

**MULTILINGUAL DOMAIN NAMES:
JOINT ITU / WIPO SYMPOSIUM**

**ITU BRIEFING PAPER:
TECHNOLOGY AND POLICY ASPECTS**

In association with the
Multilingual Internet Names Consortium



The **International Telecommunication Union (ITU)** is an international organization which brings governments and industry together to coordinate the establishment and operation of global telecommunication networks and services; it is responsible for standardization, coordination and development of international telecommunications including radiocommunications, as well as the harmonization of national policies.

To fulfill its mission, ITU adopts international regulations and treaties governing all terrestrial and space uses of the frequency spectrum as well as the use of all satellite orbits which serve as a framework for national legislations; it develops standards to foster the interconnection of telecommunication systems on a worldwide scale regardless of the type of technology used; it also fosters the development of telecommunications in developing countries.

International Telecommunication Union

ITU Strategy and Policy Unit

Office of the Secretary-General

Place des Nations

1211 Geneva 20

Switzerland

Tel. +41 22 730 5809

Fax +41 22 730 6453

spumail@itu.int

Acknowledgements

This paper was prepared by Mr Hirofumi Hotta, Director for Corporate Planning at Japan Registry Service Co. Ltd. (hotta@jprs.jp) for the ITU portion of the joint ITU/WIPO Symposium on Multilingual Domain Names held on 6-7 December 2001, at the International Conference Center of Geneva (see <http://www.itu.int/mdns/>). Important contributions to the paper have also been provided by Dr Tan Tin Wee, Vice Chairman, Multilingual Internet Names Consortium (MINC) (tinwee@pobox.org.sg), retired Chairman, Asia Pacific Networking Group (APNG), and ExCo of Asia Pacific Regional Internet Conference on Operating Technologies (APRICOT).

This paper has also benefitted from the input and comments of internal and external reviewers, to whom we owe our thanks. These include Chinyong Chong, Avita Dodoo, Ivo Essenberg, Daniel Pimienta, Jim Reid, James Seng and Yoshihisa Takada. Joanna Goodrick oversaw production of the paper and served as editor.

A team from the ITU Strategy and Policy Unit, which is headed by Dr. Tim Kelly, organized the ITU portion of the joint ITU/WIPO Symposium. Avita Dodoo, Internet Policy Analyst, was the overall project manager under the supervision of Robert Shaw, ITU Internet Strategy and Policy Advisor.

ITU would like to acknowledge the contribution by MINC of its expertise and experience in the preparation of this symposium; particular thanks go to Ms YJ Park, CEO, MINC; Dr Tan Tin Wee, Vice Chairman, MINC; and Professor Shigeki Goto, Chairman, MINC.

We would also like to thank the Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT), Japan, for its generous voluntary contributions to the ITU New Initiatives Programme and for its kind assistance in coordinating the briefing paper.

The views expressed in this paper are those of the authors and do not necessarily reflect the opinions of ITU or its membership.

TABLE OF CONTENTS

Introduction	7
Demand for Multilingual Domain Names	11
History of the Development of Multilingual Domain Names	15
Technological Challenges to the Development of Multilingual Domain Names	19
<i>Technical Aspects of the Multilingualization of Domain Names</i>	<i>19</i>
<i>Basic Concepts of the IETF Working Group</i>	<i>21</i>
<i>Character Codes of Multilingual Domain Names</i>	<i>22</i>
<i>Client-Side Versus Server-Side Solutions</i>	<i>24</i>
<i>Standardization for Compliance with the Current DNS</i>	<i>26</i>
<i>Preparation of Internationalized Host Names (Nameprep)</i>	<i>27</i>
<i>ASCII Compatible Encoding (ACE)</i>	<i>28</i>
<i>Internationalizing Host Names in Applications (IDNA)</i>	<i>30</i>
Impact on the DNS Structure	33
<i>Alternative Roots</i>	<i>33</i>
<i>Multilingual Domain Name Resolution by Alternative Roots</i>	<i>34</i>
<i>Pseudo-Roots</i>	<i>35</i>
Policy and Coordination Issues Raised by Multilingual Domain Names	37
<i>Consideration of Multilingual Domain Names in Various TLDs</i>	<i>37</i>
<i>Potential Types of Multilingual Domain Names</i>	<i>37</i>
<i>Technical and Non-Technical Issues</i>	<i>38</i>
<i>Mixed Multilingual.ASII Domain Names</i>	<i>38</i>
<i>Multilingual.Multilingual Domain Names</i>	<i>38</i>
<i>What are the Languages that Constitute Multilingual Domain Names?</i>	<i>40</i>
<i>Who is the Language Authority for Multilingual Domain Names?</i>	<i>40</i>
<i>Matrix of Authority</i>	<i>41</i>
<i>Models for a Matrix of Authority</i>	<i>43</i>
Summary	45

Annex A: Glossary of Acronyms	47
Annex B: Some Implementations of Multilingual Domain Names	49
<i>Chinese Domain Name Consortium (CDNC).....</i>	<i>49</i>
<i>China Internet Network Information Center (CNNIC)</i>	<i>50</i>
<i>i-DNS.net.....</i>	<i>50</i>
<i>Japan Network Information Centre (JPNIC) / Japan Registry Services (JPRS)</i>	<i>50</i>
<i>Korea Network Information Center (KRNIC).....</i>	<i>51</i>
<i>NativeNames</i>	<i>52</i>
<i>Neteka.....</i>	<i>52</i>
<i>Netpia.....</i>	<i>53</i>
<i>New.net.....</i>	<i>53</i>
<i>RealNames.....</i>	<i>54</i>
<i>VeriSign Global Registry Services (VGRS)</i>	<i>54</i>
<i>WALID</i>	<i>55</i>

Introduction

1. A domain name is used to identify an entity within the Internet in a format that humans can easily understand; it has been one of the fundamental addressing schemes in Internet use for over 15 years. At the most basic level, it maps a human-readable name such as “www.itu.int” to a machine-readable Internet Protocol (IP) address (e.g. 156.106.134.92). In its current form, only a limited set of ASCII¹ characters, namely letters, digits and hyphens, can be used in domain names. Envisaged originally as a system of easily remembered identifiers to help network engineers address computers, there was no initial perceived need to expand the set of supported characters to include non-ASCII scripts.

2. However, the past decade has seen a wide global adoption of the Internet. Founded on innovative technological and economic principles, the Internet has experienced dramatic growth. It took 74 years for the telephone network to reach 50 million users. It took only 4 years for the World Wide Web to reach that same number. Today, the Internet is a global network of more than 230 connected economies and more than 350 million users.

3. One consequence of this growth is that the number of users, as well as Internet content, from societies and cultures not familiar with ASCII is growing daily. To address this phenomenon, e-mail and web pages in many scripts and languages are supported by various pieces of Internet software. Yet domain names, arguably one of the most visible symbols of the Internet, are still in ASCII characters and pose a significant linguistic barrier. Although users of languages based on Latin characters, either natively (e.g. English) or in a transliterated form (e.g. Malay), do not have linguistic problems with the current domain name system, native speakers of Arabic, Chinese, Japanese, Korean, Tamil, Thai and others who use non-ASCII scripts remain at a considerable disadvantage. In an attempt to solve this problem, as well as generally provide for improved multilingual and multiscript support, a process of “internationalization” of the Internet’s Domain Name System (DNS) has been underway.

¹ ASCII (American Standard Code for Information Interchange) is the most common format for text files in computers and on the Internet. In an ASCII file, each alphabetic, numeric, or special character is represented with a 7-bit binary number (a string of seven 0s or 1s). 128 possible characters are defined.

4. Since 1998, a number of technical solutions for this problem have emerged. More than a dozen commercial companies, as well as some country code² top-level domain (ccTLD) administrators, have set up a variety of technical multilingual domain name solutions. In the commercial market, there is intense competition with no clear winners emerging with a *de facto* standard.

5. Consumer demand has been extremely strong — particularly in Asian countries. By 2000, various “test beds” had been deployed around the world to offer multilingual domain names. However, for the most part, these solutions remain technically non-interoperable among themselves. Recognizing the problem, an Internationalized Domain Names (IDN) Working Group was formed within the Internet Engineering Task Force (IETF) in early 2000 to define a technical approach and related standards.

6. There has also been an emerging realization that multilingualization of the DNS is far from being an exclusively technical problem — it is also one of administration, management and policy. By 2001, organizations such as the Multilingual Internet Names Consortium (MINC), Arabic Internet Names Consortium (AINC), Chinese Domain Names Consortium (CDNC), International Forum for IT in Tamil (INFITT), and Japanese Domain Names Association (JDNA), as well as a number of other nascent language groups have emerged to occupy a policy vacuum.

7. In parallel, there have been major ongoing developments in administration and policy with respect to conventional ASCII-based domain names. In October 1998, the Internet Corporation for Assigned Names and Numbers (ICANN), a not-for-profit corporation, was established under the laws of the State of California, in the United States of America³. The following month, a Memorandum of Understanding (MoU) was signed between the US Department of Commerce and ICANN⁴. Under the framework of this MoU, ICANN has provided for competition in the domain name registration

² Country code top-level domains are based principally on the two-letter code set of the ISO 3166-1 Standard (e.g. .fr for France, .cn for the People’s Republic of China). See <http://www.din.de/gremien/nas/nabd/iso3166ma/> for a list of these codes.

³ For details on the organization and activities of ICANN, see <http://www.icann.org>.

⁴ See <http://www.ntia.doc.gov/ntiahome/domainname/icann-memorandum.htm>.

market, a uniform domain name dispute resolution policy (UDRP)⁵, and some new top-level domains (TLDs).

8. More recently, in March 2001, ICANN formally launched a number of activities related to multilingual domain names. A recent survey conducted by an ICANN internal working group⁶ has indicated that there is strong support for the rapid deployment of multilingual domain names.

9. Nevertheless, a great number of challenges and uncertainties remain as to when and how multilingual domain names will be deployed. At the time of preparation of this briefing paper (November 2001), the IETF's IDN Working Group had not reached the consensus needed for technical standardization of multilingual domain names. Considering the related debates, even if an IETF standard does emerge, it is unclear whether it will be universally adopted. Equally unclear is whether new emerging naming technologies not based on the DNS, such as keywords, will emerge as a preferred solution. There is even the possibility that hybrid technologies merging the DNS and keywords will surface. One result is that users have been left in a state of considerable confusion by a multiplicity of technologies, "test bed" deployments, and incompatible technologies.

10. Finally, the appropriate model for the assignment, administration and management of multilingual domains, including multilingual top-level domains, will need to be developed. ICANN, having only recently approached this problem, has not indicated any clear sense of the direction to be taken on this issue. In practice, national or regional approaches may differ widely according to local language requirements. In this case, there may be some sensitivity as to which authority would be responsible for what may be seen as national, localized or regional issues. Linguistic groups have also proliferated, adding yet another necessary level of coordination. All this suggests that the establishment of multilingual domain names may result in further challenges to the technology, policy and management aspects of the DNS.

⁵ Principally developed by WIPO, see <http://arbitrator.wipo.int/domains/index.html>.

