

Internationalized Domain Name System (IDNS): Future of the Internet

Tariq Rahim Soomro and Ghassan Al-Qaimari

Department of Information Technology, Fujairah College, Fujairah, United Arab Emirates

Abstract: The Internet has changed everything, the way we live our life, the way we communicate with each other and the way we do business. It is a world-wide broadcasting capability, a mechanism for information and interaction between individuals without regard to geographical distances. However, it is no denying a fact that the majority of languages used on the Internet are non-English. Keeping in view these facts non-English speakers diverge their efforts to access their local language sites easily and comfortably. Efforts in this regard have already been made in Chinese, Japanese and Korean (CJK), in Arabic and in other languages. That's why the need for an Internationalized Domain Name System (IDNS) is foreseen. The IDNS allows people from under-developed countries to participate on the Internet in their own local languages. This platform is the main standard technology, which helps the under-developed economies to overcome the technological barriers and to take part into main stream. This study is written keeping in view the future use of the Internet by non-English speakers using IDNS platform. The study first discusses the concept of IDNS platform, followed by issues related to Urdu, Sindhi and Arabic Top Level Domains (TLDs) and how to accommodate these languages in main stream Internet and issues related to implement IDNS and finally the concluding remarks.

Key words: Internet, DNS, IDNS, meta DNS, future of the internet

INTRODUCTION

People in developed countries have better access to the Information and Communication Technologies (ICT) and use them more intensively and more effectively than do people in underdeveloped countries. It is no denying a fact that Internet usage is far more common in richer countries than in poorer countries. Furthermore, the number of users has grown more quickly in richer countries. It is also fact that the Internet and new technologies could help stimulate economic development (George *et al.*, 2005; DNS and IDNS, 2007). The Internet from its birth was in English. Since its inception in 1992, the Internet has produced tremendous social and economical wealth in the United States of America (USA). Subsequently the Internet has grown into other countries and the people speaking languages other than English has also enjoyed part of the wealth, which the Internet has brought. However, the language barrier has gradually shown its limiting impact on the further expansion of the Internet usage in non-English speaking countries, which are mostly developing or under-developing, countries.

The majority of languages used on the Internet are non-English languages (George *et al.*, 2005; DNS and IDNS, 2007) (e.g., Chinese, Japanese, Korean, German, Arabic, Persian, Urdu, Sindhi or other regional languages).

Table 1: Major language

Language users	(%)
Chinese language	15.7
Spanish language	8.7
Japanese language	7.4
German language	5.0
French language	5.0
Portuguese language	4.0
Korean language	2.9
Italian language	2.7
Arabic language	2.5
Rest of the world's languages (Which includes, Russian, Persian, Urdu, Hindi, Sindhi and others)	15.0

According to <http://www.internetworldstats.com/stats7.htm> (2007), English speaking users represent 31.2% of the total users of the Internet. The percentage of users from other major languages is shown in the Table 1.

Similarly, according to a research work published by "Byte Level Research" (2007), less than 30% of the Internet users of the world are native-English speakers. The study however predicts that this number will drop to less than 25% by 2010. It is also expected that the social, economical and political developments in China, Russia, Brazil, India, Pakistan and the Arab countries will drive millions of non-English-speaking people to the Internet (DNS and IDNS, 2007). Keeping in view these facts it was obvious that non-English Speakers would device a mechanism to access their own language sites, sending

and receiving email in their own languages with their own local scripts. Given that this is only possible by adapting an Internationalized Domain Name System (IDNS), efforts in this regards have successfully been made in Chinese, Japanese and Korean (CJK) (<http://www.i-dns.net/index.html>; www.apng.org/old/commission/idns/APNG-APTLD-idnsFeb99.ppt; <http://www.minc.org/Default.aspx?andlang=en>), in Arabic (<http://www.minc.org/Default.aspx?andlang=en>; <http://www.i-dns.net/newsroom/news/EG010129-01.html.en>; <http://192.0.34.163/announcements/adnppv02-12dec05.pdf>; www.itu.int/ITU-T/worksem/multilingual/papers/s2paper-sabouni.pdf; <http://www.actapress.com/PaperInfo.aspx?PaperID=30992>). Other languages, such as Hindi, Tamil, Urdu, Sindhi, Persian, etc, are yet to take full advantage of the Internet (<http://www.actapress.com/PaperInfo.aspx?PaperID=30992>; <http://www.actapress.com/PaperInfo.aspx?PaperID=30996>).

