International Chamber of Commerce
The world business organization

Department of Policy and Business Practices

Commission on E-Business, IT and Telecoms
Task Force on the Internet and IT Services

Issues Paper on Internationalized Domain Names

The introduction of Internet domain names in LDH (Letters Digits Hyphen)[1] characters is the subject of much controversy and debate. Some contend that it is an almost trivial exercise while others argue that it is a tremendously complex task that if done too hastily or without proper planning threatens the integrity and stability of the Internet. The Internet Engineering Task Force (IETF) has produced a number of 'Requests for Comments (RFCs)' on the topic (see http://www.rfc-editor.org/rfc/rfc3490.txt) that provide guidance on the issue.

The Task Force on the Internet and IT Services has developed this Issues Paper on internationalized domain names (IDNs) to explain the need for IDNs in a manner that ensures the flexibility, stability, and global interoperability of the Internet. Given the current existence of numerous languages and some 241 top-level domain names (TLDs), there is no question that it is a huge and complicated task. Indeed, it is one that could quickly become mired in boundless problems.

This paper examines a number of the issues surrounding IDNs, such as the need for their introduction, the technical challenges, and the risks to the current domain name system. It also addresses their impact on business, including the ability of companies to protect their intellectual property. Finally, it outlines policy issues such as the impact on IDN's on current policies governing the domain name system.

What are Internationalized Domain Names?

A domain name is the unique character-based label assigned to a numbered address. Although domain names are used in a number of different applications, in the context of this paper the Domain Name System (DNS) resolves and identifies the address and points a browser to a particular computer containing the user's requested data.

[1] The universal set of characters a-z, 0-9, hyphen and dot.

When a user enters a domain name in his browser, e.g., iccwbo.org, the user's computer accesses the global DNS directory to find the corresponding Internet Protocol Address (IP Address) of the website. On finding the IP Address, the user's computer is then able to contact and communicate with the computer bearing the IP Address corresponding to the iccwbo.org domain name. In effect, the domain name is simply the IP Address, represented in an easy to remember form, of the computer where ICC's data is held for public, or even private, access. The DNS allows the user's computer to replace the text-based label iccwbo.org with the IP Address and thus, locate and communicate with the ICC computer over the global Internet.

A domain name consists of both low level and top-level domain (TLD) name components. In the domain name "iccwbo.org", the low level domain name is "iccwbo", whilst the top level domain is ".org". There are different types of top-level domain names; generic TLDs (gTLD) such as .org, .com, .info, and country code TLD (ccTLD) such as .uk (United Kingdom), .fr (France), .jp (Japan).

Since the original language and characters available for use in computers and the Internet were based on ASCII (American Standard Code for Information Exchange)[2] codes/characters which use Latin characters, domain names were restricted initially to the LDH subset of ASCII characters or through the transliteration of non-Latin based languages into this subset.

Subsequently, the introduction of UNICODE[3], "provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language."[4] has created a uniform foundation for global software irrespective of language. UNICODE now comprises 96,382 characters from currently recognized scripts of the world. The standard is continuously upgraded to add new characters and character sets. New, and what may be described as non-transliterated Internationalized Domain Names, take advantage of this technological facility and allow the use of domain names in character sets other than LDH characters

Domain names can be 'internationalized' by allowing non-LDH characters in the domain. Domain names that may have appeared as "Urdu.com" can now appear as:
" ارد و. com"
" ارد و. com.pk"
"کوّم . ارد و . pk"
and even
"پک . کوّم . ارد و "

Different TLDs operators are currently offering Internationalized Domain Names. The question now is not how to enable internationalization of domain names but how to ensure that the processes for development, maintenance, upgrade and resolution proceed in a manner that will preserve the stability, integrity and security of the Internet. Specifically, it is necessary to have a uniform encoding of IDNs regardless of the specific

[2] ASCII (American Standard Code for Information Interchange). The ASCII characters are Latin or Roman language characters with a maximum of 128 defined alpha, numeric and special characters.
[3] For more information: www.unicode.org
[4] www.unicode.org/standards/What IsUnicode.html. Accessed June 20, 2006.

application that is using them. Browsers, as used in this paper, are only one example of such an application.