⁶ See <http://www.icann.org/committees/idn/final-report-28aug01.htm>

Demand for Multilingual Domain Names

11. As the Internet originated in the United States, the technology has, not surprisingly, been very much based on the English language. Even those outside of the US who were pivotal in the development of the Internet typically had technical backgrounds and were familiar with English. Furthermore, ASCII codes have long been used at the core of computing and the Internet, especially early on, when resources such as central processing units and memory were limited. Because of these historical circumstances, even people in countries that do not use ASCII characters in their written languages have typically used ASCII characters when accessing services on the Internet. In addition, because users in the early stages of the Internet's development were from the research and academic communities, English language exclusivity did not prove to be significant obstacles to its expansion.

12. However, in more recent years, the Internet has grown to reach all corners of the world, to people of all ages and educational backgrounds, and is used by businesses and consumers alike. It is estimated that by 2003, two-thirds of all Internet users will be non-English speakers⁷. Furthermore, over 90 per cent of the world's population speaks a primary language other than English⁸. This means that, for an increasing number of people, English and the English alphabet will be considered barriers to becoming Internet users. These people will find it extremely unnatural to use the Internet in English with the English alphabet.

13. Therefore, the demand for Internet usage in languages other than English is growing and will continue to grow. Enabling the use of the Internet in one's native language, in which one is at ease, is important in extending the benefits of the Internet to all individual users. This is one more step toward bridging the "digital divide" — an expression commonly used to refer to the uneven global pace of progress in access to information and communication technologies.

14. It should be noted that, besides the disadvantages of using an alphabet with which they are not familiar, non-English speakers often face other issues of a more complex

⁷ See <http://www.icann.org/committees/idn/final-report-28aug01.htm> and <http://www.walid.com>

⁸ See <http://www.walid.com>

nature. For example, a Japanese person's name “博文” is transcribed as “hirofumi” in Roman letters. On the Internet, where only ASCII characters can be used, he is “hirofumi”, just like other people named “hirofumi” but whose names may use different Japanese characters such as “博史” or “宏史”. In fact, there may be over 100 different Japanese representations that will end up being denoted simply as “hirofumi” in ASCII space. Consequently, in the ASCII 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.

15. This type of problem can exist, to a lesser extent, for people using Latin-based languages — for example, in the case of people with apostrophes, accents or other diacriticals in their names. The exact forms of these names cannot be represented as domain names either, as these are restricted to Latin alphanumeric characters and the hyphen. In other words, these people’s real names are subject to mapping into a space where a much more limited set of characters are available.

16. Over time, there has been a substantial evolution in the use of non-English languages in Internet content. For example, in the case of e-mail, the following developments have taken place:

- Step 1: Expression of a native non-English language in e-mail texts using phonetic mapping from the language in question into the English alphabet (transliteration);
- Step 2: Use of native language characters in e-mail texts;
- Step 3: Use of native language characters in the subject field of e-mails.

What should the next step be? It is a natural evolution for people to want the name of the sender and receiver of e-mails to appear in their native language.

17. All machines connected to the Internet are given unique Internet Protocol (IP) addresses, which are machine-readable, (e.g. 123.4.5.67 in the case of IP Version 4). An IP address can be made more human-friendly by using the Domain Name System which provides a simple, memorable string of characters, called a domain name, synonymous with a particular IP address. With the number of services that have emerged on the Internet, the need has arisen to address more than just machines. For example, with

e-mail, we address *users* of machines. With the World Wide Web, we address the locations of *documents*. Thus, in order to facilitate communication, objects on the Internet are named by means of Uniform Resource Locators (URLs) such as <http://www.itu.int/mdns/> or e-mail addresses such as itumail@itu.int.

18. A domain name is a string of characters, such as “www.itu.int” or “www.wipo.int”, in this case referring to Internet host computers. Given that domain names were devised as easily memorable strings to be used in place of IP addresses, there is no doubt that this requirement for memorability will also exist for native languages as this is part of everyday life. Furthermore, the demand will grow for the use of other significant expressions such as company names and personal names. This means that domain names have evolved to a certain extent from simple identifiers to represent identities of entities. These days, domain names are considered equivalents to brand names, product names and service names. From a technical aspect, this is a major departure from their intended original purpose.

19. In addition to domain names, there are various other methods of naming entities on the Internet. These include, *inter alia*, search engines and directories, such as the Lightweight Directory Access Protocol (LDAP) and Common Names Resolution Protocol (CNRP)⁹. However, only domain names have become so widely and consistently used, and therefore retain a role as the preferred naming scheme for the Internet.

20. The terms “multilingual domain names” and “internationalized domain names” are often used interchangeably, although Internet engineers and operators tend to prefer “internationalized domain names.” This may reflect the view that they wish to avoid the semantics of natural languages in domain names and merely want to make it possible to use characters from all over the world in domain name scripts. However, generally this paper will use the term “multilingual”, except where “internationalized” appears as a proper noun.

⁹ See <http://www.ietf.org/html.charters/cnrp-charter.html>.

History of the Development of Multilingual Domain Names

21. One of the earliest efforts to develop multilingual domain names took place in Asia in the late 1990's. Multilingual domain names were developed at the National University of Singapore (NUS)¹⁰. Following this development, a working group on Internationalization of the DNS was formed within the Asia Pacific Networking Group (APNG)¹¹ in July 1998 to coordinate the evolution of multilingual domain names. One of the working group's projects was the development of the experimental implementation of an Internationalized Multilingual Multiscript Domain Names Service (iDNS)¹². The first phase of this project, led by the Center for Internet Research at the NUS, stated its objective as "Why shouldn't domain names be internationalized too, now that the Internet has grown to reach almost every corner of the world using different languages?". Governmental, academic bodies and industry in China, Hong Kong, Japan, Korea, Singapore, Taiwan, and Thailand, as well as Bioinformatrix Pte. Ltd., together with a number of organizations involved with the Tamil language, all participated in the project. Another project, called iDomain¹³, had the objective of creating an iDNS test bed in Asia-Pacific countries. During the 1998/1999 time frame, test bed projects were set up in several Asia Pacific countries, providing the ability to support, *inter alia*, Chinese, Japanese, Korean (Hangeul), Tamil and Thai.

22. Later that year, a prototype of a working multilingual DNS was demonstrated in Asian countries, proving its technical feasibility. In August 1998, at an International Forum on the White Paper (IFWP)¹⁴ meeting in Singapore, a multilingual domain name system was demonstrated to international delegates to the meeting who were discussing a new Internet Assigned Numbers Authority (IANA)¹⁵, including those from the

¹⁰ See <http://www.nus.edu.sg/>.

¹¹ See <http://www.apng.org/>.

¹² See <http://www.apng.org/idns/>.

¹³ See <http://www.apng.org/commission/idns/idomain/>.

¹⁴ IFWP is a meeting for the discussion of Internet governance by Internet stakeholders from all over the world. ICANN was established following a series of IFWP meetings. See <http://www.domainhandbook.com/ifwp.html>.

¹⁵ See <http://www.iana.org>.

InterNIC¹⁶ and the Internet Engineering Task Force (IETF)¹⁷. By the end of 1998, several countries had expressed an interest in implementing such a system, including China, Hong Kong, Japan, Republic of Korea, Singapore and Thailand. In several international conferences in 1999, such as the Asia Pacific Regional Conference on Operational Technologies (APRICOT)¹⁸, and INET 99¹⁹, several “Birds of a Feather” meetings (BoFs) were held to discuss multilingual domain names.

23. Following these activities, on the purely technological side, a BoF on multilingual domain names was held during the 46th IETF meeting in November 1999. The purpose was to determine whether the IETF should develop technical standards related to multilingual domain names. Mailing list discussions were immediately launched following this BoF. Three months later, at a subsequent IETF meeting in January 2000, the Internationalized Domain Name (IDN) Working Group²⁰ began work. Since that date, there has been intensive and active discussion on standardization in the IETF, principally through a mailing list and periodical physical meetings.

24. On the deployment front, at the end of 1999, several companies (including a commercial spin-off of the Asia Pacific iDNS initiative from the National University of Singapore, called i-DNS.net International Inc.²¹) began to commercialize the technology that had been developed. Several test beds of internationalized domain names rapidly emerged, including one based on i-DNS.net technology²² (see § 86 - 88 below) and one offered by VeriSign Global Registry Services (see § 104 - 108 below).

25. The Multilingual Internet Names Consortium (MINC)²³ is a major global player whose activities are not confined to deployment. Established in July 2000, with 39

¹⁶ See <http://www.internic.net>.

¹⁷ See <http://www.ietf.org>.

¹⁸ See <http://www.apricot.net>.

¹⁹ See <http://www.isoc.org/inet99/>.

²⁰ See <http://www.i-d-n.net>.

²¹ See <http://www.i-dns.net>.

²² For an example, see <http://www.verisign-grs.com/idn/>.

²³ See <http://www.minc.org>.

founding members from around the world, MINC inherited some of APNG's activities. It focuses on the promotion of the multilingualization of Internet names, including Internet domain names and keywords, the internationalization of Internet names standards and protocols, technical coordination, and liaison with other international bodies. Its vision is to give all peoples of the world their best chance to succeed in the Internet world, in e-commerce, and in the future of the digital knowledge age²⁴. In addition to this, organizations that correspond to a language, country, or region are active in pursuing the deployment of multilingual domain names. Among them are the Arabic Internet Names Consortium (AINC)²⁵, the Chinese Domain Name Consortium (CDNC)²⁶, the International Forum for IT in Tamil (INFITT) and the Japanese Domain Names Association (JDNA)²⁷.

26. On the policy side, ICANN formally embarked upon its activities related to multilingual domain names in March 2001. It considered policy coordination to be vital for the introduction of multilingual domain names based on any technology standards. Accordingly, it established an IDN working group consisting of four ICANN Board Members at its March 2001 meeting. At the same meeting, the Governmental Advisory Committee (GAC)²⁸, an ICANN advisory committee, issued a communiqué²⁹ expressing its support for multilingual domain names. The communiqué read: "With regard to international domain names (IDNs), the GAC confirms the importance and interests of this development to the benefit of Internet users worldwide". The small ICANN working group began by carrying out a "fact finding" mission based on a survey covering three aspects of multilingual domain names, namely: technical, policy, and services. The results of the survey³⁰ were reported at a September 2001 ICANN meeting. In the report, it was indicated that there was great demand for multilingual domain names. Based on these results, the ICANN Board decided to set up a committee

²⁴ See Dr. Tan Tin Wee's presentation at <http://www.minc.org/events/yokohama2000/ppt/mincworkshopJul2000.ppt>.

²⁵ See <http://www.arabicdomainname.org>.

²⁶ See <http://www.cdnc.org>.

²⁷ See <http://jdna.jp>.

²⁸ See <http://www.icann.org/committees/gac/>.

²⁹ See <http://www.icann.org/committees/gac/communique-10mar01.htm>.

³⁰ See <http://www.icann.org/committees/idn/final-report-28aug01.htm>.