This study is written keeping in view the future use of the Internet by non-English speakers using IDNS platform. The flexibility of IDNS and its ability to accommodate local languages will have a crucial impact on the future use of the Internet and the development of the e-business and e-government services in developing and under-developed countries.

INTERNATIONALIZED DOMAIN NAME SYSTEM (IDNS)

In the past few years, the importance of domain names for identifying businesses, companies, departments, governments, locations, newspapers, projects, research and researchers has significantly increased. Domain Name System (DNS) stores and associates many types of information with domain names, but most importantly DNS is a translator. It translates domain names to IP addresses. It is actually providing a world-wide keyword-based redirection service. Domain name on the Internet today accommodates only a subset of ASCII characters. Current Domain Name System (DNS) names are restricted to English ASCII case insensitive [A-Z] plus numbers [0-9] and hyphen. Currently DNS server software does not support multilingual multi-script UNICODES. Therefore, DNS is unable to support Non-English languages needs of companies and e-commerce (<http://www.actapress.com/PaperInfo.aspx?PaperID=30996>).

An Internationalized Domain Name System (IDNS) is designed to use the multilingual characters as well as English alphabets, numerical characters and some symbols without any modification to existing DNS

systems (http://www.wgig.org/docs/book/Kangsik_Cheon.pdf). The main objective of IDNS is to allow the use of domain names, which are not restricted to the mere 38 characters used in the original DNS. Instead, the IDNS is associated with Unicode- (ISO-10646) based characters, which contains tens of thousands of possible characters known as "code points". The technical solution for the IDNS was introduced in the Request for Comments (RFC) 3490, 3491 and 3492, published in March 2003 (Internet Engineering Task Force (IETF) 2007). In sum, the technical solution relies on keeping the standard DNS character set "on the wire" for compatibility with the currently deployed DNS infrastructure and applications (www.itu.int/ITU-T/worksem/multilingual/papers/s2paper-sabouni.pdf). The major advantages of INDS names are: It will be easy for the users to remember domain names when the domain name character sets are in their native languages; it will be helpful for the users who are not familiar with English language; and the marketing effectiveness of a domain name will be increased given that more and more users will be able to use a common IDNS platform.

PROPOSED TOP LEVEL DOMAINS (TLDs)

Urdu and Sindhi languages contains several common alphabets of Arabic language, as well as using the same writing script. Currently Urdu and Sindhi web-sites can be found on the Internet - mostly personal and multimedia sites, in addition to cultural and historical sites, which are available in these languages along with Newspaper and Magazine sites (<http://www.actapress.com/PaperInfo.aspx?PaperID=30992>). For example, among the Urdu leading newspapers are www.jang-group.com, www.nawaiwaqat.com and www.jasarat.com and among the Sindhi newspapers are www.dailykawish.com, www.ibratgroup.com and www.awamiawaz.com. These newspapers are widely read because they are in national and regional languages. Compared to Daily Dawn www.dawn.com, one of the leading English newspapers in Pakistan the "Daily Dawn" is only read by the limited 4% Pakistani intellectuals, who can read English (Afzal, 2007).

Languages play an important role in this information era and according to linguists languages are dynamic and keep evolving over time. For example, while Japanese has changed relatively little over the last one thousand years (because it does not adopt foreign words/jargons), English has evolved rapidly in the last few centuries. Languages change for a variety of reasons, especially due to the large-scale shifts often occurring in response to social, economic and political pressure. The most

Table 2: Top Level Domains (TLDs) for Urdu, Sindhi and Arabic

English TLD	Proposed Urdu TLD	Proposed Sindhi TLD	Proposed Arabic TLD
Com	كوم	كوم	كوم
Edu	ايجو	ايجو	اڊيو
Gov	گو	گو	غوف
Net	نيت	نيت	نت
Org	اور	اور	اورغ
Mil	مل	مل	مل
Int	انت	انت	انت
Info	انفو	انفو	انفو
Biz	بز	بز	بز
Name	نيم	نيم	نيم
Coop	كوپ	كوپ	كوپ
Jobs	جوب	جوب	جوب
Pro	پرو	پرو	پرو
Mobi	موبى	موبى	موبى
Museum	ميوزيم	ميوزيم	ميوزيم
Many more

important reason for language change is the introduction of new technologies, industries, products and experiences, which simply require new words. Plastic, cell phones and the Internet didn't exist in English in Shakespeare's time, for example. By using new and emerging terms, we all drive language change (Nicole, 2007).