It is important, however, that the reader does not come away with the impression that providing for IDNs in the Domain Name System will resolve all issues with regard to the use of non-LDH characters. Mail systems are an example of an application that will require extensive modification to accommodate the use of IDNs.

The Case for Internationalized Domain Names

The demand for IDNs is based on the desire for increased access to the information and knowledge available online. Much of the world's population today does not use, or even recognize, Latin characters.

A multilingual Internet will foster an inclusive, democratic, legitimate, respectful, and locally empowering Information Society. In this regard, it can be said that what is truly needed are localized domain names – or, the ability to access the Internet in one's native language. To offer localized domain names, the Domain Name System (DNS) must be multilingual. Of course, the ability to obtain useful content in an individual's native language is a significant issue, but beyond the scope of this paper.

Non-LDH domain names are also necessary to safeguard the cultural and linguistic integrity of names, brands and trademarks represented in native character scripts. Companies and individuals in societies that use non-Latin characters will be able to globally preserve their unique web-identities.

> Example:
> A Japanese person's name "□□" is transcribed as "hirofumi" in Roman letters. On the Internet, where only LDH characters can be used, he is "hirofumi", just like other people named "hirofumi" but whose names may use different Japanese characters such as "□□" or "□□". In fact, there may be over 100 different Japanese representations that will end up being denoted simply as "hirofumi" in LDH space. Consequently, in the LDH world, the person in question is just one "hirofumi" of many other Japanese "hirofumis", although in his native Japanese characters he would be clearly differentiated."[5]

While internationalized domain names will certainly help in achieving many objectives of developing countries, they cannot be considered to be the sole bridge for the digital divide. It is also important to consider the challenges posed by the fact that many people do not have access to computers in general and the Internet in particular. It is those who have access to both but are unable to practically use computers or the Internet due to language restrictions that are the ones most affected by the lack of internationalization.

[5] Paragraph 14 of the Multilingual Domain Names: Joint ITU / WIPO symposium in association with the Multilingual Internet Names Consortium - http://www.itu.int/mlds/briefingpaper/

General Issues and Concerns

Technical Issues

There are substantial technical issues surrounding the introduction of IDNs, many of which are quite complex. The technical community is working toward their resolution. Rather than trying to include them specifically in this paper, the reader is encouraged to consult the Internet Engineering Task Force (IETF) paper at http://www.ietf.org/Internet-drafts/draft-iab-idn-nextsteps-02.txt. This document describes some of the issues in detail and outlines the areas where further work is needed.

Intellectual Property Issues

The possibility for confusion among domain names raises several intellectual property issues. Several languages contain strings of characters that have equivalent or near-equivalent meanings. Use of such character strings in IDNs might lead to domain names that are similar phonetically, visually or across various character tables.

While classification of goods and services allows the use of a trademarked brand that might be similar or identical to another as long as they relate to a separate class of goods and services, this is not the case with respect to IDNs where there is no classification of domain names. Avoiding conflict and having a uniform globally enforceable dispute resolution policy is imperative for an efficient continued working of the DNS.

Lack of interoperability and coordination between registration authorities can lead to concerns by owners of domain names who must retain the ability to protect their trademark, trade name or brand. The ability of registration authorities to transfer domain names in case of breach of good faith on the part of respondents is essential, as is effective enforcement. Failure to resolve these issues will make the implementation of IDNs prohibitively expensive for business in trying to protect their IP rights.

