सत्यमेव जयते

Ministry of Electronics &
Information Technology
Government of India

# Universal Acceptance & Multilingual Internet

Phase-I

April-2022

Universal Acceptance

# Phase I

# Table of Contents


1. *Executive Summary* ..... *iii*
2. *Preface* ..... *v*
3. *Purpose / Scope* ..... *vii*
4. *Definitions of Terms* ..... *ix*

**5. MULTILINGUAL INTERNET FROM A COMMON MAN PERSPECTIVE ..... 1**

**6. UNIVERSAL ACCEPTANCE DEFINITION ..... 3**

**7. CURRENT STATE OF UNIVERSAL ACCEPTANCE AND MULTILINGUAL INTERNET ..... 4**

**8. READINESS STATUS AND CHALLENGES OF VARIOUS SERVICES ..... 14**

a. Hardware ..... 15
b. Software / OS ..... 15
c. Browser ..... 16
d. E-Mail ..... 17
e. Websites ..... 18
f. Security Systems ..... 19
g. Apps / Social Media ..... 20

**9. TEST BED, COMPLIANCE AND MOTIVATIONAL MEASURES ..... 21**

**10. CAPACITY BUILDING - UA Champions, UA Ambassadors & UA Icons ..... 22**

**11. ELEMENTS OF THE ROADMAP ..... 23**

**12. AWARENESS CREATION AND TRAINING PROGRAMMES ..... 25**

**13. RECOMMENDATIONS ..... 27**

*a. Proposed Funding Model ..... 33*

*b. Tentative budget requirements (for first year of operation) ..... 33*

**14. Annexure I - List of scheduled Indian languages and major scripts used ..... 35**

**15. Annexure II - Readiness Status and Challenges of various services in implementation of UA Support ..... 37**

*a. Hardware ..... 37*

*b. Email ..... 38*

*c. Security Systems ..... 39*

*d. Software/Operating System ..... 44*

*e. Browsers ..... 48*

*f. Websites ..... 51*

*g. Apps/Social Media Apps ..... 52*

*h. Industry bodies in India ..... 53*

**16. ACKNOWLEDGEMENTS ..... 57**

# EXECUTIVE SUMMARY

The Internet landscape has changed dramatically over the last decade with the expansion and evolution of available Top-Level Domains (TLDs), generic Top-Level Domains (gTLDs), the Internationalized Domain Names and Email Address Internationalization (EAI). A new horizon has opened up with the possibility to have Internationalized Domain Names in one's mother-tongues and scripts.

As of Jan 2022, there are Globally there are 1488 active TLDs including 153 IDN TLDs (mostly ccTLDs), which includes India's 15 IDN ccTLDs, covering 22 Indian languages represented using 11 scripts (10 Unicode Blocks). NIXI has already started offering Indian language domain names in all 22 scheduled languages / scripts.

The next billion Internet users are not online because systems that enable their access do not support their language. India being a multilingual country and 92 percent of population is non-English, for want of proper support / availability of tools / technologies, 50 percent of India's population is yet not online. Providing access to the internet for these users will require technological solutions apart from merely Internationalized or multilingual content. Localized Domain names and email addresses need to be part of these technological solutions.

Universal Acceptance (UA) is the state in which all valid Domain Names including new generic TLDs, Internationalized TLDs and Internationalised email IDs are treated consistently, regardless of script, number of characters, or how new it is and are accepted equally by all Internet-enabled applications, devices, and systems.

The India UA programme is spread over short term (3-9 months), medium-term (6-18 months) and long term (up to 36 months) with multi-stake holder participation. The short term includes UA Program background work and Initiation, MOU with UASG, knowledge dissemination portal, establishment of Monitoring committee, support System (NIXI as Nodal org), initiation of Centre for Excellence and minimum 3 language domain names for Government and private websites.

The medium term includes acceptance of IDNs and EAI in various forms of the website, and EAI level 1 support, wherein one can send and receive via internationalised emails. The long-term activity focuses on challenges and possible solutions to DNS security, Indian languages SPAM filter development and allied R&D. The long term also focuses on recommending major email service providers for providing facility to create language email boxes.

Industry, State Government involvement is expected not only from technical implementation, but also financial support via various mechanism. Tentative Budget of Rs. 461.50 lac, Rs. 1211.00 lac and Rs. 2004 lacs for short term, medium term and long term to be provisioned.

# PREFACE

Since the inception of the Internet, Domain Names were available only in Latin characters and having Domain Names in one's language was a distant dream. But, today with the initiative of ICANN, having Domain Names in any script / language of the world has become reality.

The Internet landscape has changed dramatically over the last decade with the expansion and evolution of available Top-Level Domains (TLDs), generic Top-Level Domains (gTLDs), the Internationalized Domain Names and Email Address Internationalization (EAI)

Since 2010, the industry has seen the introduction of IDNs which are based on different languages and scripts. Over 1,200 greater varieties of new generic Top-Level Domains (new gTLDs) got registered. Email Address Internationalization (EAI) also started appearing on the scene. The gTLDs consisted of New short Top-Level Domain Names as well as Long Top-Level Domain Names which removed the restriction of 3 characters at TLD level

Though the Internet and Domain Name System (DNS) have transformed, many websites and applications have not kept themselves up with the changes. Many systems still cannot process all Domain Names or email addresses and more specifically the Internationalized Domain and Email Address Internationalization and have not realized that the growth of Internet users is dependent on this.

The Universal Acceptance (UA) initiative of Universal Acceptance Steering Group (UASG) of ICANN addresses this issue and the solutions are already available from the industry and for different technological platforms. However, everyone in the chain needs to be UA ready to achieve a truly Multilingual and inclusive Internet.

To support the new Top-Level Domains and Email Addresses, Applications and Systems must be capable of fundamental five actions: Accept, Validate, Store, Process and Display. Software and online services support Universal Acceptance when they offer the five actions listed above for all Domains and email Names.

A new horizon has opened up with the possibility to have Internationalized Domain Names in one's mother-tongues and scripts. ICANN has opened up having gTLD's other than those

previously eight viz., .com, .org, .net, .int, .edu, .gov, and .mil, .arpa which were created in the 1980s. Today one can have a gTLD with the name of an organization or even the name of a city such as .delhi or an institution such as .iitmumbai. These changes have opened up new vistas and exciting possibilities and at the same time technological challenges as well as legal and security issues.

In order to promote UA, Universal Acceptance Steering Group (UASG)–A community-led initiative was founded in February 2015 by ICANN

- Tasked with undertaking activities to promote the Universal Acceptance of all valid Domain Names and email addresses.
- Members from more than 120 companies (incl. Apple, GoDaddy, Google, Microsoft, and Verisign), governments, and community groups.

### The need:

Though 65 percent of the world's population is connected to the Internet, 92 percent of the web pages are published only in 12 languages. Also, 60 percent of Internet publications are in the English language alone[1]. It is interesting to note that there are 7,000 languages and dialects used across the globe and the next billion Internet users will likely come from non-English speaking countries. Hence there is a need for technological shift to bring this next billion plus users online.

Many of the next billion Internet users are not online because systems that enable their access do not support their language. India being a multilingual country and 92 percent of population is non-English, for want of proper support / availability of tools / technologies, 50 percent of India's population is yet not online. Providing access to the internet for these users will require technological solutions apart from merely Internationalized or multilingual content. Localized Domain names and email addresses need to be part of these technological solutions.