The Urdu and Sindhi Top Level Domains are proposed here keeping in mind the fact that languages are adopting jargons. Currently the ".com" term is very widely used in the Urdu and Sindhi languages. Therefore, the proposed TLDs give the unique phonetic option of using .com as "كوم." (c = ك ; o = و ; m = م) for Urdu, Sindhi and Arabic users instead of using شركة. or ش. in Arabic, which may leads to some confusion. The proposed (Tariq and Al-Qaimari, 2007) Top Level Domains (TLDs) for Urdu, Sindhi and Arabic are given in Table 2.

PROPOSED IMPLEMENTATION OF IDNS

Currently, the domain name space consists of a tree structure of domain names. Each node, or leaf, in this tree has an associated resource record, which is holding the information associated with the domain name. The tree is further divided into zones. A zone consists of a collection of connected nodes authoritatively served by an authoritative DNS nameserver. Note that a single nameserver may host several zones. When a system administrator wants to let another administrator control part of the domain name space within his or her zone of authority, he or she may delegate control to the other administrator. A resolver looks up the information

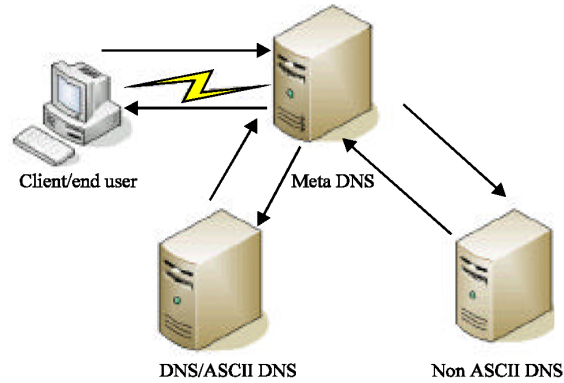


Fig. 1: Server tree

associated with nodes. It knows how to communicate with name servers by sending DNS requests and heeding DNS responses. Resolving usually entails recursion through several name servers to find the needed information (Tariq and Al-Qaimari, 2007).

As suggested by RFC-3490 (www.rfc-editor.org/rfc/rfc3490.txt), the Internationalized Domain Name Application (IDNA) works by allowing applications to use certain ASCII name labels (i.e. beginning with a special prefix) to represent non-ASCII name labels. Lower layer protocols need not be worried about; therefore, INDA does not depend on changes to any infrastructure. In particular, IDNA does not depend on any changes to DNS Servers, resolver, or protocol elements, because the ASCII name service provided by the existing DNS is entirely sufficient for IDNA. Implementing IDNA has some limitations, to name one it does not solve all linguistic issues with users inputting names in different scripts.

Keeping in view the limitations of implementing IDNA, the guidelines of RFC-3490 (www.rfc-editor.org/rfc/rfc3490.txt) and the guidelines for the implementation of Internationalized Domain Names System (<http://www.icann.org/general/idn-guidelines-14nov-05.htm>), the concept of Meta Domain Name Systems (Meta DNS) was proposed by authors (Tariq and Al-Qaimari, 2007). The proposed Meta DNS basically works like the current DNS. The request to resolve DNS or IDNS will go to the Meta DNS at root level. It will check first with the help of IsAscii() to find out whether the request is for ASCII domain name or for non-ASCII domain name. If the request is identified for ASCII domain name it will be forwarded to DNS/ASCII DNS (or ASCII-Based DNS) Server tree (which is currently known as DNS Server) to resolve. On the other hand, if the request is identified for non-ASCII domain name, it will be forwarded to Non-ASCII DNS (NADNS) Server tree to resolve with the help of special prefix, gTLDs and ccTLDs as shown in Fig. 1.

