

iEmail - Multilingual E-mail

S.Maniam

i- DNS. Net International Pte Ltd
4 Kent Ridge Road, Singapore 119213
E-mail: maniam@i-dns.net

Abstract

The presently exploding demand for internationalized, multilingual domain names will only be exceeded by the demand for internationalized, multilingual e-mail addresses. Traditional, English-ASCII e-mail addresses are inappropriate and insufficient for the vast majority of non- English populations embracing the Internet, and a new solution that meets their needs is required.

i-DNS.net's core multilingual e-mail addressing technology - *iEmail*, is that solution. Enabling complete support for *all* Unicode-supported languages in partial or complete e-mail addresses, *iEmail* allows for true multilingual addressing while retaining total compatibility with all existing standards-compliant Mail Transfer Agents (MTAs) in use on the Internet. Capable of supporting any and all of the IETF official Internationalized Domain Name (IDN) solutions, the *iEmail* technology is implementable in a wide number of forms, with existing products ranging from client-based solutions, to client-server as well as server-only solutions. Leveraging the known reliability of UNIX operating systems as well as the superior price performance of Industry-Standard Intel-compatible architecture, *iEmail* provides superior functionality, high price-performance, all traditional e-mail services, complete standards compliance and *true* multilingual addressing

1. INTRODUCTION

Far more so than the World Wide Web, e-mail is at the fundamental core of the global communications breakthrough of the Internet; unprecedented in its reach, it is a fast, reliable and efficient direct communications channel for anyone, anywhere in the world. The demand for access to this means of communication is hard to overstate – with cyber cafés appearing even in remote corners of the world, the need to send and receive e-mail is evident throughout all countries and all people of the world. As increasing segments of the world population embraces the Internet revolution and come online, the use of, and the demand for e-mail can only increase exponentially.

One aspect of e-mail, however, in its present form fails to meet the needs of the increasingly global users of the Internet – specifically, there is no provision for true multilingual e-mail addressing. This may at first glance appear to be a non-obvious, perhaps even as a trivial problem – after all, when one considers e-mail, it would seem that it is the content of the mail that is most important. But nothing could be further from the truth.

Technology exists in order to meet real, human needs, and will only flourish if it answers these needs. Without talking in terms of market demand, such as the exploding adoption of multilingual domain-names, or political attempts to protect a language, such as the recurring actions of the French government, one need only consider the fact that an e-mail address is effectively one's very identity on the Internet, to realize the significance and importance of an e-mail address. The need for a non-English, multilingual e-mail address is non-critical only to the English-speaking. There is additionally a pragmatic, functional reason for the use of non-English e-mail addresses by non-English speakers. Communications is about understanding and meaning, and language is the tool that facilitates this.

Imagine if everything in America was unchanged as it is, with English being spoken and written everywhere – but all e-mail addresses were issued in ancient Egyptian hieroglyphs, that nobody knows how to pronounce or understand, and are only useable after painstaking memorization of letter shapes. Exchanging addresses with a potential client would be an exercise in frustration. Verbally informing someone is the fastest, most convenient method of messaging available would, however, be itself a near-impossible task. This is the situation as presently faced by the huge segments of the world population that are getting online – but have only marginal proficiency in English. The need and demand for multilingual e-mail addresses is self-evident.

iEmail technology is the solution.

2. TECHNOLOGY DESIGN

2.1 The Problem

There are an exceedingly large number of languages in use amongst population of the world. When one considers the subset that possess in a written script, the number drops somewhat, but remains extremely large – and these are not adequately supported by traditional Western character sets.

While there are similarities between many languages in their written script, or “glyphs” in Unicode terminology, differences abound and the basic 7-bit ASCII character-set is already barely adequate for the related Western European languages, and is completely inadequate for the representation of the glyphs, scripts and characters of the vast majority of languages across the world.

Existing Mail User Agents (MUAs) and Mail Transfer Agents (MTAs) do not allow the use of **8-bit characters** used in localized encodings such as ISO-8859-6 and CP1256 for Arabic, as well as the Unicode multilingual standard. Additionally, many of these languages require more than one byte to represent each character.