### Benefits:

The Top-Level and Internationalized Domains have evolved and matured enough as far as the technology is concerned. For increasing business reach and greater opportunities, the UA for applications, services are crucial. People are generally comfortable in trusting and communicating in their local language. Having a local language identity (i.e., email address) is easier to use for the non-English speaking user for participating in any government, social, banking and other online applications. UA allows customers to expand their customer base by offering products / technologies / services to various countries in their own languages. Businesses can now communicate, share information, provide products, technologies and services in the customer's language, creating trust and build a huge business potential while bringing the next billion plus users online. Govt. services can also communicate with the user in their local language creating inclusiveness and better adoption.

1 https://w3techs.com/technologies/history_overview/content_language/ms/y

# PURPOSE / SCOPE

The purpose of this document is to prepare a roadmap of fundamental issues of multilingual internet especially focusing on Universal Acceptance. The document will define the current practices and recommendations comprising of testing frameworks, policy recommendations etc. to identify and fill the gaps in the current Internet system that hinder the growth and acceptance of IDNs. Phase wise reports & implementations are proposed towards realizing the goal of "Multilingual Internet and Universal Acceptance".

The current report (Phase-I) focuses on Universal Acceptance. However, the subsequent report (Phase-II), which is under preparation, will dwell upon the challenges / issues in relation to having truly "Multilingual Internet" with Multi-Stakeholder participation. This first phase of "Universal Acceptance" is proposed to be realized under short term, medium term and long-term activities.

# DEFINITIONS OF TERMS

| | |
|---|---|
| **IDN** | Internationalized Domain Names |
| **ASCII** | American Standard Code for Information Interchange |
| **BIS** | Bureau of Indian Standards |
| **ccTLD** | Country Code Top Level Domain |
| **gTLD** | Generic Top-Level Domain |
| **C-DAC** | Centre for Development of Advanced Computing |
| **CLDR** | Common Locale Data Repository |
| **DNS** | Domain Name System |
| **EAI** | Email Address Internationalization |
| **GIGW** | Guidelines for Indian Government Websites |
| **GOI** | Government of India |
| **gTLD** | generic Top-Level Domain |
| **HTML** | Hypertext Markup Language |
| **ICANN** | Internet Corporation for Assigned Names and Numbers |
| **IDNA** | Internationalized Domain Names in Applications |
| **IDNSBL** | Internationalised Domain Name System blacklist |
| **IEA** | Internationalized Email Addresses |
| **IETF** | Internet Engineering Task Force |
| **IMA** | Internet Message Access |
| **IMAP** | Internet Message Access Protocol |

| | |
|---|---|
| **ISCII** | Indian Script Code for Information Interchange |
| **EAI Level 1 (L1 Support)** | Sends and Receives from EAI Addresses |
| **EAI Level 2 (L2 Support)** | L1 level plus facility to create EAI addresses |
| **MDA** | Mail Delivery Agent |
| **MSA** | Mail Submission Agent |
| **MSP** | Mail Service Provider |
| **MTA** | Mail Transfer Agent |
| **MUA** | Mail User Agent |
| **NIXI** | National Internet Exchange of India |
| **NLP** | Natural Language Processing |
| **OS** | Operating System |
| **PSU** | Public Sector Unit |
| **RFC** | Request for Comments |
| **SAP** | Systems Applications and Products |
| **TLD** | Top Level Domain |
| **UA** | Universal Acceptance |
| **UASG** | Universal Acceptance Steering Group |
| **UT** | Union Territories |
| **UTF-8** | UCS Transformation Format 8 |
| **W3C** | World Wide Web Consortium |
| **WG** | Working Group |
| **WHATWG** | Web Hypertext Application Technology Working Group |

# MULTILINGUAL INTERNET

## MULTILINGUAL INTERNET, FROM A COMMON MAN PERSPECTIVE

**A set of tools / services by which one can easily create, communicate, transact, process and retrieve information with ease in digital medium without language barrier.**

The Internet has become all pervasive and has become a part and parcel of our lives. We all have witnessed the power of the internet more specifically in the pandemic, which helped us to stay connected as well as do business as usual. As of Jan 2021, the global active internet user stands at 4.66 billion while 90-95% consumption of internet is from social networking usage alone

The convergence of AI (Artificial Intelligence) and IoT has redefined the way industries, business, and economy's function. Speech to speech technologies, facial recognition, virtual assistants, Machine translation systems, natural language processing, natural language generation and many more are now growing part of our lives and help dissolving language barriers.

50 billion devices expected to get connected in 2022, within 50 years we will have the technology for embedding internet transceivers into human brains, and that by 2069 the brain-machine interface will be fully developed, wherein the internet ecosystem will be catalytic to human advancement. The residential internet speeds will be touching 10 gigabits per second – 10 times faster than today's networks

Multilingualism becomes an extremely crucial aspect to bring the next one billion users on the network. There are 7,000 languages and dialects used across the globe. In India we have 22 scheduled languages, and we have one to many and many to many relationships between the scripts and languages. As an example, Devanagari script alone covers 10 scheduled languages, such as Boro (Bodo), Dogri, Konkani, Hindi, Maithili, Marathi, Nepali, Santali, Sanskrit, Sindhi while Sindhi is written in Devanagari as well as Perso-Arabic script.

The next billion Internet users will likely come from non-English speaking countries, providing access for these users will require more than supporting internationalized or multilingual content. Localized domain names and email addresses are required.

Consumption as well as creation of multilingual contents is also on rise, which is a boon to advancements in human inspiring systems. The advancement in Machine Learning has led to remarkable progress in Natural Language Processing (NLP), the field of Artificial Intelligence that gives computers the ability to understand human language.

# UNIVERSAL ACCEPTANCE DEFINITION

Universal Acceptance (UA) is the state in which all valid Domain Names and email addresses are accepted, validated, stored, processed and displayed correctly and consistently, regardless of script, number of characters, or recently introduced in the Unicode and are accepted equally by all Internet-enabled applications, devices, and systems.

To achieve Universal Acceptance, internet applications and systems must treat all Top-Level Domain (TLDs) in a consistent manner, including new generic TLDs and all Internationalized TLDs. This includes supporting all country code Top-Level Domains (ccTLDs), new and long generic Top-Level Domains (gTLDs), and Internationalized Domain Names (IDNs).

All Domain names should be validated against the Internationalized Domain names in applications IDNA2008[1] standard.

1 https://datatracker.ietf.org/doc/html/rfc5895

# CURRENT STATE OF UNIVERSAL ACCEPTANCE AND MULTILINGUAL INTERNET

## National Scenario

As of Jan 2022, there are 1488 active TLDs including 153 IDN TLDs (mostly ccTLDs), which includes India's 15 IDN ccTLDs, covering 22 Indian languages represented using 11 scripts (10 Unicode Code Charts).

NIXI has already started offering Indian language domain names in all 22 scheduled languages.[2] .भारत (.bharat) IDN ccTLD (using Devanagari script) covers 8 languages Bodo (Boro), Dogri, Hindi, Konkani, Maithili, Marathi, Nepali, and Sindhi-Devanagari, while .ভারত IDN ccTLD covers 2 languages Bengali and Manipuri, also includes ccTLDs from RTL scripts viz. Urdu, Sindhi and Kashmiri. These domain names are being offered by accredited registrars[3] and the user/registrant can register the Indian language domain names in his/her choice of language. Email in one's own language offerings are also on rise. The following table shows the IDN ccTLDs in the scripts mentioned and the languages supported. Annexure I include "List of scheduled Indian languages and major scripts used".