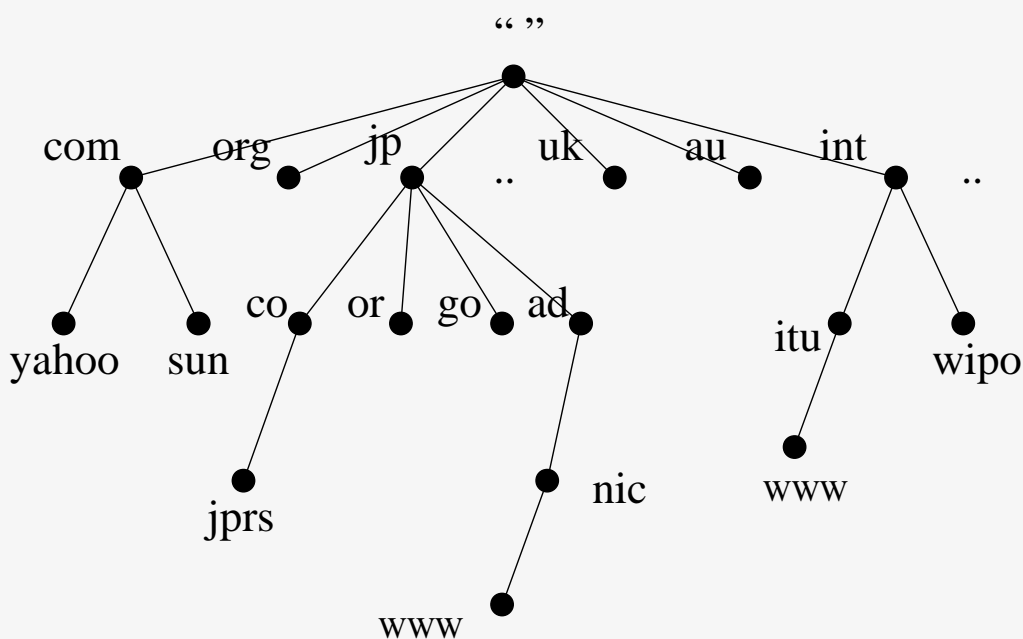
consisting of experts from various fields. This committee's mission would be to provide recommendations on non-technical policy issues, including interoperability, cybersquatting/dispute resolution, top-level domains, consumer protection and competition.

Technological Challenges to the Development of Multilingual Domain Names

Technical Aspects of the Multilingualization of Domain Names

27. The DNS domain name space has a hierarchical structure (see Figure 1 below) used to identify entities in the Internet. Each node in the structure corresponds to an entity in the Internet. A name given to a node in the structure is called a domain label. All nodes are given labels with one exception: the root node, as shown at the top of Figure 1, which has no label. The domain name of an entity (node) is a sequence of node labels starting from itself up to the root, where labels are separated by periods. As to the length, a domain label should not exceed 63 octets³¹ and an entire domain name should not be longer than 255 octets.

Figure 1: The Structure of Domain Names

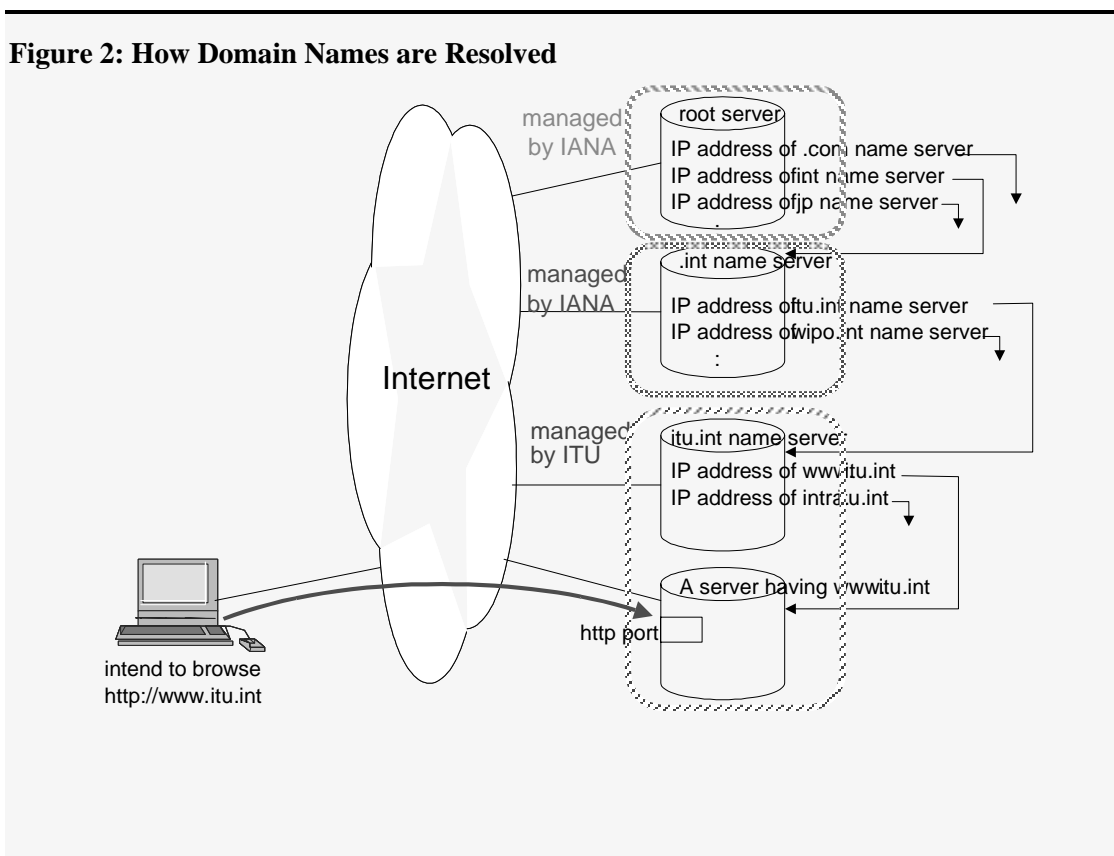


28. Figure 2 (below) shows how an entity named by a domain name is identified on the Internet. Each node of the DNS structure can be considered as a table, called a name

³¹ In computers, an octet (from the Latin *octo* or “eight”) is a sequence of eight bits. An octet is thus an eight-bit byte. Since a byte is not eight bits in all computer systems, *octet* provides an unambiguous term.

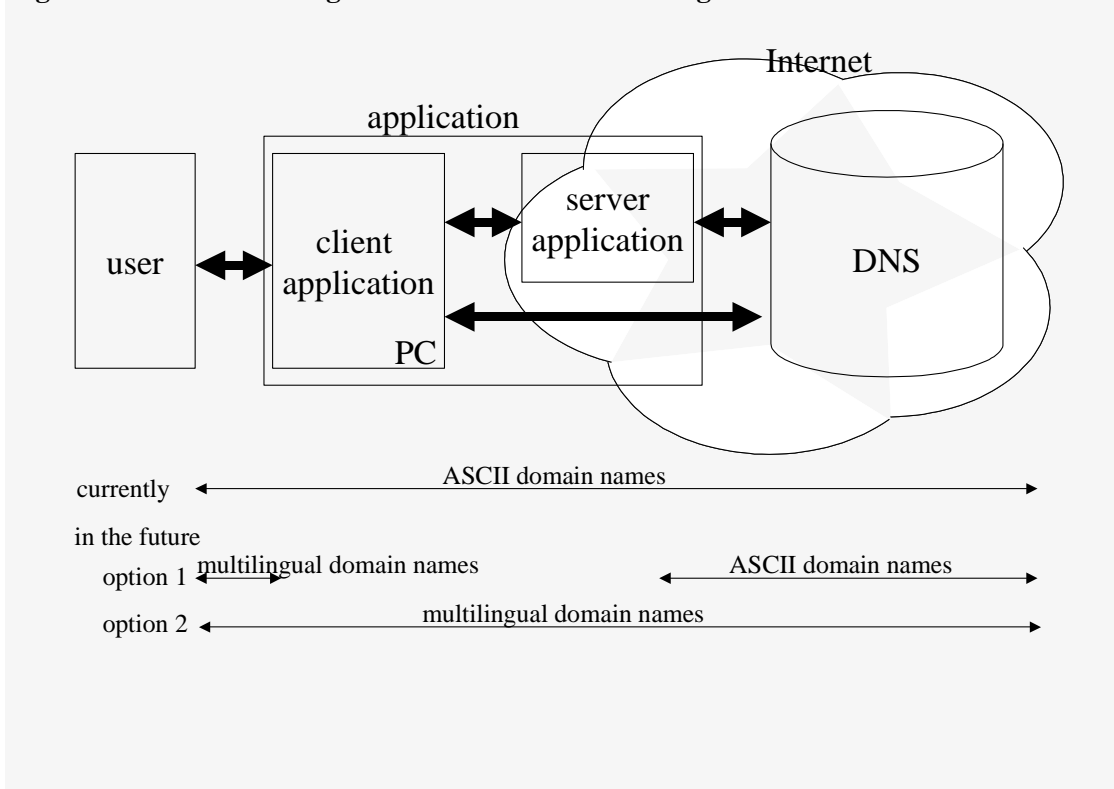
server, maintaining pairs of the node labels directly underneath the node and the corresponding IP addresses. Name servers correspond to organizations or units that are *authoritative* to manage the domain name corresponding to the node. For example, the root server is the authoritative source for the .int or .com names; the name servers for .int are the authoritative source for the .itu.int and .wipo.int names, and the name servers for .itu.int are authoritative for www.itu.int. The DNS is therefore, in effect, a large globally distributed database from both an engineering and management viewpoint.

Figure 2: How Domain Names are Resolved



29. From the standpoint of the relationship between the Internet user and the DNS, a domain name is handled as shown in Figure 3 (below). With current protocols restricted to working with ASCII, users would be forced to limit themselves to using the ASCII characters permitted in domain labels. This effectively means that ASCII domain names would be used at all points, from the user to the website. However, with the introduction of multilingual domain names, the protocol between the user and the personal computer would be based on non-ASCII characters, while the current DNS is based on ASCII.

Figure 3: Where Multilingual Domain Names are Recognized



30. The key technical questions are:

- How should non-ASCII codes be represented?
- Where should non-ASCII codes be recognized, in the client application or in the DNS server?
- What is the technical mechanism that maps multilingual domain names to current DNS technology?

The basic concepts of IETF’s work on this problem are described in § 31 - 33 below. The first question is discussed in § 34 - 37; the second is discussed in § 38 - 42; the third is discussed in § 43 - 52.

Basic Concepts of the IETF Working Group

31. As the DNS is one of the fundamental technologies deployed in the Internet, compatibility and interoperability of multilingual domain names is of critical importance. Any new technology should entail a minimal number of changes to the Internet, should coexist with the current domain names, and should allow a domain

name to consistently designate the same unique entity throughout the Internet. This is achieved by means of appropriate standardization and compliance to standards by systems in the Internet. Standardization involves establishing a common protocol that promotes interaction between entities within the Internet; in the case of the DNS, this is carried out by the IETF.