The restrictions on allowable characters are not the result of some patchable program flaw – the restrictions exist on *all* MTAs and MUAs because they originate from the RFC standards regarding e-mail themselves. It could even be said that to design e-mail software to be standards compliant is in fact to design e-mail software to be multilingual intolerant.

While theoretically the proposal and adoption of a new set of mail-RFCs would be one method to allow the use of multilingual e-mail addresses, in a practical sense this is not a solution at all, as it will be impossible to force the obsolescence of every last existing Mail Transfer Agent and Mail User Agent on the planet, many of which have been operating reliably from the beginning of the Internet and are sophisticated and effective solutions in their own right.

An elegant solution that retains compatibility with the existing installed-base of mail-software and servers across the Internet is required.

2.2 Solution Requirements

The *primary requirement* will necessarily be to enable the use of multilingual e-mail addresses while retaining compatibility with *all* existing MUAs and MTAs, with full RFC-standards compliance. Equally crucial design requirements exist, however:

- The solution must allow for immediate use of multilingual e-mail addresses.**

While defining a theoretical standard is a meaningful academic exercise, the ability to deploy and employ multilingual e-mail addresses *immediately*, is materially more effective in furthering the aim to truly globalize the Internet.

- The solution must be compatible with all Internationalized Domain Name (IDN) standards**

Deploying a solution before IDN systems are finalized by the IETF would appear to carry inherent risk of obsolescence and incompatibility. The way to counter this is to be modularly compatible with *all* proposed IDN solutions.

- The solution must support the widest possible range of languages**

If the system does not work with certain languages/language-encodings, then it is at best a partial solution, and at worst no solution at all. If the goal is true globalization of Internet e-mail addressing, then *all* languages must be supported. Not only must the evolving standard in multilingual encoding, Unicode, be supported, but the various localized language encoding schemes that have arisen throughout the world independently must be supported as well, as they are already widely entrenched and part of the installed base.

- The solution must not *demand* the use of non-English characters.**

If the system only works with non-English characters, it would necessitate the deployment of two different sets of mail systems on the Internet, one for English mail, and one for non-English addresses. This is an inelegant and inefficient solution. A multilingual e-mail addressing solution must be able to seamlessly handle both traditional English addresses as well as non-English, multi-byte languages not supported by existing MTAs.

It is also highly desirable to allow for mixed character-set addresses, in order to allow users the freedom to choose whichever e-mail address suits their requirements.

- The solution must be technically efficient and elegant

A solution that is inefficient is doomed to eventual replacement, and is ultimately no solution at all. Only functional, effective solutions will stand the test of time. We at i-DNS.net are working for the ultimate goal of true globalization of the Internet, and we have designed our solution to last.

- The solution must be deployable on open platforms

Compatibility and interoperability go beyond interacting with other software. A solution that requires the use of proprietary hardware would carry unstated, but very real limitations. A solution that operates on the widest-possible range of hardware implementations, from lowest-cost PC hardware to high-availability netserver, would be able to meet the cost-performance requirements of the widest possible range of needs, allowing the use of true multilingual addressing for the greatest number of people throughout the world.

- The solution must be modular, robust, and scalable

Good software design is never out of place in any solution. Providing unprecedented e-mail addressing capability without also offering stability, reliability and maintainability is a proposition that we at i-DNS.net are not willing to consider.

iEmail technology is designed to meet *all* of these requirements.

Our technology operates fully on any UNIX-based operating system and is compatible with a wide range of hardware implementations, such as Sun SPARC/Solaris platforms, Compaq Alpha systems running Tru64 or Linux, to cost-efficient PC-based hardware operating with a wide range of Linux distributions, BSD implementations, and more.

Capable of supporting any mix of multilingual and English e-mail addresses, i.e. anything from fully-multilingual (both a multilingual UserID as well as a multilingual domain) such as <மணியம்ஐ-டிஎன்எஸ்.வணி> to mixed-multilingual-English addresses such as மணியம்@i-dns.net reiterate, any language, any combination. *iEmail* technology supports *all* existing (and future) Unicode languages¹ as well as the various localized encodings in use worldwide (e.g. BIG5 in China, S-JIS in Japan, and more).