Internationalized Domain Names status in India

| Internationalized Domain Name (IDN) | Punycode | Script | Script Code | Language(s) supported |
|---|---|---|---|---|
| .भारत | xn—h2brj9c | .Bharat in Devanagari Script | Brx-deva<br>Dgo-deva<br>Hin-deva<br>Mai-deva<br>Mar-deva<br>Nep-deva<br>snd-deva | Bodo(Boro), Dogri, Hindi, Konkani, Maithili, Marathi, Nepali, and Sindhi-Devanagari |

2 https://registry.in/home

3 https://www.registry.in/accredited-registrars

| Internationalized Domain Name (IDN) | Punycode | Script | Script Code | Language(s) supported |
|---|---|---|---|---|
| .ভারত | xn—45brj9c | .Bharat in Bengali Script | Ben-beng<br>mni-beng | Bengali and Manipuri |
| .భారత్ | xn—fpcrj9c3d | .Bharat in Telugu Script | tel-telu | Telugu |
| .ભારત | xn—gecrj9c | .Bharat in Gujarati Script | guj-gujr | Gujarati |
| . ترابھ | xn—mgbbh1a71e | .Bharat in Arabic Script | urd-arab | Urdu |
| .இந்தியா | xn—xkc2dl3a5ee0h | .Bharat in Tamil Script | tam-taml | Tamil |
| .ਭਾਰਤ | xn—s9brj9c | .Bharat in Gurmukhi (Punjabi) | pan-guru | Punjabi |
| .ಭಾರತ | xn—2scrj9c | .Bharat in Kannada Script | kan-Knda | Kannada |
| .ଭାରତ | xn—3hcrj9c | .Bharat in Oriya Script | ory-Orya | Oriya |
| .ভাৰত | xn—45br5cyl | .Bharat in Bengali (Unicode) Script | asm-Beng | Assamese |
| .भारतम् | xn—h2breg3eve | .Bharat in Devanagari Script | san-Deva | Sanskrit |
| .भारोत | xn—h2brj9c8c | .Bharat in Devanagari Script | sat-Deva | Santali |
| .تراب | xn—mgbbh1a | .Bharat in Arabic Script | kas-Arab | Kashmiri |
| .ترابٻ | xn—mgbgu82a | .Bharat in Arabic Script | snd-Arab | Sindhi |
| .ഭാരതം | xn—rvc1e0am3e | .Bharat in Tamil Script | mal-Mlym | Malayalam |

## Global scenario

The Universal Acceptance Steering Group (UASG) was formed in 2015 to advocate and enthuse the relevant stakeholders to make their applications UA ready.

UASG has formed following Global Working Groups to take care of various activities of UA readiness

a. **UA Technology WG**–focuses on remediation of standards and technology and developing technical training. They focus on Technology Enablers and Technology Developers as the stakeholders.

b. **UA Email Address Internationalization (EAI) WG**–focuses on identification of relevant technology and its gap in supporting EAI, remediation of the technology and providing training materials for email software and service providers to promote EAI support and deployment.

c. **UA Measurements WG**–plans, oversees and directs the gap analysis efforts of the UASG for various frameworks and technologies and reports on progress on UA readiness.

d. **UA Communications WG–**develop communication strategy for the UASG and oversee its execution, in collaboration with other WGs.

As per the mandate UASG will continue to:

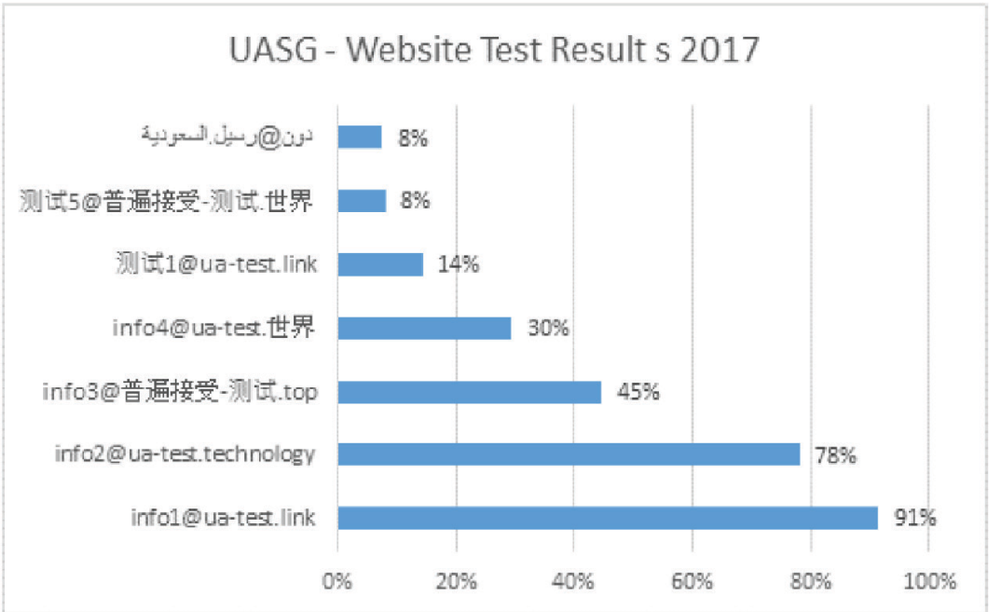
1. Raise awareness of the issue among the relevant stakeholders
2. Help technology and email providers make their tools, systems and services
3. UA ready by providing documentation and training
4. Encourage businesses and governments to ask their developers and
5. suppliers to provide UA ready solutions for the benefit of end users
6. Measure progress on UA readiness

UASG continues to measure and publish UA readiness reports from time to time. Reports covering the following can be found on uasg.tech site.

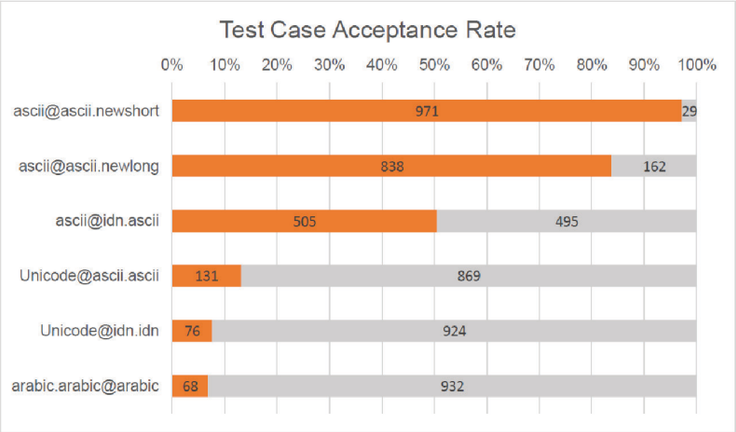
1. UA readiness of technology, including social media, Content Management
2. Systems, Programming Languages, etc.
3. Top websites globally accepting a wide range of email addresses.
4. Email software and services supporting internationalized email addresses.
5. Email deployments supporting internationalized email addresses.

### UA Readiness for Top Global websites

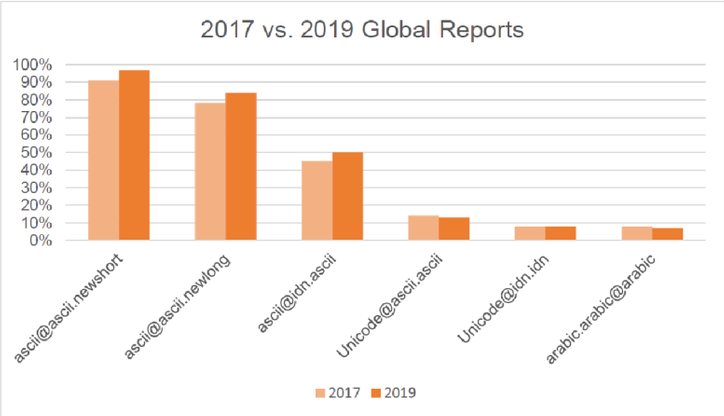
UASG conducted Global Evaluation of Websites for Acceptance of E-mail Addresses in the year 2017, 2019, 2020. Around 1000 Top Global websites were evaluated for EAI acceptance.