32. In January 2000, the IETF set up the IDN Working Group for the standardization of multilingual domain name technology. Its charter can be summarized as follows³²:

- The goal of the group is to specify the requirements for internationalized access to domain names and to specify a standards track protocol based on those requirements;
- A fundamental requirement in this work is not to disturb the current use and operation of the domain name system anywhere to resolve any domain name;
- The group will not address the question of what, if any, body should administer or control usage of names that use this functionality.

33. In processing the standardization of the technology of multilingual domain names, the basic requirements of the Internet Architecture Board (IAB)³³ are as follows:

- RFC 2825: Preservation of compatibility with current domain names;
- RFC 2826: Preservation of uniqueness of domain name space;
- The Internet must not be divided into islands.

Character Codes of Multilingual Domain Names

34. Only the letters of the basic Latin alphabet (non case-sensitive A-Z), the decimal digits (0-9), and the hyphen are permitted in domain names (RFC 1034³⁴ and RFC1035³⁵). Multilingualization of domain names entails the extension of this character set to include non-ASCII characters. To ensure that applications uniformly recognize

³² See http://www.minc.org/events/yokohama2000/ppt/IETF_JamesSeng.ppt.

³³ See <http://www.iab.org>.

³⁴ See <http://www.ietf.org/rfc/rfc1034.txt>.

³⁵ See <http://www.ietf.org/rfc/rfc1035.txt>.

and process the multilingual domain names, encoding and representations of such non-ASCII characters must be uniquely determined. To do this, a globally agreed-upon code set is desirable for multilingual domain names so that all applications and systems relating to domain names scattered throughout the Internet can have technical interoperability.

35. However, for various historical reasons, the fact is that many language scripts currently used in information systems have adopted national or proprietary standards. To give an example, the most popular Japanese character set used in Japanese devices is based on Japanese Industrial Standards (JIS) X 0208 and X 0201. Therefore, many PCs, personal digital assistants (PDAs), as well as Internet-enabled mobile phones in Japan can only display JIS and ASCII characters. This causes overlapping of codepoints and a lack of ability to uniquely define a type of encoding used, resulting in compatibility problems.

36. The most promising solution is the adoption of Unicode³⁶ (ISO/IEC 10646), which specifies the code sets of many scripts and therefore languages. Although Unicode may be the best current solution, it may have to be further developed to accommodate actual usage. Furthermore, where applications do not directly use Unicode for a representation of local characters, conversion of commonly used local code sets to and from Unicode is required somewhere in the computing environment (e.g. in the case of Japanese, JIS).

37. There is also the possibility that mere adoption of Unicode will not be appropriate for domain names. For example, some Chinese characters have two representations — a traditional Chinese character and a simplified Chinese character. The fact that the correspondence between a traditional Chinese character and a simplified Chinese character is not one-to-one makes the situation much more complicated. Furthermore, although they are usually used in mainland China in place of traditional Chinese characters, simplified Chinese characters are seldom used in Taiwan or Hong Kong. The point has been raised as to whether or not these two character sets should be considered as one³⁷. Some have argued that they should be treated as different characters if domain names are simply identifiers. Others argue that they should be regarded as the same characters if, in reality, domain names correspond to the identity

³⁶ See <http://www.unicode.org>.

³⁷ This is often referred to as the TC/SC equivalence problem.

of entities. Even if they are regarded as the same characters, other issues may arise in respect of whether it is merely a local code issue or a universal protocol issue; and whether a distinction should be made for such characters where used for traditional or simplified Chinese.

Client-Side Versus Server-Side Solutions

38. As regards the question of *where* non-ASCII codes should be recognized in Figure 3 on page 21, approaches to the solution of this problem are typically based on one of the following scenarios:

Client-Side Solution

39. In a client-side solution, translation between the multilingual script and the ASCII-compatible representation is performed in user applications (e.g. a Web browser). The client application translates multilingual scripts into ASCII strings, which can then be processed in the current Internet: i.e. the domain names are subsequently processed as ASCII domain names throughout the Internet. This category actually includes the case of an application that consists of both client-side and server-side software. But for the sake of convenience, the term “client-side” is used in the interest of consistency with the ICANN survey report³⁸.

40. Technically, a client-side solution is needed regardless of which approach is chosen. It is unlikely that an ASCII-only application will work immediately with multilingual domain names. Some form of upgrade will be necessary, either through provision of fonts, input methods or additional technical functionality to support internationalization.

Server-Side Solution

41. In a “server-side” solution, domain names are sent natively over the Internet by the client application in a local encoding, such as UTF-8³⁹, GB or BIG5⁴⁰, or Unicode. Applications and services communicate with each other using non-ASCII domain names all the way along the communications path between them (sometimes referred to

³⁸ See <http://www.icann.org/committees/idn/final-report-28aug01.htm>.

³⁹ See <http://www.ietf.org/rfc/rfc2279.txt>.

⁴⁰ GB and BIG5 are coding schemes for Chinese characters.

as “on the wire”). Note that the first implementations of IDN were actually proxy server solutions that intercepted local encoding from client applications and converted the encoding into an ASCII-compatible encoding so that DNS servers remained unaltered.

42. Some of the services, experiments and test beds currently deployed employ client-side, and others, server-side solutions. There is ongoing debate among technical experts as to the practical feasibility of using non-ASCII characters natively in the DNS and how this would interact or interfere with other Internet protocols. Currently, the IETF is moving towards standardization of a purely client-side solution. This is supported by the following arguments:

- First, the DNS is a huge, robust and distributed database, but one which works on the basis of a delicate balance. Too many pieces of Internet software and protocols make use of the DNS in its current form. Other than by carrying out exhaustive testing, modification of the DNS at such a fundamental level may lead to a collapse of the entire system. In view of this, many Internet engineers think it is inadvisable to modify the core of the DNS, as this may have disastrous consequences for the Internet. It is argued that a client-side solution not requiring any significant changes to the DNS is much safer for the stability and growth of the Internet.
- Second, in view of the rapidly growing demand, the ability to use multilingual domain names should be made available as soon as possible. In general, deployment of servers would take much longer than deployment of client applications. In client-side solutions, only the entities intending to communicate using multilingual domain names will need to be adapted to support multilingual domain names. Conversely, server-side solutions require that all components along the communications route, including the client, server and anything else in between, must be prepared for multilingual domain names. The deployment of a server-side solution may require reconfiguration of all of the servers throughout the Internet to accommodate multilingual scripts, which would take a considerable amount of time.
- Third, given the non-negligible time it would take to achieve server-side deployment, this approach could result in only limited areas of the Internet being able to support multilingual domain names. This might lead into separation of

the Internet into “islands” and possibly the emergence of alternative roots⁴¹. This may result in confusion and inconsistency for users. The GAC expressed its concern about this in its March 2001 communiqué supporting multilingual domain names, stating “preserving the universal connectivity and accessibility in domain name system is vital to the continuance of the Internet as a global network”.

Standardization for Compliance with the Current DNS

43. Ideally, in technical standardization, all languages and characters that could potentially be used in multilingual domain names should be taken into account. However, many issues relating to a particular language are only identifiable by those who use the languages and characters in practice. Standardization will therefore be evolutionary, as all issues involved cannot be identified and solved at one time.

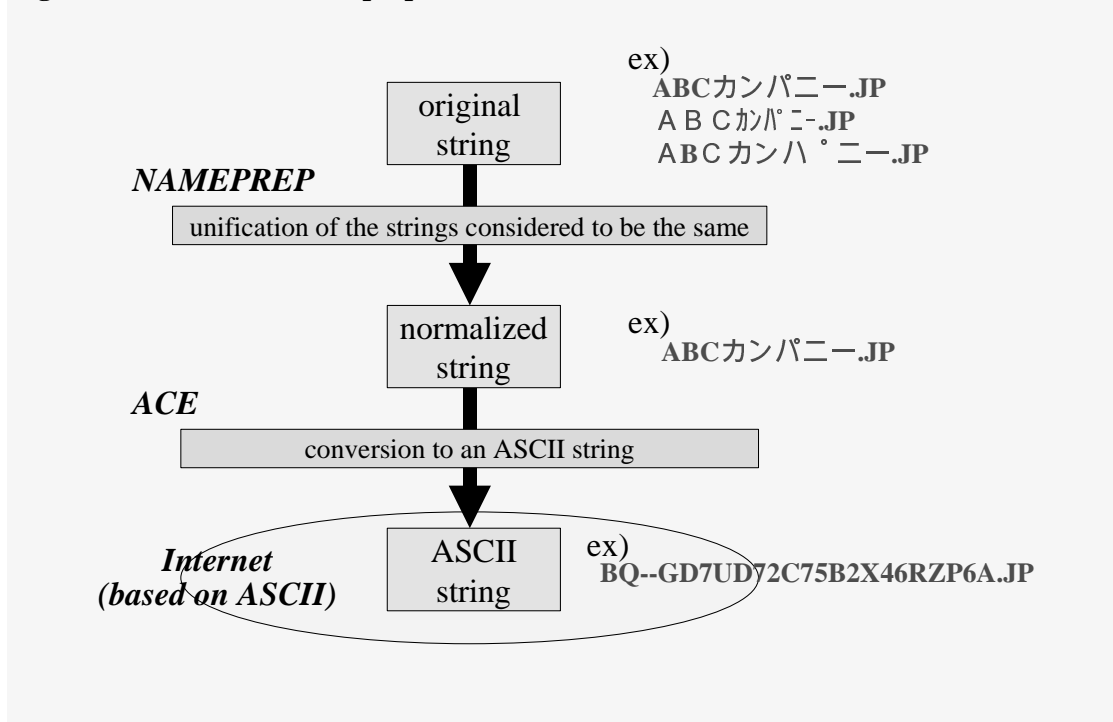
44. The IETF is currently working on standardization based on a client-side solution, as described above. The technical elements that need to be standardized include:

- Preparation of Internationalized Host Names (Nameprep);
- ASCII Compatible Encoding (ACE);
- Internationalizing Host Names in Applications (IDNA).

45. In Nameprep, multiple multilingual string representations, which technically should be regarded as the same string, are combined into one string. After Nameprep, ACE converts the multilingual representation into an appropriate ASCII domain name. The roles of Nameprep and ACE are shown in Figure 4 (below). The architecture for application software to apply these two translations to the original multilingual domain names so as to be properly incorporated into the current Internet is called IDNA.

⁴¹ Alternative root: a method of creating a separate domain name space from that of ICANN, possibly by operation of replacement or additional root servers.

Figure 4: The roles of Nameprep and ACE



Preparation of Internationalized Host Names (Nameprep)

46. The main functions of Nameprep are:

- Case folding: since the difference between uppercase and lowercase letters is insignificant in constituting ASCII-based domain names, the cases are merged or *case folded* into a single form. This needs to be done not only for ASCII letters but also for non-ASCII letters. Other types of case folding may be needed for non-ASCII characters. Case folding is also called “a map” because it maps (a) character(s) onto (an)other character(s) which is(are) regarded as equivalent. The specifications of case folding are based on Unicode Technical Report #21⁴².
- Normalization: many characters have several representations even if the human eye cannot see the difference. In domain names, these characters should be *normalized* into one representation in order to be regarded as the same character. For example:
 - the ligature “ä” and “a +” are canonically equivalent;

⁴² See <http://www.unicode.org/unicode/reports/tr21/>.