The Internet Corporation for Assigned Names and Numbers (ICANN) and the World Intellectual Property Organization (WIPO) have already had to deal with the various issues that arise out of trademark and intellectual property disputes in IDNs. ICANN has contributed to the resolution of disputes in this area through the adoption of the Rules for Uniform Domain Name Dispute Resolution Policy (UDRP) on 24 October 1999, which has been used by WIPO in deciding cases on IDNs. The UDRP applies equally to both registered as well as unregistered trademarks. WIPO has, to date, decided 45 cases of non-ASCII Domain Names using Chinese, Dutch, French, German, Japanese, Korean, Norwegian, Spanish and Swedish languages.

However, some feel the UDRP still needs reform in the area of IDNs. For instance, use of the UDRP is driven by the existence of bad faith and lack of legitimate right to a domain name. A problem arises when both parties are bona fide and have conflicting legitimate rights and wish to or are using the domain name in good faith. This can have a substantial impact on business when faced with competitors to domain names that may phonetically sound or visually depict confusingly similar trademarks. This issue needs to be considered in order to have a system that addresses the disparate Internet scenarios that exist.

Security Issues

The possibility of confusion between phonetically similar or visually similar IDNs may be used for spoofing[6] or phishing[7] as well as cyber-squatting[8]. While such practices are possible within the LDH DNS, recently cyber criminals have taken advantage of the increased vulnerability in IDNs and the IDN system to confuse users regarding which web address or web page they are visiting. It is important for business to be aware of such vulnerabilities when managing their companies. In addition, there will be a need for industry support for development of solutions to the problem through technical and policy initiatives.

Language issues

The foundation for internationalization of computers and the Internet depends upon the availability and usage of character sets and character tables that are mapped to a universally recognized system such as UNICODE.[9] However, the diaspora effect on languages, character tables and sets may lead to differences in national, official, regional, local and diasporan languages, causing further confusion and conflict for intellectual property in trade names, trademarks and brands.

For example, some Chinese characters have two representations – a traditional Chinese character and a simplified Chinese character. Correspondence between a traditional Chinese character and a simplified Chinese character is not one-to-one, and while they are usually used in mainland China in place of traditional Chinese characters, simplified Chinese characters are seldom used in Taiwan or Hong Kong. Such issues clearly will clearly be difficult to decide, dictate or solve by regional, local or linguistic groups, since there will inevitably be conflict within such groups.

There are a number of organizations and consortiums attempting to resolve these differences. The Multilingual Internet Names Consortium (MINC)[10] is one such group that has been active in raising the issue and in coordinating and compiling a number of language groups. The Unicode Consortium is a transparent, open and global institution that maintains the global Standard[11] for Languages character sets and tables. The Consortium cooperates with the World Wide Web Consortium (W3C) and International Organization for Standardization (ISO)[12] and liaises with ISO to ensure synchronization of the Unicode Standard with the International Standard ISO/IEC 10646. The Unicode Consortium includes major computer corporations, software producers, database vendors, research institutions, international agencies, various user groups, educational institutions, governments and interested individuals. The Unicode Consortium enables a globally unified and effective solution to problems related to character sets that ensure

[6] A technique used to gain unauthorized access to computers, whereby the intruder sends messages to a computer with an IP address indicating that the message is coming from a trusted host.
[7] A technique whereby the websites of known institutions are entirely or partly copied and e-mails are used to obtain private or confidential data of the customers of those institutions. The request to provide those data is often motivated by so-called safety measures or the need to update data banks.
[8] Cyber-squatting is the act of registering a popular Internet address, usually a company name, with the intent of selling it to its rightful owner.
[9] Character Sets is a widely used term and may mean any or a combination of the following three: character repertoire, character code, and character encoding. A tutorial on character code issues is available at http://www.cs.tut.fi/~jkorpela/chars.html
[10] www.minc.org
[11] Unicode Standard 4.0
[12] www.iso.org

that the standard is implemented not just over regional or local jurisdictions but throughout software, computers and the Internet.