**August 2019:**

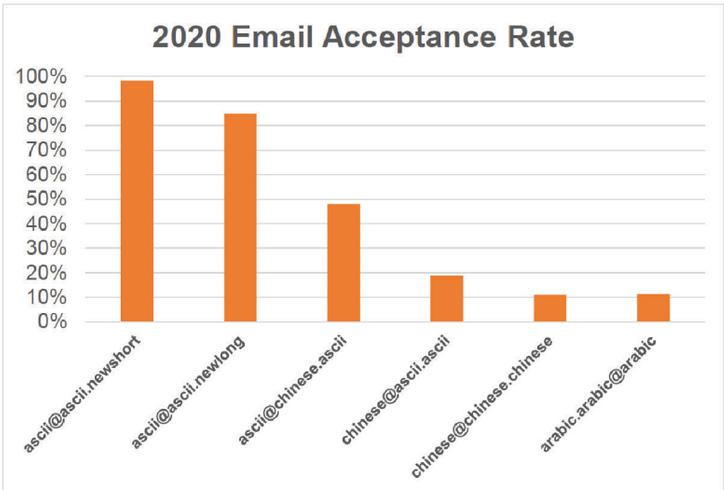


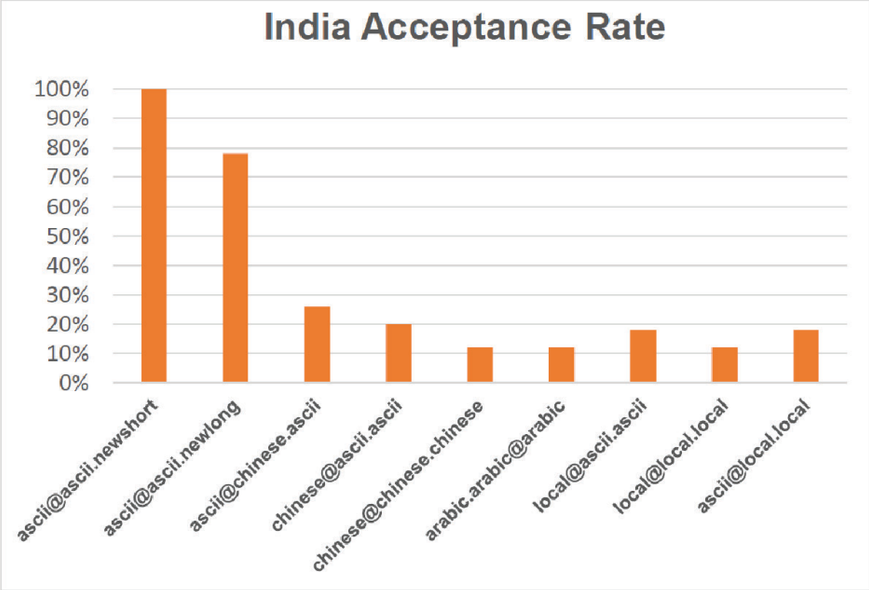

**2017 vs. 2019 Test Totals Comparison**




Country-Based Evaluation of Websites for Acceptance of Email Addresses in 2020, 21 April 2020, wherein total of 1,117 websites were tested over the 20 different countries

**April 2020:**






### UA compliance of Programming Languages and Frameworks

UASG conducted study related to compliance of programming languages and frameworks, the report of the same is published in the documents UASG018, UASG018A (sept 2020)

- GNU Libidn. Implementation of IDNA2003 in C. Bindings available for Perl and Ruby.
- GNU Libidn2. Implementation of IDNA2008 in C by the author of GNU Libidn.
- International Components for Unicode. Versions are available for Java and for C with C and C++ bindings.
- Python encodings.idna. Part of the Python standard library. Test in Python and Python3.
- Python idna module. A replacement for the Python standard library encodings. idna module that supports IDNA2008. Test in Python and Python3.
- PHP IDN functions. Part of the PHP standard library, supporting IDNA2003 and IDNA2008.
- Go idna package. Part of the Go standard library supporting IDNA2008.
- Javascript idna-uts46 npm module. Supports IDNA2003 and IDNA2008. Bundled with Node.js.

### Status of the evaluation:

| |
|---|
| UA ready |
| UA ready but developer needs to be careful |
| UA not ready |

| Language | Lib Name | Compliance on dataset (%) | Datasets |
|---|---|---|---|
| c | Libcurl | 84.3 | HEs |
| C | Libidn2 | 95.2 | LA2U, LU2A |
| Csharp | Mailkit | 84.3 | HEs |
| Csharp | Microsoft | 83.9 | LA2U, LU2A |
| Go<br>Go | Idna<br>mail | 79<br>100 | LA2U, LU2A<br>HEs |
| Go<br>Java<br>Java | Smtp<br>Commons- validator<br>Guava | 19.6<br>85.5<br>77.8 | HEs<br>HEs, HDns<br>HDns |
| Java<br>Java | Icu<br>Jakartamail | 93.5<br>82.4 | LA2U, LU2A<br>HEs |
| Java | Jre | 71 | LA2U, LU2A |
| Js<br>Js | Idna-uts46<br>Nodemailer | 85.5<br>84.3 | LA2U, LU2A<br>HEs |
| Js | Validator | 94.2 | HEs, HDns |
| Python3 | Django_auth | 48.1 | HEs, Hld |
| Python3 | Email_validator | 86.3 | HEs |
| Python3 | Encodings_idna | 67.7 | LU2A, LA2U |
| Python3<br>Python3<br>Rust | Idna<br>Smtplib<br>Idna | 100<br>84.3<br>87.1 | LA2U, LU2A<br>HEs<br>LA2U, LU2A |
| Rust | Letter | 7.8 | HEs |

### UA readiness in Major Email Software and Services:

Studies were undertaken in Sept 2018, August 2019 and December 2020 (Documents UASG021A, UASG021B, UASG028)

The study was focused on components from the email software ecosystem to determine the degree to which each component is "EAI Ready", that is, whether it currently handles EAI messages, addresses, and domain names correctly

### Email Software and Service Selections:

- Component 1: evaluated in its capacity as a Mail User Agent ("MUA").
- Component 2: evaluated in its capacity as a Mail Submission Agent ("MSA"), Mail Transfer Agent ("MTA"), and Mail Delivery Agent ("MDA").
- Component 3: evaluated in its capacity as a Mail Service Provider ("MSP").

| Software | MUA | MSA | MTA | MDA | MSP | Webmail | Region |
|---|---|---|---|---|---|---|---|
| XgenPlus Email Server | X | X | X | X | X | X | AP |
| Axigen Mail Server | X | X | X | X | | X | EUR |
| MDaemon Email Server | X | X | X | X | | X | NA |
| Oracle Beehive | X | X | X | X | | X | NA |
| Zimbra | X | X | X | X | | X | NA |
| Apple iCloud | X | | | | X | X | NA |
| Coremail | X | | | | X | X | AP |
| FastMail | X | | | | X | X | AP |
| Gmail | X | | | | X | X | NA |
| Mail.ru | X | | | | X | X | EUR |
| Microsoft Outlook.com | X | | | | X | X | NA |
| NetEase 163.com | X | | | | X | X | NA |
| Oath Mail | X | | | | X | X | NA |
| Rediffmail | X | | | | X | X | AP |
| Sina | X | | | | X | X | AP |
| Sohu | X | | | | X | X | AP |
| Tencent QQ | X | | | | X | X | AP |
| Yandex Mail | X | | | | X | X | EUR |
| IBM Notes | X | | | | | X | NA |
| Roundcube | X | | | | | X | |
| Apple Mail | X | | | | | | NA |
| Microsoft Outlook | X | | | | | | NA |
| Microsoft Windows Mail | X | | | | | | NA |
| Mozilla Thunderbird | X | | | | | | NA |
| Courier | | X | X | X | | | |
| IBM Domino | | X | X | X | | | NA |
| James Enterprise Mail Server | | X | X | X | | | NA |
| Microsoft Exchange Server | | X | X | X | | | NA |
| Oracle Communications Messaging Server | | X | X | X | | | NA |
| EXIM | | X | X | | | | |
| Halon | | X | X | | | | EUR |

| Software | MUA | MSA | MTA | MDA | MSP | Webmail | Region |
|---|---|---|---|---|---|---|---|
| OpenSMTPD | | X | X | | | | NA |
| Postfix | | X | X | | | | |
| Sendmail | | X | X | | | | NA |
| Dovecot | | | | X | | | |
| Fetchmail | | | | X | | | |
| Procmail | | | | X | | | |