- full-width “A” and half-width “A” are equivalent.

The specifications of normalization are based on Unicode Standard Annex #15⁴³.

- Prohibition: many characters in the Unicode character set are control sequences, formatting sequences or spacing characters, which are not appropriate and *prohibited* for domain names.

The above demonstrates that Nameprep translates various representations regarded as the same original string into a unique representation in the multilingual string space. If the outputs of Nameprep are the same, input strings are regarded as the same domain name. If the outputs are different, they are regarded as different domain names. To meet this requirement, Nameprep should precede ACE. The IETF is nearing the final stages of Nameprep standardization.

ASCII Compatible Encoding (ACE)

47. ACE encodes a non-ASCII string represented in Unicode into an ASCII string, which complies with the existing ASCII domain name format. This enables multilingual domain names to be properly processed as the corresponding ASCII domain names. At the 49th IETF meeting in November 2000, the IDN Working Group was steered in the direction of choosing ACE, although arguments claiming the necessity of natively using UTF-8 have still been a matter of debate in mailing list discussions. The IETF is now reaching the final stages of ACE standardization.

48. RACE (Row-based ASCII Compatible Encoding)⁴⁴ was one of the earlier candidates among the proposed ACE algorithms. It was used in the registration and resolving services provided by, *inter alia*, VeriSign Global Registry Services (VGRS)⁴⁵ and Japan Network Information Center (JPNIC)⁴⁶ / Japan Registry Service (JPRS)⁴⁷.

⁴³ See <http://www.unicode.org/unicode/reports/tr15/>.

⁴⁴ See <http://www.i-d-n.net/draft/draft-ietf-idn-race-03.txt>.

⁴⁵ See <http://www.verisign-grs.com>.

⁴⁶ See <http://www.nic.ad.jp>.

Following RACE, other algorithms have been proposed and evaluated by engineers as to their advantages and disadvantages using actual multilingual domain names that were registered in various test bed scenarios.

49. At the August 2001 IETF meeting, an ACE system called AMC-ACE-Z⁴⁸ received significant support owing to its compression efficiency. For example, AMC-ACE-Z can represent at least 18 Japanese characters as a domain label, while RACE can represent up to 17 such characters. As one example, the ASCII output strings for “日本語ドメイン名例.JP” (meaning Japanese domain name example), produced by RACE and AMC-ACE-Z⁴⁹ respectively are:

- RACE: BQ--3BS6KZZMRKPDBSJQ4EYKIMHTKQGU7CY
- AMC-ACE-Z: ZQ--ECKWD4C7CU47R2WFQW7A0ECL32K

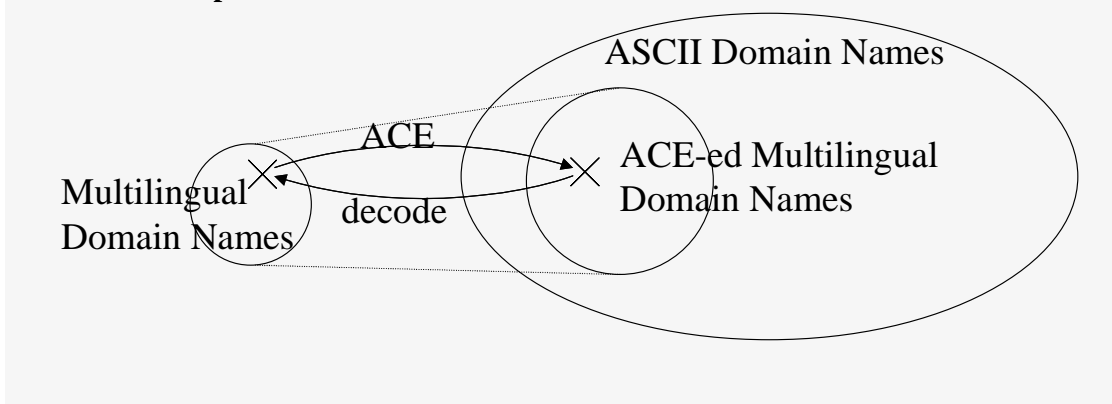
50. An ACE encoding maps multilingual domain name space into a subspace of ASCII domain names. In the reverse direction, it should be possible for the ASCII domain name using ACE to be uniquely re-mapped to a multilingual domain name. Therefore, a subspace should be reserved for multilingual domain names within the existing ASCII domain name space, as shown in Figure 5 (below). For this, a prefix, suffix or “tag” for a resulting ACE string needs to be defined. All strings having such an ACE tag will constitute a subspace defining multilingual domain names. The ACE tag has to be chosen taking into account the following conditions: there must be a 0 per cent possibility of coincidental existence of ASCII domain names with such a prefix or suffix, and the length of the prefix or suffix must be short enough to leave maximum space for multilingual domain names. Under these conditions, the prefix or suffix could be simple strings, i.e., “??--“, or “--??”, where ? is an alphanumeric character. For example, if RACE is chosen, domain names starting with prefix “bq--” would indicate a multilingual domain name.

⁴⁷ See <http://jprs.jp>.

⁴⁸ See <http://www.ietf.org/internet-drafts/draft-ietf-idn-amc-ace-z-01.txt>.

⁴⁹ The prefix of AMC-ACE-Z is assumed as “zq--” although it has not yet been specified.

Figure 5: Mapping from Multilingual Domain Name space to Subspace of ASCII Domain Name Space

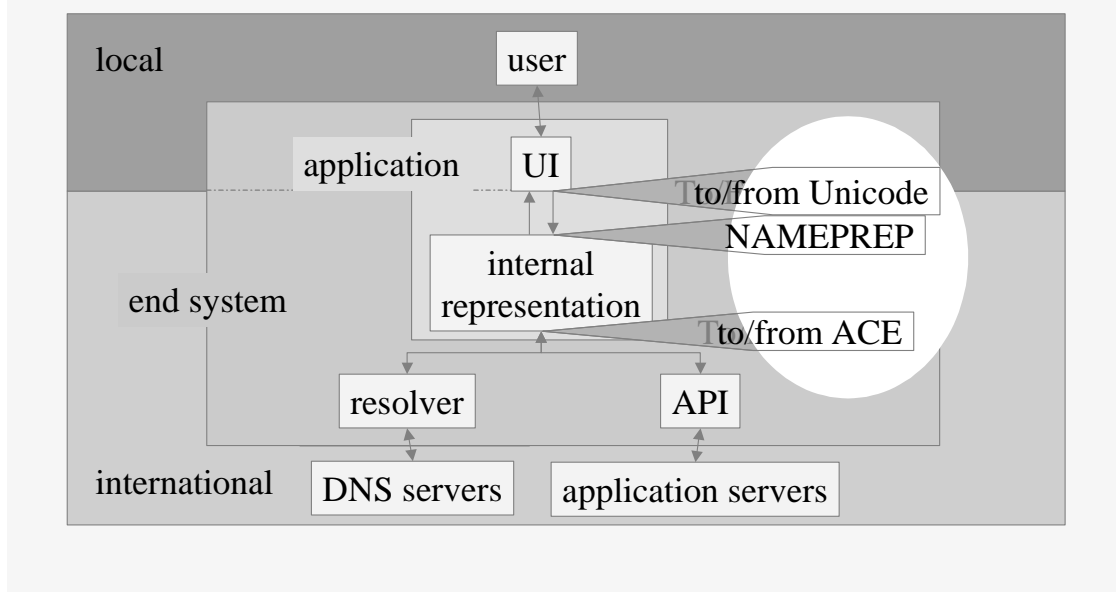


51. Although ACE is promising, a number of issues still need to be resolved. First, ASCII domain names should not be registered in the subspace reserved for multilingual domain names. For example, registration of ASCII domain names starting with “bq--” must be blocked if RACE is chosen. Second, as a domain label should not exceed 63 ASCII characters, it can only accommodate a limited number of multilingual characters — for example, 18 Japanese characters. This will restrict multilingual domain labels to shorter lengths than ASCII domain labels. In addition, deeper domain hierarchies cannot be achieved, as the length of a full domain name cannot exceed 255 characters.

Internationalizing Host Names in Applications (IDNA)

52. To use the Internet as it currently stands, translations by Nameprep and ACE should be carried out before sending the domain name “down the wire” to the DNS or application server. The application architecture in which Nameprep and ACE are performed following the mapping from local code to Unicode is called IDNA, as shown in Figure 6 (below). At the August 2001 IETF meeting, many attendees supported the IDNA client-side solution.

Figure 6: The architecture of IDNA



Impact on the DNS Structure

53. A basic requirement of the DNS is the ability to identify entities on the Internet. To meet this requirement, the structure of the hierarchical domain name space must be administratively coordinated. This is currently performed by ICANN with final oversight by the US Department of Commerce⁵⁰. This means that the authority of the DNS hierarchy root shown in Figure 1 on page 19 is generally ICANN. This root is sometimes called the *authoritative root*.

Alternative Roots

54. An increasing number of software solutions now offer so-called *alternative root* systems. These encapsulate the public DNS and extend it by offering additional top-level domains, thereby enabling Internet users to view domain names other than those recognized by ICANN. Unless there is some sort of global administrative coordination of top-level domains⁵¹, this could result in a fragmentation of the Internet into disparate name spaces.

⁵⁰ The stated policy of the US Administration has been to transfer management of the DNS to ICANN. In practical terms, *inter alia*, this would entail transferring both policy and technical control of the authoritative domain name system server, where existing or new top level domains are defined and maintained, to ICANN or its subsidiary, IANA. On later occasions, the US Department of Commerce has stated that they have “no plans to turn over policy control of the authoritative root server” (see <http://www.gao.gov/new.items/og00033r.pdf>). Currently, the primary root server, “a.root-servers.net”, is maintained by VeriSign Global Registry Services, a subsidiary of VeriSign, Inc. (<http://www.verisign-grs.com>), located in the United States of America. The final authority for change control of the *root zone file* (e.g. addition, modification or deletion of top level domains) is held by the US Department of Commerce. See *Cooperative Agreement No. NCR-9218742, Amendment 11*, (Oct. 6, 1998) where it is stated: “While NSI continues to operate the primary root server, it shall request written direction from an authorized USG official before making or rejecting any modifications, additions or deletions to the root zone file. Such direction will be provided within ten (10) working days and it may instruct NSI to process any such changes directed by NewCo when submitted to NSI in conformity with written procedures established by NewCo and recognized by the USG.” See <http://www.ntia.doc.gov/ntiahome/domainname/proposals/docnsi100698.htm>.

⁵¹ Note that there does not necessarily have to be technical coordination.

55. In response to this concern, ICANN has recently issued position papers⁵² arguing the need for a unique authoritative public DNS root, which should be managed as a public trust, and asserting that ICANN has assumed this public trust role. There is general agreement among technical experts that a unique public name space is necessary in order to maintain the integrity and global connectivity of the DNS. Here, a related statement of the Internet Architecture Board (“IAB”), documented in RFC 2826⁵³, is worth citing:

“To remain a global network, the Internet requires the existence of a globally unique public name space. The DNS name space is a hierarchical name space derived from a single, globally unique root. This is a technical constraint inherent in the design of the DNS. Therefore it is not technically feasible for there to be more than one root in the public DNS. That one root must be supported by a set of coordinated root servers administered by a unique naming authority”.