Introduction of Uniform Resource Identifiers (URI) may further assist in the internationalization of the Internet. A Uniform Resource Identifier (URI) is a compact string of characters for identifying an abstract or physical resource.[13] URIs may contain information from all kinds of protocols or formats that use characters beyond ASCII. However, the URI syntax currently only allows a subset of ASCII – about 60 characters. Internationalized Resource Identifiers (IRIs) are sequences of characters from the Universal Character Set that can be mapped to URIs, which means that IRIs can be used instead of URIs where appropriate to identify resources.

The World Wide Web Consortium (W3C) develops protocols and guidelines to maximize 'Web interoperability'. By publishing open (non-proprietary) standards for Web languages and protocols, W3C seeks to avoid market fragmentation and thus Web fragmentation.[14]

The Internet Engineering Task Force (IETF) is concerned with the evolution of the Internet architecture and the smooth operation of the Internet through best practices for the Internet Community on the standardization of protocols and procedures. It also addresses the intellectual property rights and copyright issues associated with the standards process.[15]

Maintaining a Unified Domain Space

When implementing IDNs, there is a clear need for one domain space, the preservation of compatibility with current domain names, preservation of the uniqueness of the domain name space and the need to ensure that the Internet is not divided into islands.[16] If the domain space is managed by a variety of entities, it will result in an uncoordinated, conflicting and fractured Internet.

As a part of the introduction of IDNs, ICANN has a central function in preventing a breaking up of the domain spaces into regional or local authorities. Otherwise, the Internet will be transformed into islands of information that may very well conflict. For instance, domain space that may be allocated by one region may find that in another domain space managed by an independent domain space authority these web addresses are allocated to other users, owners or organizations.

This would particularly be the case where there are conflicts within language character tables and sets as a result of divergent views and usages. This will not just confuse Internet users but negate the usefulness of the Internet as an efficient tool to target and locate unique web addresses from a global, unified, singular directory. It would threaten the objective of an Internet compatible with globally unique domain names in a universally resolvable public space.

[15] www.ietf.org

[16] These are requirements of the Internet Architecture Board – http://www.iab.org

Conclusion

This paper has raised a number of issues and indicated a number of areas where it may be in the best interests of business around the world to provide further input and guidance to interested parties and policy-makers. Potential areas of interest are:

1. Policies regarding mixed IDNs.
2. UDRP reform.
3. Maintenance of a unified domain space.
4. Means of achieving consolidated language tables and character sets.

IDNs are an important next step in ensuring that information through the Internet is accessible to all users around the world.
However, if not carefully and centrally implemented, IDNs threaten to destabilize the Internet and disenfranchise the global user from his right to access correctly, efficiently and securely a singular and interconnected database of the global Internet currently available to the global citizen. There is a real concern that internationalized domain names may lead to different resolutions and results in a fragmented Internet.

* * * * *

ICC Commission on E-Business, IT and Telecoms (EBITT)
Business leaders and experts drawn from the ICC membership establish the key business positions, policies and practices on e-business, information technologies and telecommunications through the EBITT Commission. With members who are users and providers of information technology and electronic services from both developed and developing countries, ICC provides the ideal platform to develop global voluntary rules and best practices for these areas. Dedicated to the expansion of cross-border trade, ICC champions liberalization of telecoms and development of infrastructures that support global online trade. ICC has also led and coordinated the input of business around the world to the World Summit on the Information Society, Geneva 2003, Tunis 2005, and continues this effort in the activities established in the Tunis Agenda.
http://www.iccwbo.org/policy/ebitt/

About ICC
ICC is the world business organization, a representative body that speaks with authority on behalf of enterprises from all sectors in every part of the world. The fundamental mission of ICC is to promote trade and investment across frontiers and help business corporations meet the challenges and opportunities of globalization. Business leaders and experts drawn from the ICC membership establish the business stance on broad issues of trade and investment, e-business, IT and telecoms policy, as well as on vital technical and sectoral subjects. ICC was founded in 1919 and today it groups thousands of member companies and associations from over 130 countries. www.iccwbo.org