The below chart lists the legend used to visualize the test results and the types of EAI support (Levels 1 and 2). The blank cells in the results tables indicate a component that does not exist

| | | | | |
|---|---|---|---|---|
| EAI level 1 (L1) – sends to and receives from EAI addresses | All or Most | Part* | Few** | Not tested |
| EAI level 2 (L2) – L1 plus provides local EAI addresses | All or Most | Part* | None | Not tested |

***Notes:*** ** Part: Some tests passed; component has partial EAI support.*

*** Few: Few tests passed; component does not have usable EAI support*

The EAI support test results summary

| Name | MUA | MSA | MTA | MDA | MSP | Webmail |
|---|---|---|---|---|---|---|
| Coremail | Few | All L2 | Most L2 | Few | All L2 | Most L2 |
| MS Outlook.com | Most L1 | Most L1 | Most L1 | None | None | Most L1 |
| Yandex Mail | Few | None | None | Few | Part | Few |
| Roundcube | Most L2 | | | | | |
| Apple Mail | Few | | | | | |
| Apple iOS Mail 14.X | Most L2 | | | | | |
| Mozilla Thunderbird | Few | | | | | |
| MS Outlook | Most L1 | | | | | |
| MS Exchange Server (hosted) | | All L1 | All L1 | Few | | |
| Exim | | Most L2 | All L2 | | | |
| Postfix | | All L2 | All L2 | | | |
| Courier | | All L2 | All L2 | All L2 | | |
| Gmail | All L1 | All L1 | All L1 | Few | | |
| Xgenplus | | Most L2 | Most L2 | Most | All L2 | Most L2 |
| Sendmail 8.17 Alpha | | Most L2 | Most L2 | | | |
| Halon | | Most L2 | Most L2 | | | |
| Thunderbird 89 beta | Most L1 | | | | | |
| Dovecot | | | | None | | |

***Source*** *:https://uasg.tech/download/uasg-030a-eai-software-test-results-en/*

## Universal Acceptance of popular browsers: (UASG016, 26 Sept 2017)

Desktop platform results

**Result key:**
**Y - Passes all tests**
**A - Fails to load the correct page**
**B - Fails to display URL correctly in location bar**
**C - Fails to display URL correctly in title bar**

| Test ID | Test Data | Chrome | Firefox | Opera | Safari | Edge | IE | Vivaldi |
|---|---|---|---|---|---|---|---|---|
| 1 | ua-test.link | Y | Y | Y | Y | Y | Y | Y |
| 2 | ua-test.technology | Y | Y | Y | Y | Y | Y | Y |
| 3 | 普遍接受-测试.top | Y | Y | Y | Y | Y | Y | B |
| 4 | ua-test.世界 | Y | Y | Y | Y | Y | Y | B |
| 5 | 普遍接受-测试.世界 | Y | Y | Y | Y | Y | Y | B |
| 6 | 普遍接受-测试。世界 | Y | A | Y | A | Y | Y | A |
| 7 | ua test.xn rhqv96g | Y | Y | Y | Y | Y | Y | B |
| 8 | xn----f38am99bqvcd5liy1cxsg.top | Y | Y | Y | Y | Y | Y | B |
| 9 | xn f38am99bqvcd5liy1cxsg.xn rhqv96g | Y | Y | Y | Y | Y | Y | B |
| 10 | اختبار-القبولالعالمي.top | C | Y | C | C | C | Y | B |
| 11 | اختبار-القبولالعالمي.شبكة | Y | Y | Y | Y | Y | Y | B |
| 12 | ua-test.link/我的页面 | Y | Y | Y | Y | Y | Y | Y |
| 13 | ua-test.technology/我的页面 | Y | Y | Y | Y | Y | Y | Y |
| 14 | 普遍接受-测试.top/我的页面 | Y | Y | Y | Y | Y | Y | B |
| 15 | ua-test.世界/我的页面 | Y | Y | Y | Y | Y | Y | B |
| 16 | 普遍接受-测试.世界/我的页面 | Y | Y | Y | Y | Y | Y | B |
| 17 | 普遍接受-测试。世界/我的页面 | Y | A | Y | A | Y | Y | B |

Mobile platform results (Testing was done on iOS 10.3 and Android 7.0)

Result key:
Y - Passes all tests
A - Fails to load the correct page
B - Fails to display URL correctly in location bar
C - Fails to display URL correctly in title bar
D,E - Other non-failing issues seen

| | | Chrome | | Firefox | | Opera | | Safari | Samsung Browser |
|---|---|---|---|---|---|---|---|---|---|
| Test ID | Test Data | Android | iOS | Android | iOS | Android | iOS | iOS | Android |
| 1 | ua-test.link | Y | Y | Y | Y | Y | Y | Y | Y |
| 2 | ua-test.technology | Y | Y | Y | Y | Y | Y | Y | Y |
| 3 | 普遍接受-测试.top | Y | Y | Y | B | B | B | Y | Y |
| 4 | ua-test.世界 | Y | Y | Y | B | B | B | Y | Y |
| 5 | 普遍接受-测试.世界 | Y | Y | Y | B | B | B | Y | Y |
| 6 | 普遍接受-测试。世界 | Y | Y | A | A | B | A | A | Y |
| 7 | ua-test.xn--rhqv96g | Y | Y | Y | B | B | B | Y | Y |
| 8 | xn---- f38am99bqvcd5liy1cxsg.top | Y | Y | Y | B | B | B | Y | Y |
| 9 | xn---- f38am99bqvcd5liy1cxsg.xn--rhqv96g | Y | Y | Y | B | B | B | Y | Y |
| 10 | اختبار-القبولالعالمي.top | C | C | C | B + C | B + C | B + C | C | B |
| 11 | اختبار-القبولالعالمي.شبكة | Y | Y | Y | B + C | B + C | B + C | Y | Y |
| 12 | ua-test.link/我的页面 | Y | Y | B | B | Y | B + E | Y | E |
| 13 | ua-test.technology/我的页面 | Y | Y | B | B | Y | B + L | Y | L |
| 14 | 普遍接受-测试.top/我的页面 | Y | Y | B | B | B | B + E | Y | E |
| 15 | ua-test.世界/我的页面 | Y | Y | B | B | B | B + E | Y | E |
| 16 | 普遍接受-测试.世界/我的页面 | Y | Y | B | B | B | B +L | Y | L |
| 17 | 普遍接受-测试。世界/我的页面 | Y | Y | A | A | B | A | A | E |

— As the results indicated, all browsers – except for Internet Explorer on desktop – showed certain issues resolving searches and displaying results properly.

— The findings indicate that while browser developers are making progress toward becoming UA-ready, there is still more work to do.