Our software solution is scalable, high-performance, and modular to allow for easy extensibility of capabilities and future enhancements. Fully capable of supporting any IDN solution eventually ratified by the IETF as a standard, *iEmail* technology is ready for immediate deployment with most current commercial IDN implementations.

¹ Presently Unicode v3.0.1 defines around 60 languages

3. IMPLEMENTATION

3.1 Technology Overview

Through the use of ASCII-compatible encoding, the *iEmail* system detects and translates multilingual mail addresses and converts it to a fully RFC-compliant mail address that will survive travel between all existing MTAs on the Internet.

Multilingual e-mail addressing requires dealing with display/rendering, representation and transport issues.

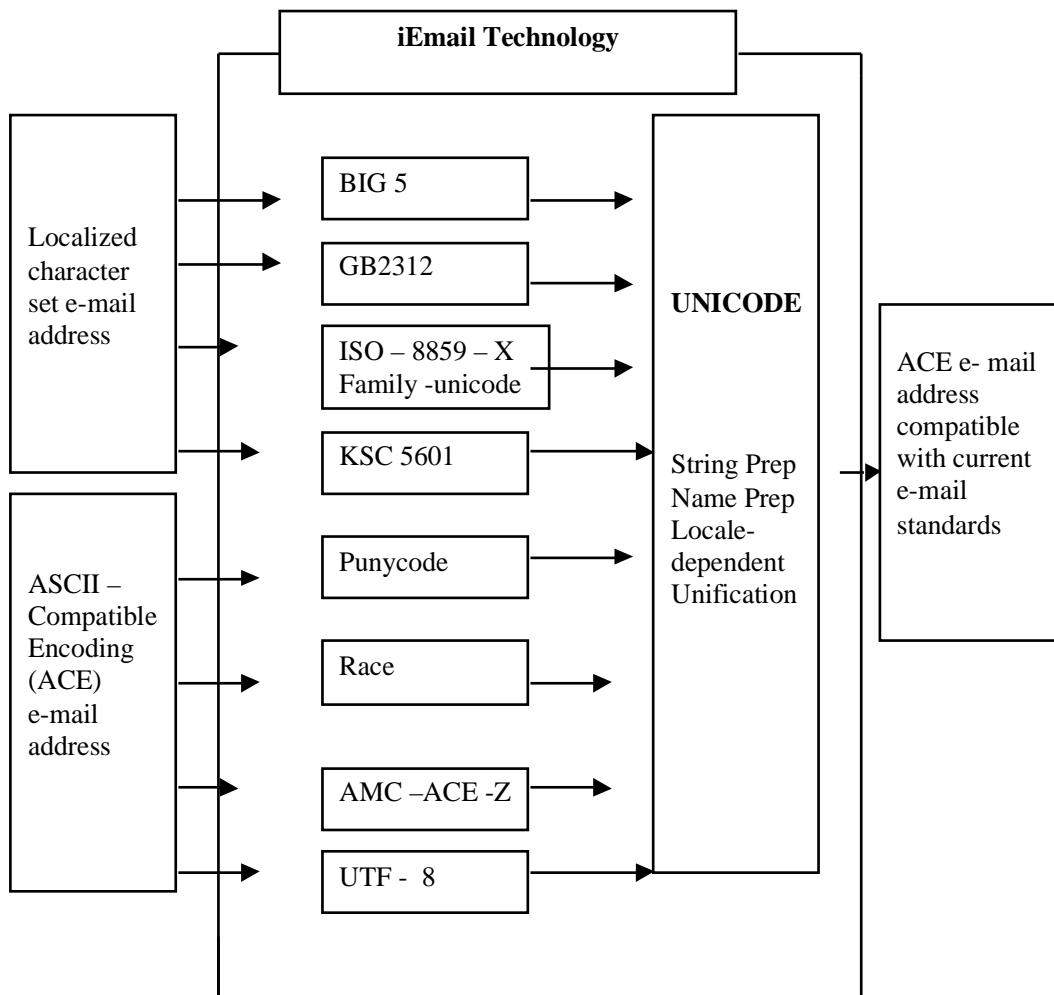
Bitmap renderings of multilingual script are of course not the way they are internally represented. 8-bit character-encoding schemes (CES) are employed to represent the individual characters used in non-English script. While a “universal” encoding format exists (UTF-8), there are a wide range of localized encoding formats. For example in Tamil there are 3 formats - TAB , TAM & TSCII. While the Tamil Nadu Government has approved TAB and TAM as the standards, the global INFITT community has since the Tamil Internet 2000 conference added TSCII primarily in the interests of overseas Tamil users. E-mail addresses must be useable and displayable in any of these formats. While none of these 3 formats should be ignored, we are making progress towards a universal solution based on UNICODE. Given that UNICODE is fast becoming reality, we are offering iEmail Solutions in UNICODE encoding