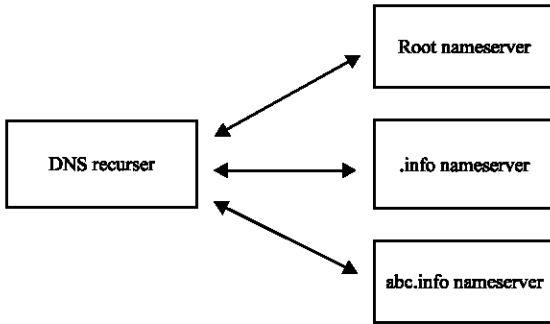


Fig. 2: Three nameserver

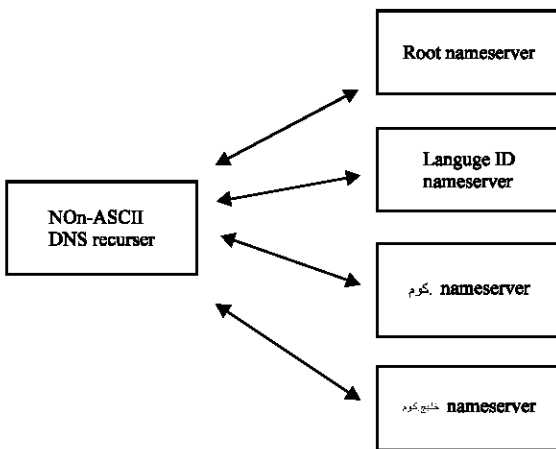


Fig. 3: Four nameserver

This proposed solution agrees with (www.rfc-editor.org/rfc/rfc3490.txt) and (<http://www.icann.org/general/idn-guidelines-14nov-05.htm>), that the current DNS mechanism should not be altered. That is why we had not proposed any change to the current DNS, instead we had introduced a parent node to the current DNS. Note that all the resolving requests should work as normal, even after introducing Meta DNS. The resolving of DNS or ASCII-Based DNS should work the same way it is working now, except it is now a Child instead of Root/Parent. As shown in Fig. 2, a DNS recursor consults three nameservers to resolve the address `www.abc.info`, for example, these are root nameserver, `.info` nameserver and `abc.info` nameserver.

The resolving of Non-ASCII DNS or NADNS will work the same way currently the DNS is working, except in the increase in recursion process i.e., the LanguageID nameserver. A Non-ASCII DNS recursor consults four instead of three nameservers to resolve the address `www.خليج.كوم`, for example, as shown in Fig. 3. The difference between both approaches is LanguageID nameserver, which was proposed to resolve Non-ASCII Languages (Tariq and Al-Qaimari, 2007).

CONCLUSION

The number of Non-ASCII users is rapidly increasing on the Internet. More and more Non-ASCII languages' Websites, such as Chinese, Japanese, Korean, Arabic, French, German, Hindi, Urdu, Sindhi and others will be available on the Internet. Keeping in view this fact, standardization of IDNS is required, because currently many different implementations are available around the world. The study presented and proposed Urdu, Sindhi and Arabic Top Level Domains (TLDs). ASCII TLDs used short form, e.g. ".com" instead of ".commercial", ".org" instead of ".organization". This is to reduce the typing efforts and avoid spelling mistakes. The same approach has been used in proposing Urdu, Sindhi and Arabic TLDs. Also there will be unique phonetic TLDs either ASCII based or Non-ASCII based, for example.com will remain.com in Urdu (كوم), in Sindhi (كوم) and in Arabic (كوم). This study also presented Meta Domain Name System (Meta DNS), which is capable of being used as standardized solution to implement IDNS. It has several advantages over current implementations (Tariq and Al-Qaimari, 2007) especially the lack of dependency on a particular Browser, on Client-based software, on Server-based software and on ISPs.

The implementation of a common IDNS will open new doors for developing countries to boost their economies and to compete with other developed countries using same technology. It is expected that after implementing a common IDNS platform, the rate of population penetration of under-developed countries will more than double, as more and more people will be able to use their own local languages easily and comfortably. What we are endorsing in this study is a multilingual approach to Domain Name System, where Websites will be available not only in English (ASCII-Based), but also available for local customers in their own local script (UNICODE-Based). A common IDNS should not only be flexible enough to enable English language along with non-English languages, but it should also provide a common platform to accommodate other technological opportunities like ftp, Email, IRC etc. The implementation of a common IDNS will enable developing economies to run their own e-business and e-government applications in their local languages, which will be of great help to people who are not familiar with English language. The future use of the Internet will see more and more multilingual users on the Internet and IDNS will be part of the Internet in near future. We intend in future publications to propose Country Code TLDs (ccTLDs). We also intend to conduct a number of usability experiments with the help of targeted users from Arabic, Urdu and Sindhi background to assess the practicality of our proposed ideas.