# READINESS STATUS AND CHALLENGES OF VARIOUS SERVICES

Today, the Internet has expanded to include domain names represented in different languages and scripts of the world including Indian languages. India already has ccTLD in 22 scheduled languages of India and one can have emails. UA is a best practice that ensures all applications, devices and systems accept and process all domain names and email addresses regardless of the chosen language or identity.

Due to the rapidly changing domain name prospect, many applications, devices and systems, due to the various gaps still do not recognize or appropriately process local language domain names or associated email addresses. These gaps results are demotivating, provide inconsistent experiences for Internet users, and limit an organization's ability to connect with users globally. In order to have a seamless experience, all components must support UA.

Below are the details of various services which need to be UA ready for an ideal experience of UA. These services have various Internet layers/levels involved. Each Service may have one or more layers.

The stack of technologies needs to be upgraded and reviewed to make the whole internet system Universal acceptance (UA) ready. The following are the majority of the top to bottom levels:

- Applications and Websites (Wikipedia.org, ICANN.org, Amazon.com, custom websites globally, PowerPoint, Google-Docs, Safari, Acrobat, custom apps)
- Social Media and Search Engines (Chrome, Bing, Safari, Firefox, local (e.g., Chinese) browsers, and Facebook, Instagram, Twitter, Skype, WeChat, WhatsApp, Viber)
- Programming Languages (JavaScript, Java, Swift, C#, PHP, Python and Angular, Spring, .NET core, J2EE, WordPress, SAP, Oracle)
- Platforms, Operating Systems(OS, and System Tools (iOS, Windows, Linux, Android, App Stores, Active Directory, OpenLDAP, OpenSSL, Ping, Telnet)
- Standards and Best Practices (IETF RFCs, W3C HTML, Unicode CLDR, WHATWG, Industry-based standards (health, aviation, ...)

These layers are part of the various services /services discussed in this document and needs to be upgraded and reviewed. Below are the brief summary of the various services and its readiness state in India.

## A. HARDWARE

Hardware includes both "traditional" computers (such as laptops and desktops) as well as mobile devices (such as phones and tablets). Hardware is UA ready, if the technology stack of the Operating systems is capable to accept / input, validate, process, store, and display all domain names and email addresses including those in Indian languages.

## B. SOFTWARE / OS

An operating system is a computer program that works as an interface between user and hardware and provides standard services for computer programs. Application Software is one type of software that runs or executes as per user request. High-level languages such as java, c, c++, etc. are used to develop the application software.

UA Readiness supported by Programming languages:

| Language | Library Name | Type of Test |
|---|---|---|
| C | libcurl | Email Syntax |
| C | libidn2 | ASCII to/from Unicode |
| C# | mailkit | Email Syntax |
| C# | microsoft | ASCII to/from Unicode |
| Go | idna | ASCII to/from Unicode |
| Go | mail | Email Syntax |
| Go | smtp | Email Syntax |
| Java | commons-validator | Email Syntax, Domain Name Syntax |
| Java | guava | Domain Name Syntax |
| Java | icu | ASCII to/from Unicode |
| Java | jakartamail | Email Syntax |
| Java | jre | ASCII to/from Unicode |
| JavaScript | idna-uts46 | ASCII to/from Unicode |
| JavaScript | nodemailer | Email Syntax |
| JavaScript | validator | Email Syntax, Domain Name Syntax |
| Python3 | django_auth | Email Syntax, Unicode ID |
| Python3 | email_validator | Email Syntax |
| Python3 | encodings_idna | ASCII to/from Unicode |
| Python3 | idna | ASCII to/from Unicode |
| Python3 | smtplib | Email Syntax |
| Rust | idna | ASCII to/from Unicode |
| Rust | lettre | Email Syntax |

***Table** 1: Level of UA Support by Programming Language Libraries*

UA-READINESS OF NETWORKING COMMAND LINE TOOLS

| Tool | MacOS 10.14 (BSD/Mach) | FreeBSD 12 (BSD) | Ubuntu 18 (linux) | Centos7 (linux) | Windows 10 |
|---|---|---|---|---|---|
| Host | No | No | No | Yes*(D) | |
| Ping | Yes* | No | Yes*(D) | Yes*(D) | Yes |
| ping6 | Yes* | No | Yes*(D) | Yes(D) | |
| Traceroute | Yes* | No | Yes*(D) | Yes(D) | |
| traceroute6 | Yes* | No | Yes*(D) | Yes(D) | |
| Dig | No | | No | Yes*(D) | |
| nslookup | No | | No | Yes*(D) | No |
| telnet | Yes* | No | No | | |
| openssl | Yes* | No | Yes* | No | |
| gnutls-cli | | Yes | Yes | | |
| tracert | | | | | Yes |

***Table** 2: Level of UA Support by Some Networking Tools*

## C. BROWSER

Browser, being application software, which enables users to navigate the Internet, is a key component in multilingual internet adoption. User both in the web and mobile user browsers to search for information and to view and navigate websites. Browser has also become a container in which many other applications run. For example, Web-mail applications, cloud-based office and productivity software, online learning apps, etc.

Hence, UA readiness of the browsers and proper handling of multi-lingual support is necessary for users to access the internet as well as the cloud applications in their local language with ease.

### UA-Readiness of Browsers

Light green represents the most successful browsers in terms of UA-readiness.

| | 360 | Amigo Mail | Atom Mail | Chrome | Edge | Epic Privacy Bro | Firefox | Internet Explorer |
|---|---|---|---|---|---|---|---|---|
| Windows | | | 1st | 3rd | 7th | 3rd | 6th | |
| Mac OS | | | | 2nd | 5th | 6th | 3rd | |
| Linux | | | | 2nd | | | 1st | |
| Android | | | | 3rd | 6th | 2nd | 4th | |
| iOS | | | | 2nd | 6th | 4th | 3rd | |

| | Opera | Safari | Samsung Browser | Sogo | UC Browser | Yandex |
|---|---|---|---|---|---|---|
| Windows | 4th | | | 5th | | 2nd |
| Mac OS | 4th | 4th | | | | 1st |
| Linux | 2nd | | | | | |
| Android | 2nd | | 5th | | 7th | 1st |
| iOS | 5th | 2nd | | | | 1st |
| | | | | | | |

*Yandex and Chrome score highly across all platforms.

**UA-Readiness of Browsers:** Desktop vs. Mobile Device

| | Passed all tests | Failed tests |
|---|---|---|
| Windows | 86 | 66 |
| Mac OS | 42 | 91 |
| Linux | 9 | 48 |
| Android | 48 | 104 |
| IOS | 31 | 102 |

*More Universal Acceptance difficulties with mobile device environments.

*Most failures were due to the following tests

*Confirm that the URL display is in the correct format as added, and

*Confirm that the URL is displayed correctly in the bar.

Top Web-Browser market share in India[4]

| Chrome | Firefox | Edge | Safari | Opera | UC Browser | IE | Edge Legacy | Chromium | Mozilla | Other |
|---|---|---|---|---|---|---|---|---|---|---|
| 86.36 | 5.93 | 3.55 | 1.78 | 1.45 | 0.34 | 0.29 | 0.19 | 0.05 | 0.02 | 0.03 |

Other browsers used in India:

| | |
|---|---|
| JioPages | https://www.jio.com/en-in/apps/jio-pages |
| Epic Privacy Browser | https://www.epicbrowser.com/ |

## D. E-MAIL

- Electronic mail (email or e-mail) is a method of exchanging messages ("mail") between people using electronic devices.
- The e-mail service hosted to a particular domain name, is configured and operated by the domain name owner; it is up to the domain owner whether to comply with the Internationalized Email updated RFCs.
- Not every email service provider uses its in-house software implementation.

