mgba5b5cceu
Republic of Iraq	العراق	u+0627 u+0644 u+0639 u+0631	xn--mgba3a5azci

		u+0627 u+0642	
Sultanate of Oman	عمان	u+0639 u+0645 u+0627 u+0646	xn--mgb9awbf
Palestine	فلسطين	u+0641 u+0644 u+0633 u+0637 u+064A u+0646	xn--ygb2ammx
State of Qatar	قطر	u+0642 u+0637 u+0631	xn--wgl6a
State of Kuwait	الكويت	u+0627 u+0644 u+0643 u+0648 u+064A u+062A	xn--mgbg8edvm
Lebanese Republic	لبنان	u+0644 u+0628 u+0646 u+0627 u+0646	xn--mgb7fjb
Socialist People's Libyan Arab Jamahiriya	ليبيا	u+0644 u+064A u+0628 u+064A u+0627	xn--mgb7fyab
Arab Republic of Egypt	مصر	u+0645 u+0635 u+0631	xn--wgbh1c
Kingdom of Morocco	المغرب	u+0627 u+0644 u+0645 u+063A u+0631 u+0628	xn--mgb0a9azcg
Islamic Republic of Mauritania	موريتانيا	u+0645 u+0648 u+0631 u+064A u+062Au+0627 u+0646 u+064A u+0627	xn--mgbah1a3hjkrd
Yemen Arab Republic	اليمن	u+0627 u+0644 u+064A u+0645 u+0646	xn--mgb2ddes

4. ICANN GUIDELINES FOR IMPLEMENTING IDN

In 2003, ICANN issued a guideline (Version 1.0) for implanting IDN. Currently, ICANN is preparing the second IDN implementation guideline (Version 2.0). However, both guidelines are still based upon a handicapped IDN solution (i.e., IDN.English) that does not support full IDN on a TLD level (i.e., IDN.IDN). This implementation scenario is not suitable for languages that are not Latin-based, for example, languages written from right-to-left (e.g., Arabic, Farsi, Urdu, ...) or ideographic languages (e.g., Chinese, Japanese, Korean, ...).

Furthermore, the ICANN IDN guideline Version 2.0 reflects the experiences of the IDN registries which have implemented Version 1.0 (i.e., registries which provide IDN.English, such as gTLD registries). This excludes the experiences collected by different entities around the world (e.g., ccTLD managers) who strive to support their languages in domain names. Focusing on IDN.English solutions has introduced numerous problems due to the use of multiple languages under one label. This can be eliminated by applying a full IDN.IDN solution to be used by ccTLD managers for their corresponding languages. It will be the responsibility for each ccTLD (who has the control over their IDN ccTLD name) to draft their own guidelines and resolve any problems that may occur related to their own language.

To clarify our point we will introduce two terms:

- ◆ Poly-IDN: That represents an international domain name in which each label in the domain name can be expressed using different language character set (e.g. "Arabic.English" or "Chinese.Arabic" ...etc).
- ◆ Mono-IDN: That represents an international domain name in which all labels in the domain names are expressed using the same language character set (e.g. "Arabic.Arabic" or "Chinese.Chinese" ...etc).

Thus, we believe that while ICANN is focusing on supporting Poly-IDN and solving their problems, this should not prohibit the implementation of Mono-IDNs by ccTLD administrators who will develop their language guidelines and share them with ICANN.

5. RECOMMENDATIONS

- ◆ A more practical approach, even if only for testing purposes, is to start the IDN support at a ccTLD level rather than on a gTLD. So that the TLD is written in a specific language (e.g. Arabic) that will be supported also on the SLD controlled by the same character set table (i.e., Mono-IDN).
- ◆ It is strongly believed that concerns and issues that are raised by the guidelines regarding IDN implementations will not be applicable when internationalized ccTLDs are supported.
- ◆ The Arabic domain names pilot project (www.arabic-domains.org) support the following principles that have been stated in the proposal submitted by the Chinese Domain Name Consortium (CDNC) (www.icann.org/announcements/idn-tld-cdnc.pdf) to ICANN, namely:
 - Give priority to internationalizing ccTLDs. To ensure the system stability, it's recommended to internationalize ccTLD before internationalizing gTLD.
 - For convenience purposes, only one form of language character variant of internationalized ccTLD is accepted. Considering that some countries or regions may have character variants, only one form of character sets shall be chosen for IDN use by each sponsored registry.
 - Supported by their own governments, ccTLD registries or authorized agencies shall make their own choices regarding which IDN character sets will be implemented in their ccTLDs.
 - Register and operate the internationalized ccTLDs in the root DNS server in the form of IDNA Punycode.