REFERENCES

- A multilingual The Internet for everyone, everywhere, IDNS Technical presentation, 2006, Retrieved 2007 form: <http://www.i-dns.net/index.html>.
- Arabic Domain Name System (ADNS) Status and Issues, ITU and UNESCO Global Symposium on Promoting the Multilingual The Internet, Geneva, 2006, Retrieved 2007 form: www.itu.int/ITU-T/worksem/multilingual/papers/s2paper-sabouni.pdf.
- Arabic Domain Names, Pilot Project, ICANN 2005, Vancouver, Retrieved 2007 form: <http://192.0.34.163/announcements/adnpp-v02-12dec05.pdf>.
- Afzal, M., 2007. Overview of Urdu Informatics and Standardization, Urdu The Internet Name System Workshop-IIUCN, 2001. Retrieved 2007 from: <http://urduworkshop.sdnpk.org/proceedings.htm>.
- Tin Wee, T., 2007. iDNS, Internationalized DNS, Asian Pacific Networking Group, Retrieved 2007 form: www.apng.org/old/commission/idns/APNG-APTLD-idnsFeb99.ppt.
- English-Speaking Web Users a Minority on the Internet, 2005, Byte Level Research, Retrieved 2007 from: http://www.bytelevel.com/news/web_languages_nov3_2005.html.
- From DNS to IDNS, 2007. The Impact on Developing Economies, Tariq Rahim Soomro and Ghassan Al-Qaimari, 8th International Business Information Management Association (IBIMA) Conference on Information Management in the Networked Economy, Dublin, Ireland.
- George, R.G., Clarke and J. Scott, 2007. Wallsten, Has the Internet increased trade? Developed and developing country evidence, Economic Inquiry Advance Access published online on December 1, 2005, Oxford Journals Oxford University Press, Retrieved 2007, form: <http://ei.oxfordjournals.org/cgi/reprint/cbj026v1>.
- Guidelines for the Implementation of Internationalized Domain Names, Version 2.0, 2005, <http://www.icann.org/general/idn-guidelines-14nov-05.htm>.
- Internationalized Domain Names in Applications (IDNA), RFC-3490, 2003, Retrieved 2007 from: www.rfc-editor.org/rfc/rfc3490.txt.
- Internet Engineering Task Force (IETF) 2007. Retrieved 2007 form: www.ietf.org.
- Kangsik, C., 2007. Multilingualism and the Domain Name System, Retrieved 2007 form: http://www.wgig.org/docs/book/Kangsik_Cheon.pdf.
- Multilingual the Internet Names Consortium, 2006. Retrieved 2007 form: <http://www.minc.org/Default.aspx?and lang=en>.
- Nicole, M., 2005. Language and linguistics, A special report, Language Change, 2005. Retrieved 2007 from: http://www.nsf.gov/news/special_reports/linguistics/change.jsp.
- Tariq, R.S. and G. Al-Qaimari, 2007. Accommodating Urdu, Sindhi and Arabic in Internationalized Domain Name Systems (IDNS), the 6th IASTED International Conference on Communication, Internet and Information Technology, CIIT, Banff, Alberta, Canada. <http://www.actapress.com/PaperInfo.aspx?PaperID=30992>.
- Tariq, R.S. and G. Al-Qaimari, 2007. Implementation issues of Internationalized Domain Name Systems (IDNS), the sixth IASTED International Conference on Communication, Internet and Information Technology, CIIT, Banff, Alberta, Canada. <http://www.actapress.com/PaperInfo.aspx?PaperID=30996>.
- The Internet World users by language, 2007. Retrieved on August 09, 2007 form: <http://www.internetworldstats.com/stats7.htm>.
- When can we write URLs in Arabic? The Language Barrier Holds Most of the Arab Users from Using the Internet, 2001. Retrieved 2007 form: <http://www.i-dns.net/newsroom/news/EG010129-01.html.en>.