4 https://gs.statcounter.com/browser-market-share/all/india

- Most of the e-mail services use some of the Standard Email Protocol implementing softwares e.g. Dovecot, Postfix, Open Exchange, Sendmail.
- As far as standard implementations of Email Protocols are concerned, some have upgraded their current versions to comply with the Internationalized E-mail protocols. However, to upgrade to the next version of this softwares is a call that rests still with the Email service provider.
- An e-mail is a two point communication, unless both the sending and receiving e-mail services comply with the Internationalized E-mail protocols, it could seem as a case of non-acceptance/non-functioning.
- This holds true even if one of them would be fully complying with the IMA protocols.
- Overall ecosystem issue which cannot be resolved till the time "most" of the Email service providers start complying with the same.

Below is the UA Readiness status of Email services and Email software solution in India

**Email services:**

- Zoho–No support for EAI L1 and L2
- Rediff–No support for EAI L1 and L2
- RajMail–Full support EAI
- XgenPlus–Full Support for EAI
- DataMail–Full support for EAI
- bsnl.in–Full Support for EAI
- dataone.in–Full Support for EAI
- Karnataka mail–Full Support for EAI

Email software solution (SMTP, IMAP, POP, webmail):

- XgenPlus–Full EAI support with downgrading capable
- Microsoft Exchange : L1 support
- Zimbra : No support

UA Case Study on EAI:

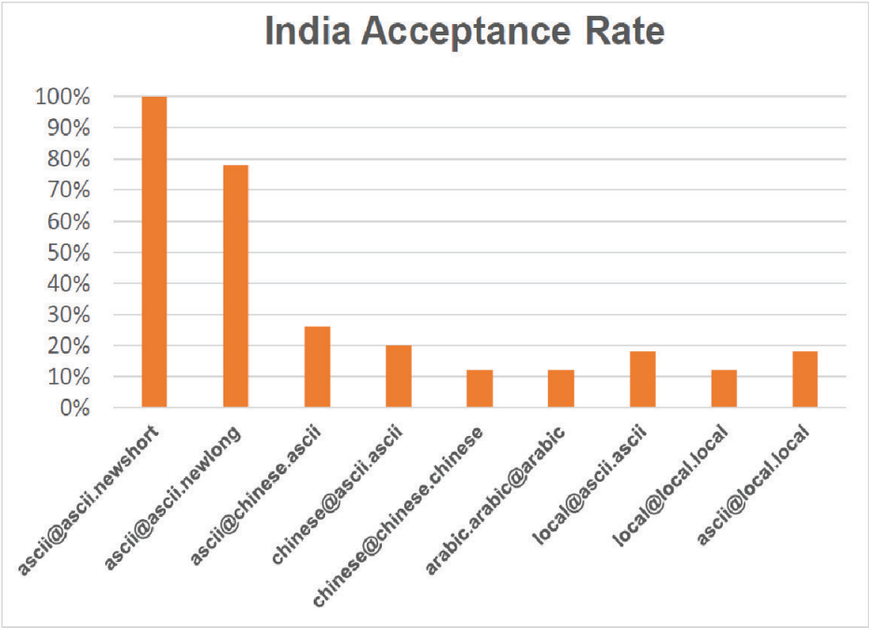
https://uasg.tech/download/uasg-013e-government-of-rajasthan-rajmail-en

https://uasg.tech/download/uasg-013c-2-icann-case-study-en

https://uasg.tech/download/uasg-013d-data-xgen-technologies-pvt-ltd-en

## E. WEBSITES

A website is a collection of web pages and related content that is identified by a common domain name. UASG conducted Global Evaluation of Websites for Acceptance of E-mail Addresses in the year 2017, 2019, 2020.




## F. SECURITY SYSTEMS

As the Internet has become a critical resource with constant security attacks and threats, the DNS has also been attacked and threatened. However, use of new protocols, developments and operational best practices has increased the resilience, stability and security of the DNS protocol and the global DNS infrastructure.

Security is an important element in ensuring safe and secure access to applications and services. Since the advent of the Internet, cyber security has evolved to a robust level through multilevel protections like perimeter security, end system security, application specific security, mechanisms for data protection, communication security, event monitoring & management, incident handling and cyber forensics.

- Domain name service based blackhole lists (DNSBLs): to see if your hostname or IP addresses are listed on major anti-spam DNS blacklist databases.
- OS patches and upgrades on a regular basis and constantly trained for encountering new threats and updating antivirus signatures.
- To protect a non-English language website/ web application, some of the perimeter security solutions (e.g. Web Application Firewall) have a feature called utf8toUnicode that helps to normalize data for inspection.

Antispam Test Results:

| Name | Spam |
|---|---|
| Spamassassin 3.4.5 | All L2 |
| Mailchannels | Part L1 |
| Spamjadoo (Xgenplus) | All L2 |

***Source:*** *https://uasg.tech/download/uasg-030a-eai-software-test-results-en/*

## G. APPS / SOCIAL MEDIA

An app (short for application) is a type of software that can be installed and run on a computer, tablet, smartphone or other electronic devices. Apps are a framework of services that application programs rely on for standard operations. A Social Media app is used for sharing of ideas, thoughts, and information through virtual networks and communities.

Below is the UA Readiness status of Apps (including Social Media) in India

**Social Media:**

- Koo–UA Ready
- Chingari–Not UA ready

**Top apps:**

- Paytm–Not UA ready
- Flipkart–Not UA ready
- MakeMyTrip–Not UA ready
- MyAadhar–Not UA ready
- Videomeet–UA ready

A detailed report covering the readiness state and challenges of covering various services has been given in the Annexure II.

# TEST BED, COMPLIANCE AND MOTIVATIONAL MEASURES

**Test Bed and Compliance:**

- Defining the Process / methodology for UA compliance testing
- Availability of Automatic, semi automatic Tools for compliance testing
- Portal for dissemination including recommendations, best practices, policies, guidelines, support team (UA ambassadors, experts), links to existing technical documents, discussion forum, support email, showcasing, UA champions, fellows, tools and technologies

**Motivational measures:**

The UA of all domain names and email addresses requires that all software accept, validate, process, store, and display them correctly. The Universal Acceptance Readiness Framework guides on implementing UA-readiness and testing it using a gating approach to verify UA conformance of an application.

NIXI along with the UASG may like to issue "UA Readiness Index", and / or "UA logo badge / UA Ready Badge" program which can be displayed on websites which have become UA ready. UA readiness levels can be defined such as UA-A, UA-AA & UA-AAA (similar to W3C).

# CAPACITY BUILDING – UA Champions, UA Ambassadors & UA Icons

### UA Champions

Tentative list of Indian apps which can become UA champions (but not limited to)

- Social Media Apps
  - » Sharechat
  - » Koo
  - » Chingari
  - » Josh
- Utility Apps / Citizen Centric Apps
  - » Flipkart
  - » Bharatpe
  - » Bhim
  - » Aadhaar
  - » GEM
  - » e-Sign

### UA Ambassadors (Appointed by ICANN):

People who are well-informed on UA and have spoken publicly and privately to their peers in the IT world on the topic.

### UA Icons (Appointed by MeitY& NIXI)

Promotion and awareness through social media (join hands with well-known names/ entities), for encouraging people to use Indian languages, including UA, on the Internet.