56. While the arguments stem from a variety of different perspectives as well as economic interests, there appears to be general agreement on the need for a DNS name space visible to a maximum of Internet users: a severely fragmented name space is of little value to anyone. As evidence, the managers of “unsanctioned” top level domains in alternative root systems have argued both a) for inclusion in the “authoritative root” and b), against ICANN introducing TLDs identical with their TLDs used in alternative inclusive roots. They also contend that it is possible to have an administratively coordinated root function that avoids collision between different top-level domains based on multiple root systems. This suggests that the debate remains more about *who* is the *root* or *coordinating naming authority* rather than about the merits of a single coordinated name space.

Multilingual Domain Name Resolution by Alternative Roots

57. Multilingual domain names cannot be supported by existing standard specifications. The deployment of multilingual domains with proprietary technology could encourage

⁵² See <http://www.icann.org/stockholm/unique-root-draft.htm>, <http://www.icann.org/icp/icp-3-background/lynn-statement-09jul01.htm>, and <http://www.icann.org/icp/icp-3.htm>.

⁵³ <http://www.ietf.org/rfc/rfc2826.txt>.

the emergence of alternative roots. From the user's perspective, this could result in one domain name referring to completely different entities in different name spaces under different root structures. In particular, because it is an extremely long process to introduce new top-level domains, there is some question as to whether the market will simply overtake the current administrative arrangements.

58. One argument put forward by proponents of alternative roots for the resolution of multilingual domain names is that ICANN's authority is principally drawn from the United States, having historically been considered the source of ASCII-based Internet domain names. It is argued that, as multilingual domain names originated elsewhere, alternative roots supporting multilingual top-level domains may be more acceptable than some contend. Other proponents support the concept of an "inclusive" root, which allows for top-level domains not under ICANN's authority to be used for national or commercial deployment. In this case, as long as users point their applications to the inclusive root, they will be able to resolve ICANN domain names as well as non-ICANN domain names — giving direct access to new multilingual top level domains. Again, some see problems with this model in that there may be more than more than one party arguing that it manages the "inclusive root". This could lead to name space collisions that would need to be resolved by negotiation, arbitration, or possibly litigation. In the worst case, this may lead to fragmentation of the Internet name space as forecast by the IAB in RFC 2826.

Pseudo-Roots

59. There is a somewhat more subtle way to create a multilingual domain name space. This is achieved by making an 'imaginary non-ASCII top-level domain' in the authoritative domain name space. This method, called *zero level domain*, was suggested in IETF draft documents as early as 1997. It conceals the upper part of the domain name space, assuming one top node of the unconcealed space as a virtual top level domain, and using the subspace governed by the virtual top level domain as the entire domain name space. For example, after creating a space {non-ASCII-string}.TLD under the authoritative top level domain '.TLD', users can access the Internet by using domain names like xxx.{non-ASCII-string} if the users' client application automatically detaches and/or re-attaches '.TLD' with each access to the Internet. This can make a (virtual) multilingual top-level domain for users of such client applications. Even if zero level domains are somewhat more acceptable than alternative roots, users still need to

be conscious of the problem that different entities may apparently be designated by the same domain name if different client applications are used.

60. It is not multilingual domain names *per se* that lead to the creation of alternative or pseudo roots. Rather, it is the combination of commercial interests and user demand for early deployment of new TLDs; whether in English or multilingual scripts. If policies for the creation of new TLDs are able to meet user and commercial demands, the risk of fragmentation is greatly reduced. This suggests that it is extremely important that ICANN find methods to address this demand effectively.

Policy and Coordination Issues Raised by Multilingual Domain Names

61. Technology is always the start of a process, not the end. Before a technology can be fully employed, it needs to be supported by policy and business. This section discusses the major policy issues related to multilingual domain names.

Consideration of Multilingual Domain Names in Various TLDs

62. In the present ASCII-based DNS, there are two basic kinds of top-level domains: generic top-level domains (gTLDs), such as .com and .info, and country code top-level domains (ccTLDs), such as .uk and .jp. There are less than 15 gTLDs, and their policies are, for the most part⁵⁴, controlled by ICANN. There are currently about 245 ccTLDs⁵⁵, and the policies of each are, for the most part, controlled by a ccTLD management organization, typically in the respective country or region⁵⁶.

Potential Types of Multilingual Domain Names

63. Several kinds of multilingual domain names may emerge, depending on the kind of TLDs they come under or represent. They could be same-language, same-script, or mixed-language, mixed-script, multilingual domain names. These might be represented as follows:

- {non-ASCII-string}.{ASCII-ccTLD};
- {non-ASCII-string}.{ASCII-gTLD};
- {any-string}.{non-ASCII-ccTLD};
- {any-string}.{non-ASCII-gTLD}.

64. The above notation is not formally defined here, as it is sufficient to have a grasp of the underlying principles. Furthermore, it is entirely possible that other types of multilingual TLDs could emerge. For example, language-related TLDs that indicate the

⁵⁴ Some 'gTLDs', such as .mil, .gov, and .edu, are clearly not under policy control of ICANN.

⁵⁵ See <http://www.iana.org/cctld/cctld-whois.htm>.

⁵⁶ However, there are a significant number of cases where management control of a ccTLD is outside the related country or territory.

language of the associated domain names: for example, {Chinese string}.{CHINESE} or {Japanese string}.{JAPANESE}, where “CHINESE” and “JAPANESE” represent the Chinese and Japanese characters for the name of the language.

Technical and Non-Technical Issues

65. While obstacles to implementation of these multilingual domain names are mainly non-technical ones, a potential technical hurdle is the increased load on the DNS. This is because a {non-ASCII-string} is unusually long when encoded into an ACE format. Other technical hurdles include the necessity of multilingualization of related systems such as the *Whois* system, an application that displays associated attributes of domain names (e.g. registrant information). Non-technical obstacles, on the other hand, include:

- issues related to responsibility for domain name registration;
- issues to be resolved in the process of registration and usage.

The second of these obstacles will be discussed in subsequent sections. The first is described in this section by classifying the issues based on the kinds of top-level domains under consideration.

Mixed Multilingual.ASCII Domain Names

66. A number of organizations are already operators with regard to {non-ASCII-string}.{ASCII-ccTLD} and {non-ASCII-string}.{ASCII-gTLD}. For example, VGRS is offering {Chinese-string}.com registrations⁵⁷, and JPNIC/JPRS is offering {Japanese-string}.jp. These services are provided on the basis that the organization involved has “authority” over a ccTLD or gTLD and, if the DNS is internationalized, that authority is sufficient grounds to delegate {non-ASCII}.{ASCII} multilingual domain names under the corresponding TLD.

Multilingual.Multilingual Domain Names

67. One example of {non-ASCII-ccTLD} is “.日本” (“日本” represents “Japan” in Japanese Kanji). If a {non-ASCII-ccTLD} and its management organization are

⁵⁷ It should be noted that objections have been put forward by the People’s Republic of China concerning this service.

coordinated with ICANN, there may not be a problem regarding authority decisions as long as there is no dispute as to that organization being the legitimate authority. In the case of Japanese, therefore, as the seat of the language is in Japan, and where no other country has designated the Japanese language as its official language, that decision appears to be clear-cut. However, it should be noted that the same Japanese characters “日本” are also used in the Chinese character set and their glyphs are identical. Those particular characters normally could also not be designated as a Chinese TLD and assigned to another organization. The Japanese language also uses two other scripts, namely Katakana and Hiragana, but as other countries do not use these scripts, they are unlikely to give rise to complications.

68. For other languages, the issues will be much more complex. If a country or region corresponding to a country code has two or more official languages, it may need to decide in which language is used to represent its country “code”{non-ASCII-ccTLD}, assuming that “country code” has an equivalent in that language. Even if a rule is established that two or more {non-ASCII-ccTLD}s can be assigned to one country or region, the issue arises as to the number of {non-ASCII-ccTLD}s to be assigned to the country or region for however many languages are official or used in that jurisdiction. For example, in the case of India, there are more than 20 commonly used languages, each with their own script.

69. An example of {non-ASCII-gTLD} is “.企業” (“企業” is a traditional Chinese character string meaning “a company”). One problem is that multiple languages may share characters. Because of this, identical strings may represent the same or different meanings in different languages. Also, similar characters exist in different languages. For example, both China and Japan use the word “企業”, so people cannot tell whether the top level domain “企業” is in Chinese or Japanese. In other words, multilingual domain names may confuse people in spite of the stated goal to make domain names more memorable. It is very difficult to decide who should be designated to manage these kinds of top-level domains (and in which country). Given the difficulties experienced for simply introducing new ASCII top-level domains, it is not hard to imagine the challenges that will be involved when introducing multilingual top-level domains.

What are the Languages that Constitute Multilingual Domain Names?

70. One of the issues that should be examined is the definition of languages from the viewpoint of multilingual domain names. Some languages have two or more kinds of scripts, and some languages have mixed scripts in the written form of the language. For example, Japanese written documents may mix Chinese Han characters, Japanese Katakana and Hiragana, Arabic numbers, as well as the English alphabet. In this case, can all the possible strings in a Japanese written document be multilingual domain names? In which language are Chinese Han characters when used as a multilingual domain name in a Japanese document?

71. In addition, local rules such as the unification of traditional Chinese characters and simplified Chinese characters, as described in § 37, will need to be addressed: even from the perspective of “whether they are the same language or different languages.” For example, would “folding” (see § 46) of traditional and simplified Chinese Han characters affect the usage of Han characters in other non-Chinese languages?

Who is the Language Authority for Multilingual Domain Names?

72. A further question is whether the issues described in § 70 - 71 are local issues or international issues. In the interest of eliminating confusion for the users, some advocate that the rules with respect to multilingual domain names should be the same even if they are under different top-level domains. Therefore, a single domain name registry⁵⁸ should not be the ultimate authority for the rules for multilingual domain names. As an example, should the representation rules and conversion rules for Chinese domain names in .com and in .cn⁵⁹ be the same? In this example, the rules definition for Chinese multilingual domain names would inherently be an international issue. However, should the international community that does not use the Chinese language be able to define localization issues for Chinese speaking people? And as the Chinese language is diasporic, used in different jurisdictions, countries and economies, how localized are these decisions?

⁵⁸ The registry of a domain name is an organization that is responsible for managing the registration of domain names under the domain name. For example, the registry of .com is VGRS.

⁵⁹ .cn is the ISO 3166 alpha-2 country code for the People’s Republic of China.

73. It is extremely difficult if not impossible, for those whose language is not concerned by this discussion to comprehend the sensitivities involved. Understanding whether the issues in § 70 - 71 are code problems or protocol problems is very difficult. But this understanding is necessary to lead to an acceptable decision as to what extent such issues need to be standardized internationally. Someone must decide which issues exist and how they are to be resolved. Perhaps a pragmatic first step is resolving who is the likely relevant decision-making authority.