As stated before, existing MTAs are hostile to 8-bit character encodings. An ASCII-compatible encoding (ACE) circumvents this problem. By using an ACE conversion method that encodes multilingual, 8-bit CES characters into ASCII characters, one can obtain e-mail addresses that conform fully to all relevant mail-related RFCs.

iEmail technology is able to accept a wide range of different language formats and automatically perform the necessary functions in order to generate a completely standards-compliant e-mail address.

The preceding schematic gives a graphical representation of the *iEmail* conversion process, and the wide range of encoding formats it is able to accept. This includes major localized character sets such as **BIG5** (Taiwan/Hong Kong), **GB2312** (PR China), **SJIS/EUC-JP** (Japan) and **KSC5601** (Korea). The IETF-sanctioned ACE **PunyCode** (draft-ietf-idn-punycode) is also supported and is used as the default output ACE to ensure compatibility to the worldwide Internationalized Domain Name standard.

A number of locale, encoding-specific and language-specific issues abound across the various world languages (e.g. differing representations for similar glyphs based on contextual positions in Arabic languages) and *iEmail* technology is designed to automatically and seamlessly resolve these issues without requiring manual intervention.



***iEmail* Technology**

3.2 Technology options

iEmail technology is deployable in many different forms.

iClient™ - Client-based Implementation

iClient™ is an *iEmail* technology-enabled client software application that resides in a client machine, operating invisibly in the background. Whenever any MUA is invoked and a multilingual e-mail address is detected, *iClient*™ intercepts the mail process and performs the modifications as required in order to enable multilingual-addressing capability without breaking RFC821/822 compatibility. A central function of iClient is to allow email address forwarding whereby mail directed to a multilingual email address is forwarded to an existing email box with an English email address, thereby accelerating iEmail technology adoption by today's English-centric email service vendors.

iEmail.Web - Web-mail (CGI) Implementation

iEmail.Web is an *iEmail* technology-enabled web-mail server that provides multilingual messaging capability to any client with a modern web browser. This implementation avoids both the need for distribution of client software as well as the need to provide client-software

configuration support. The web-mail server will provide all mail functions, both for the user interface (allowing users to read, compose and manage mails) and for the back-end (handling SMTP processes and operating as a fully RFC-compliant MTA). This can be considered an all-in-one solution.

Other Implementations

iEmail technology can be employed anywhere a requirement for multilingual e-mail addressing exists – directly included into end-user MUA client software, modification and enhancement of existing MTAs, mail-retrieval (POP3/IMAP4 or other) software, calendaring/scheduling, groupware, PKI servers and more. Any messaging system that involves multilingual usage will be enhanced by the inclusion of *iEmail* technology.

4. CONCLUSION

It is hoped that this white paper provides a clear overview of the operation and possibilities of the iEmail technology.

Our technology is effective, flexible, complete, immediately deployable, and will help take the Internet one step further towards the inevitable *true* globalization that is at the core of what the Net promises.

5. ACKNOWLEDGMENTS

The author thanks current and former staff of i-DNS.net, who have made many material contributions to this manuscript. I also thank i-DNS.net for permitting me to use proprietary information and material that have been included in this manuscript.