Matrix of Authority

74. So far, a number of combinations of country/economy, language, script, and encoding systems have emerged and examples are listed in Table 1. Table 1 suggests that a “one size fits all” policy approach is very unlikely to succeed.

Table 1

Script	Language	Encoding	Country/Economy	Comment on Administrative Model
Chinese Traditional and Simplified	Chinese	GB BIG5 HW	China, HongKong, Taiwan, Macau, Malaysia, Singapore USA, Canada, UK, etc.	Diasporic language Official language of several economies Chinese Domain Name Consortium (CDNC)?
Hiragana Katakana Kanji	Japanese	JIS SJIS EUCS	Japan	>90% Japanese speakers in Japan JDNA/JPRS/JPNIC are obvious candidates Kanji needs coordination with CJK countries
Hangeul	Korean	KSC	People’s Republic of Korea (South) Democratic People's Republic of Korea (North)	>80% Korean speakers in Koreas KRNIC is a potential candidate Hanji needs coordination with CJK countries

Script	Language	Encoding	Country/Economy	Comment on Administrative Model
Arabic	Arabic Urdu Farsi Jawi		Algeria, Bahrain Djibouti, Dubai Egypt, France Jordan, India, Iraq Iran, Kuwait Lebanon, Libya Morocco, Malaysia Mauritania, Oman Palestine, Pakistan Qatar, Saudi Arabia Spain, Somalia Sudan, Syria Tunisia, Turkey UAE, Yemen and others	Diasporic language Multi-Country official language Arabic Internet Names Consortium (AINC) Arabic Languages WG, MINC Urdu Language WG, MINC
Tamil	Tamil	TAM TAB TSCII Many other proprietary fonts	India (Tamil Nadu state), Mauritius, Sri Lanka, Malaysia, Singapore, USA Canada, UK, etc.	Diasporic language minority in all countries Official language in a few Tamil Nadu State in India is recognized as seat of Tamil Language International Forum for IT in Tamil (INFITT) Working Group WG02
Thai	Thai	TSC	Thailand	>90% of Thai speakers in Thailand
Khmer	Khmer	Many proprietary fonts	Kingdom of Cambodia Thailand (Surin) Vietnam	>90% of Khmer speakers in Cambodia Official language in one
Lao	Lao	A few proprietary fonts	Lao PDR Thailand	10 times more Lao speakers in Thailand
Cyrillic	Russian		Russia and about a dozen other former USSR republics	>90% in Russia Russia recognized as seat of Russian language
Hebrew	Hebrew		Israel	>95% in Israel

Models for a Matrix of Authority

75. The table above suggests that it will be important for language stakeholders to coordinate among themselves. Where needed, regional or international organizations may be appropriate forums. Generally, as a matter of principle and where possible, it seems appropriate that decisions affecting language users should be made by the language users themselves. Table 2 suggests some of the models that may need consideration.

Table 2

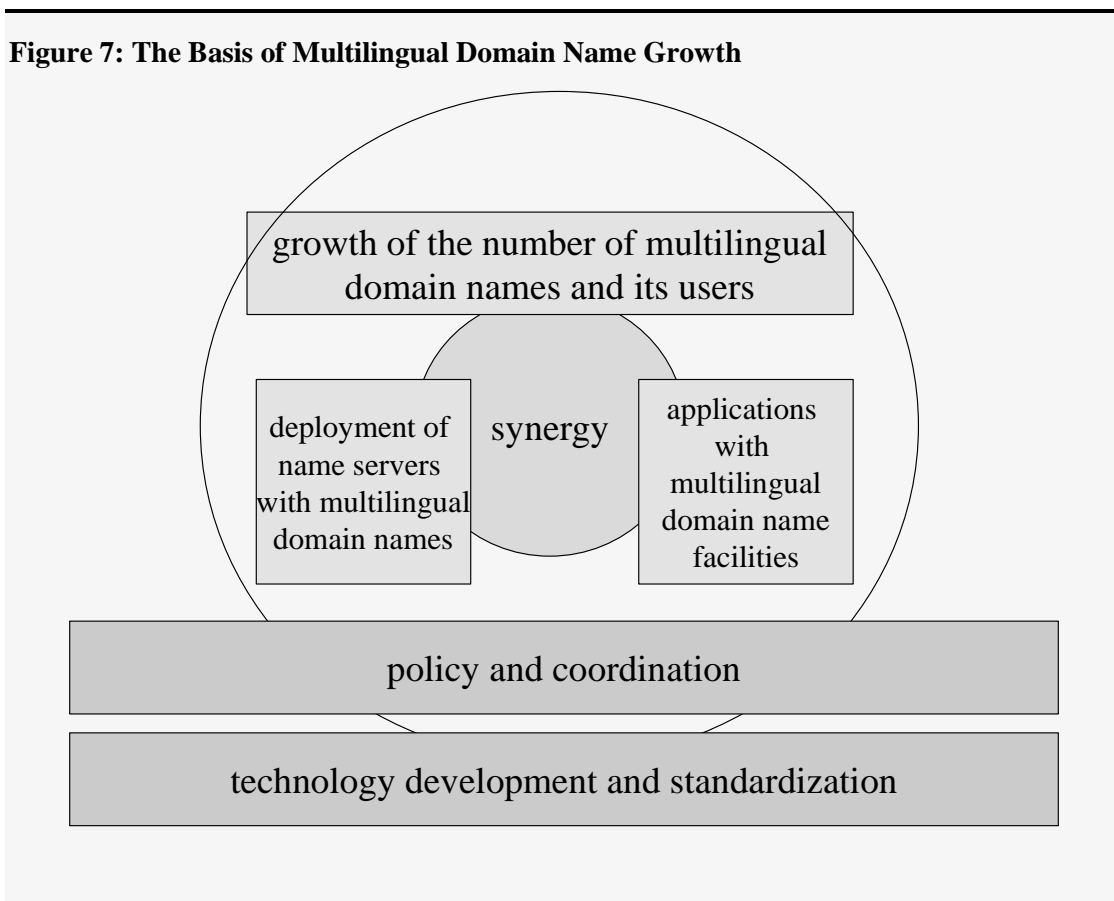
Model	Language
One language-one script-one country model	Hebrew, Thai, Russian
One language-one script-no country model	Tamil
One language-one script-many countries model	Arabic, Lao
One script-many languages-many countries model	Arabic-Urdu-Farsi-Jawi system, Han
One language-many scripts-one country model	Japanese, Korean
One language-many scripts-many countries model	Chinese (TS-SC), Urdu (Arabic-Hindi)
One country-many scripts-many languages	Many countries

Summary

76. To make multilingual domain names fully usable on the Internet, technical standardization will be but the tip of the iceberg. In order to meet user requirements, it will be necessary to also complete the following steps:

- standardization of technology;
- policy and coordination of registration and management rules;
- deployment of applications and name servers.

The relationship between these steps, necessary for deployment of multilingual domain names, is illustrated below in Figure 7.



77. Concerning technical standardization, standardization of Nameprep, ACE, and IDNA (see § 43 - 52) is expected to be completed in the first half of 2002, according to the proposed milestones of the IDN Working Group. However, as all languages of the

world have yet to be considered, the specifications of the standard will necessarily need to further evolve. In addition, as the DNS itself is evolving, longer-term solutions such as server-based solutions or additional software layers may emerge (e.g. keywords) and prove to offer better solutions.

78. The policy and coordination issues discussed in § 61 - 75 will need to be resolved in the very near future. However, with national, regional and international cooperation, solutions can be found.

79. The deployment of applications and name servers must rely on the dynamics of the business sector. In order to achieve satisfactory usage, it is important to promote deployment of both servers and applications. It is vital that application development be catalyzed and widely promoted. As a practical example, the Japanese Domain Names Association (JDNA), established in July 2001, has Japan-based members such as application vendors, network service providers, and domain name registries. Within JDNA, local necessary specifications such as detailed representation of URLs and e-mail addresses can be determined.

80. To summarize, there is substantial market and user demand for multilingual domain names. To satisfy this demand, the entire environment will need to be developed to take into account technology standardization, policy and administrative arrangements, as well as new applications. The future of multilingual Internet names is imminent. We should not underestimate the significance of this activity, as it is part of a far nobler goal: the ongoing internationalization of the Internet.

Annex A: Glossary of Acronyms

ACE	ASCII Compatible Encoding
AINC	Arabic Internet Names Consortium
AMC-ACE-Z	Adam M. Costello-ASCII Compatible Encoding-Z (26 th Version)
APNG	Asia Pacific Networking Group
APRICOT	Asia Pacific Regional Conference on Operational Technologies
ASCII	American Standard Code for Information Interchange
BoF	Birds of a Feather meeting
ccTLD	Country Code Top Level Domain
CDNC	Chinese Domain Name Consortium
CNNIC	China Internet Network Information Center
CNRP	Common Names Resolution Protocol
DNS	Domain Name System
GAC	Governmental Advisory Committee
gTLD	Generic Top Level Domain
HKNIC	Hong Kong Network Information Center
HTTP	Hypertext Text Transfer Protocol
IAB	Internet Architecture Board
IANA	Internet Assigned Numbers Authority, part of ICANN
IC	Identification Code
ICANN	Internet Corporation for Assigned Names and Numbers
IDN	Internationalized Domain Name
IDNA	Internationalizing Host Names in Applications
iDNS	Internationalized Domain Names Service
IETF	Internet Engineering Task Force
IFWP	International Forum on the White Paper
INET	Internet networking
INFITT	International Forum for IT in Tamil
IP	Internet Protocol
ISOC	Internet Society

ITU	International Telecommunication Union
JDNA	Japanese Domain Names Association
JIS	Japanese Industrial Standard
JPNIC	Japan Network Information Center
JPRS	Japan Registry Service
KRNIC	Korea Network Information Center
LDAP	Lightweight Directory Access Protocol
LDH	Case insensitive letters-digits-hyphen used in the DNS
MPHPT	Ministry of Public Management, Home Affairs, Posts and Telecommunications
MINC	Multilingual Internet Names Consortium
MONIC	Macau Network Information Center
MoU	Memorandum of Understanding
NIC	Network Information Center
NUS	National University of Singapore
PC	Personal Computer
RACE	Row-based ASCII Compatible Encoding
TLD	Top Level Domain
TWNIC	Taiwan Network Information Center
UDRP	Uniform Dispute Resolution Policy
URL	Uniform Resource Locator
VGRS	VeriSign Global Registry Services
WIPO	World Intellectual Property Organization

Annex B: Some Implementations of Multilingual Domain Names

81. Market demand often does not wait for technically perfect solutions, which is why some implementations of multilingual domain names have already emerged. Currently, implementations typically rely on proprietary technology or incomplete standards specifications. However, many solution providers have stated that they will comply with any future standards once standardization has been completed. Some of the known implementations in the market are listed in alphabetic order below. As many multilingual domain name solution providers use Internet keyword technologies for resolution services, companies focused in this area are also listed. The information provided is, in most cases, provided by the solution provider. As developments take place rapidly in this area, this list is by definition incomplete. Further information or clarification on solutions offered in the market is solicited.

Chinese Domain Name Consortium (CDNC)

82. On May 19th, 2000, Chinese domain name consortium (CDNC) was set up in Beijing by four Network Information Centers (NICs) around the Taiwan Strait, who are China Internet Network Information Center (CNNIC), Taiwan Network Information Center (TWNIC), Hong Kong Network Information Center (HKNIC) and Macau Network Information Center (MONIC). As an independent non-profit organization, CDNC will mainly take charge of the coordination and regulation of Chinese domain names worldwide. Since the domestic domain name plays a more and more important role in China, plenty of organizations and companies have shown interest and are actively joining in the research and popularization of Chinese domain name. However, because of the lack of communication and coordination between them, there are very many differences in approaches and technologies to support a Chinese domain name system, which would heavily delay popularization. To avoid these problems, the four NICs advocated and finally set up CDNC and will improve the coordination and cooperation of Chinese domain names.

83. CDNC will evaluate all Chinese domain name resolution issues, strictly complying with international criteria, making the technical standards for Chinese domain names and the corresponding regulations for Chinese domain name registration. It also coordinates its running in the other countries or regions, communicates and cooperates with all corresponding international organizations so that CDNC can make international standards in near future.

China Internet Network Information Center (CNNIC)

84. CNNIC⁶⁰ provides trial Chinese domain name registration using technical solutions based on internationalized domain name technical requirements and Chinese domain name users' requirements.

85. The resolution is “server-side” using HTTP forwarding. They also provide a keywords client download for resolution.

i-DNS.net

86. i-DNS.net⁶¹ is an Internationalized Domain Name (IDN) solutions provider and registry for {Native-Character}.{Native-Character} domain names. The generic top-level domains (gTLDs) supported by i-DNS.net are local language versions of .com, .net and .org, selected in consultation with local Network Information Centres (NICs) and in-country linguistic experts.

87. All names registered and hosted in i-DNS.net’s registry database are compatible with, and enjoy full and total delegation under, the existing DNS. These names are globally resolvable via a wide range of resolution methods, including the popular iClient software - a Windows-based client-side resolution plug-in.

88. i-DNS.net’s IDN offerings are compliant with the recommendations and standards promulgated by the IDN Working Group of the Internet Engineering Task Force (IETF), that is - client-side, Nameprepped and ACE-based. Through its registrar and strategic partners, i-DNS.net has launched its registration services across the globe in more than 30 languages.

Japan Network Information Centre (JPNIC) / Japan Registry Services (JPRS)

89. JPNIC⁶²/JPRS⁶³ provides registration and resolution services for Japanese domain names⁶⁴ with a client-side solution using almost the same technology as VGRS (see §

⁶⁰ See <http://www.cnnic.net.cn/>.

⁶¹ See <http://www.i-dns.net/>.

⁶² See <http://www.nic.ad.jp/>.

104-108 below). JPNIC/JPRS accepts Japanese script multilingual domain names under its ccTLD .jp and charges the same amount for the registration of multilingual domain names as for ASCII domain names.

90. The following functions are provided :

- They use RACE (and in the near future ACE-AMC-Z) to encode Japanese domain names into ASCII strings;
- They set the ASCII strings to the ordinary DNS name servers as domain names;
- They provide development kits for applications such as web browsers to make it possible for them to refer to DNS with Japanese domain names;
- Over 60'000 domain names have been registered and can be used. In addition, RealNames keyword resolution technology is used, similarly to VGRS.
- Before 'first come, first served' registration, JPNIC/JPRS conducted some defensive measures such as prefix blocking, reserved words, and a sunrise period in order to avoid problems related to intellectual property and false starts.

Korea Network Information Center (KRNIC)

91. KRNIC⁶⁵ has taken experimental registrations of {Korean-string}.test.kr and {Korean-string}.실험.kr between March 16, 2001 and April 25, 2001 to test the feasibility of deploying Hangeul domain names.

92. KRNIC implemented the following for the services:

- used BIND 8.2.3 with a few modifications;
- uploaded the zone files in EUC-KR, UTF-8 and RACE format;

⁶³ See <http://jprs.jp/>.

⁶⁴ See <http://jprs.jp/eng/GUJP-Eng.files/frame.htm>.

⁶⁵ See <http://www.nic.or.kr/>.

- responded to queries with IP addresses directly;
- developed several standards under RFC-KR such as the second level domains;
- developing several more standards under RFC-KR now;
- started testing multilingual TLDs;
- developing Nameprep for Korean characters.

KRNIC will undergo further tests and decide when to begin the formal registration of Hangeul domain names.

NativeNames

93. NativeNames⁶⁶ offers Arabic, Farsi, Urdu and Cyrillic name equivalents of the existing gTLDs .com, .net and .org, as well as offering the equivalent of new TLDs in these languages⁶⁷. According to Pyramid Research, NativeNames is the market leader in a rapidly growing Arab Middle East internationalized DNS market⁶⁸ and concludes that the “growing presence of Arabic character domain names stands to boost Internet adoption across the Middle East and North Africa”.

Neteka

94. Neteka⁶⁹ is not a registry or registrar itself, but provides a solution for multilingual domain names that is a combination of server-side and client-side solutions. The solution provides for registration of {non-ASCII-string}.gTLD, {non-ASCII-string}.ccTLD, and {non-ASCII-string}.SLD, where SLD means a second level domain.

⁶⁶ See <http://www.nativenames.net>

⁶⁷ See <http://www.nativenames.net/english/technology/tld.asp>

⁶⁸ Pyramid Research, The Economist Intelligence Unit, “Arab Middle East: Native Domain Names Selling Rapidly,” April 19, 2001.

⁶⁹ See <http://www.neteka.com/>.

Netpia

95. Netpia is a provider of Internet Keyword services. The heart of Netpia's Internet keyword service is that people can access Internet web sites in their own native languages without remembering the cumbersome English Domain Name.

96. Multilingual Internet Keyword Name is a next-generation domain name system, a proprietary solution developed by Netpia.com in 1997. The new system's primary strength is to support current Internet address system (DNS) while allowing multilingual recognition (MSS: Multilingual Scan System) system. The dual support marks a new paradigm in the fast-evolving Internet environment. With traditional country-to-country barriers falling fast due to digital revolution, Netpia plans to expand its multilingual and keyword-based Internet domain business to other countries where English is not an official language, a source of fresh business opportunities for Netpia.

97. Netpia is expected to standardize multilingual Internet keywords. As native language becomes ccTLD, so does Internet keyword ccTLD. As a result, .kr, .jp, .cn would not be necessary when one is surfing the Web. Netpia's vision is to allow people to surf the Web in their own languages by localizing the Internet address system.

New.net

98. New.net is a market-based domain name registry and registrar operating more meaningful, descriptive domain names in multiple languages. In 8 months, New.net has built a voluntary network of 73 million Internet users who can access and resolve domain names in 6 different languages. New.net has released 30 English language extensions including .shop, .family, .mp3 and .club and translated extensions for the Spanish, Portuguese, French, Italian and German speaking communities, such as .tienda, .reise and .amor.

99. To enable users to access these domain names, New.net forms partnerships with ISPs who make minor changes to their nameserver software. All the customers of New.net's partner ISPs are then enabled to see New.net domain names. For those who do not connect to the Internet using a New.net partner ISP, there is a small downloadable plug-in for enabling individual users' PCs.

100. New.net will be releasing a IDNA solution in the first quarter of 2002 that will be compliant with the standards promulgated by the IDN Working Group of the Internet

Engineering Task Force for resolution of IDN.IDN domain names. New.net will also be taking registrations and accrediting registrars to encourage the registration and use of these domain names.

101. New.net's stated intention is "to continue to work with the existing DNS to provide practical solutions to Internet naming for Users around the world. This includes investigating longer-term server side solutions to the issue of IDN resolution."

RealNames

102. RealNames Corporation is a global infrastructure provider of Keywords, a superior Web naming and navigation platform that improves on the existing Domain Name System. Keywords replace complicated URLs with simple names and brands, and work in the consumer's native language, making the Internet easier to use. Founded in 1996, RealNames is based in Redwood City, California with offices in London, Tokyo and Seoul.

103. The RealNames Keyword Resolution service works across all Internet-enabled devices, and many applications and services. It has been integrated into Microsoft's Internet Explorer browser and Openwave Systems Mobile Access Gateway, as well as in leading search and portal sites.

VeriSign Global Registry Services (VGRS)

104. VGRS⁷⁰ is currently offering an Internationalized Domain Name (IDN) test bed that presently provides registration and resolution services for multilingual domain names using a client-side solution. In the VGRS test bed, only the second level domain is internationalized; the native language domain is followed by the ICANN authorized TLD .com, .net or .org to form a mixed language domain name. VGRS accepts more than 39 Unicode scripts for IDNs. It charges the same amount for the registration of multilingual domain names as for ASCII domain names, although recently all registrations of IDNs made during the first year of the test bed were extended without charge for an additional six months.

⁷⁰ See <http://www.verisign-grs.com/>.

105. The VGRS IDN test bed uses ASCII Compatible Encoding (ACE) as currently proposed by the IETF IDN Working Group to encode IDNs into ASCII strings. The original ACE used was Row-based ASCII Compatible Encoding (RACE) and more recently ACE-AMC-Z (Z). IDNs have not yet been put into the .com, .net and .org zones; resolution has been provided at the third level. Nearly one million domain names have been registered. In addition, RealNames keyword technology is employed, making it possible for Microsoft Internet Explorer users to access websites with URLs containing multilingual domain names.

106. The IETF publicized the draft of the RACE algorithm in March 2000. VGRS launched the internationalized domain name test bed registration service based on RACE in November 2000. Before the launch of the registration service, some people encoded multilingual domain names into ASCII domain names beginning with "bq--" and registered them as ASCII domain names. This meant that people registered ASCII domain names that corresponded to the RACE version of IDNs before the registration service had started. That is to say, they essentially blocked the corresponding IDN registration. More recently, to accommodate the change in direction from RACE to Z, the prefix was changed to "zq--".

107. In the future, it is anticipated that a final ACE algorithm will be proposed as a standard with a new prefix. All unused four character prefixes ending in two dashes have been reserved for .com, .net and .org, thereby eliminating the problem incurred with RACE names at the beginning of the test bed.

108. From the beginning of the test bed, VGRS committed to cooperating with the standards development process: that commitment continues. When a standard is proposed, it will be implemented and the test bed would end. In the interim, to minimize multiple conversion efforts by IDN registrars, registrars continue to submit IDN registrations in RACE form and VGRS converts the names into Z form.

WALID

109. WALID⁷¹ provides a {non-ASCII-string}.{non-ASCII-string} registration service together with client software for resolving the registered multilingual domain

⁷¹ See <http://www.walid.com>

names. WALID technology is based on the IDNA/ACE recommendation endorsed by the IETF IDN working group, and supports all Unicode based languages for both registration and resolution. WALID also provides fully customizable solutions with multilingual capabilities for registries and registrars worldwide. WALID technology is part of the VeriSign multilingual test bed (see § 104-108 above).