# ELEMENTS OF THE ROADMAP

**For achieving Universal Acceptance–Below steps would ensure that domain names and the identifiers built on them are useable in all services/applications and all services/ applications impacting the user must work**

- Objective I: Acceptance of IDN / EAI in various forms (registration, subscription, various services, etc)

  **Example**: mygov.in site Register page needs to accept EAI as valid email.

- Objective II: Acceptance and processing (sending / receiving) of EAI–L1 level compliance

  **Example:** ceo@nixi.in could receive email from EAI address and successfully be able to reply to the sender too.

**Creating the Ecosystem for UA–the goal of these steps would be that more websites and email-ids will be in Indian Languages and thus support UA across all applications.**

- Objective I: All Govt websites to have Internationalised Domain Names including all the resources, linkification to IDN in the website contents

  **Example:** mygov.in site Hindi content could be on मेरीसरकार.भारत

- Objective II**:** All Govt apps to be in at least three languages i.e. English, Hindi and Regional language of the state. If it is a GOI app, it should be in English and All official languages.

  **Example:** https://socialsecuritymission.gov.in/ may host their Malayalam website on "സാമൂഹ്യസുരക്ഷമിഷ്.സർക്കാർ.ഭാരതം" and also available their content in Hindi

  **Example:** Vikaspedia (https://vikaspedia.in/) has their content available in 22 scheduled Indian Languages. Language content should be host on local language domain names

- Objective III: Acceptance and processing (sending / receiving) of EAI–L1 level compliance plus creation of mailboxes for email service providers.

**Example:** ceo@nixi.in could also have सीईओ@नक्सी.भारत email address and people from all sort of valid email addresses could communicate with सीईओ@नक्सी.भारत email addresses.

**The above steps will ensure that a "sense of urgency" and a continuous momentum will be created among various stakeholders for supporting Universal Acceptance. However for multilingual Internet, more study, brainstorming will be needed for putting forward the recommendations. Below is just a mere example of the same. More elaborate recommendations will be covered in the next report (Phase-II).**

- Accessing current status of language implementation, usage and gap analysis
- Standards and compliances (Storage, inputting, display, Transliteration and others)
- Linguistic resources and standards (Dictionary formats, POS, corpus, and many more)
- Drive for content creation
- Discoverability of the multilingual contents (search, cross language, )
- Human inspiring systems (Voice driven, Machine Translation, Natural language understanding, Natural language Generation)
- Governance services at door step in one's own language
- Reaching the unreached
- Initiation of Research in various domains
- Awareness / Training / increased usage, language technology in basic education
- All round presence and usage of languages in Digital Medium (platforms, programming languages, databases, social media and many more)
- Security considerations
- Providing tools and technologies for creating and accessing the internet ecosystem in local languages in a safe and secure manner.

# AWARENESS CREATION AND TRAINING PROGRAMMES

**Training Programmes:**

In order to have awareness as well as technological implementation a series of awareness and technical programs needs to be planned. The training programs will raise awareness of Universal Acceptance (UA) challenges and remediation efforts and engage with key industry stakeholders

Training programmes can be planned by NIXI, UASG in association with industry experts to raise awareness of UA challenges in the Indian context, help build capacity to address those challenges, and assist ccTLDs and other stakeholders in developing and implementing robust IDN and UA-related systems and policies. Each program may have a series of training sessions on the technical aspects of UA and serve as a forum to discuss how to effectively address UA issues specific to the Indian region.

The online sessions, delivered by leading industry experts, will address various UA and IDN-related topics each targeted at audiences including policymakers, technical administrators, ccTLD managers, registrars and their resellers, and local regulators and businesses

Given the complexity of Indian writing systems, safeguards will be of primary importance to ensure that the common man is secure as far as possible from pharming and spoofing attacks. The syllabic structure of Indian languages, variant tables, restriction rules are some of the major focal areas of IDN, which will be thrown open to the major players for their comments. The key aspect here is to reach a consensus on the mode of implementation so that the user community can benefit at the earliest.

UA programmes become more crucial in multilingual and multicultural country like India having 22 scheduled Indian languages, and 15 ccTLDs (covering 22 Scheduled Indian Languages) with a large number of potential IDNs.

**Target Audience:**

- Software Developers
- Software Platform Developers (e.g. Java, PHP, etc)

- E-Commerce (e.g. Paytm, Flipkart)
- Social Media Apps
- Government
- Registrars & Registries
- Academia (Colleges and Universities)
- Suggestive Training modules:
- Demystifying Universal Acceptance - An introduction to the fundamentals of UA and EAI.
- Internationalized Domain Names (IDN) in context of Indian languages
- Email Address Internationalization (EAI) and its implementation.
- Programming for supporting Universal Acceptance
- UA readiness – for Registries & Registrars
- UA for Developers–Technical (software developers)–A detailed training on how to design and develop applications and systems to support UA.

**Additionally, for decision makers following may be covered:**

- Policies (Registry, Registrar, decision maker, others)
- Recommendations
- Various studies and its report
- .IN Reserved Names
- Variant generation

**Additionally, for Developers following may be covered:**

- Indian languages nuances
- Encoding mechanisms, Unicode, UTF-8, ASCII, ISCII
- Fonts, inputting mechanism
- IDNA libraries, Challenges and Solutions

# RECOMMENDATIONS

This part covers the various recommendations for stakeholders for UA and Multilingual Internet ready

This section is divided into two parts viz. Implementation plan which will cover the details of funding and related stakeholders and plan for short term, medium term and long term as described above.

## Short Term

### UA Program backgroumd work & Initiation

- MOU with USAG (ICANN) for support and use of resources
- Development of Basic knowledge dissemination portal
- Operationalization of support Team & SOP
- Initiation of Centre for excellence in "UA"
- Infographics, brochures, flyers
- Social media, reach out,
- Announcements of awareness campaigns
- Public announcement of "UA Program" by MoIT and Stakeholders

### Support System (NIXI as Nodal org)

- Formation of "Central Monitoring and Measurement Committee", industry participation and interface by NASSCOM, FICCI and others.
- NIXI & NIC to have a mechanism for registering and making functional IDNs under sarkar.bharat for Govt. sites
- Support to Govt. websites authorities for possible IDNs as per policy including variants (if any)
- Registering and making functional websites with IDNs - minimum 3 languages. Joint efforts of NIXI & NIC.
- Support to early movers, private players, social media platforms and others

## Long Term

### EAI – L2 level

- EAI –L1 plus creation of EAI mail boxes creation.
- SPAM filter in Indian languages
- Accessing DNS security challenges and possible remediation
- Research, design, development and deployment of indigenous "India specific emailing system"

01 02 03

### IDN / EAI Acceptance

- Establishment of Centre for excellence in "UA" and capacity building
- Assessing resource requirements & suggesting possible remediation to website authorities
- IDN / EAI acceptance in various forms viz. registration, feedback, support, subscription and others.
- Policy for UA readiness - (as example GeM procurements, private players internal procurement policies)

### EAI - L1 level

- Portal with advanced tools / technologies and reference implementation (s) with various libraries support.
- Sending and receiving emails from and to "Internationalised email addresses"
- Private – 1000 and Govt (state/Centre) – 4000 websites / applications, specifically e-commerce / services.
- Compliance and motivational measures
- Initiation of development of SPAM filters for major Indian languages

## Medium Term

| | |
|---|---|
| Short | 3 to 9 months |
| Medium | 6 to 18 months |
| Long | 12 to 36 months |

**Implementation plan:**

The recommendations are divided into three main categories viz. short term, medium-term and long term. Some recommendations are continuous and will also stand valid across the roadmap period.

## SHORT TERM (3-9 MONTHS)

| Activities | Detailed Description of the Activities | Proposed Implementation agency (ies) | Duration | Tentative budget in Rs. lac | Remarks |
|---|---|---|---|---|---|
| UA Program background work and Initiation<br><br>Support System (NIXI as Nodal org) | Formation of initial working Team to plan and kick off the activities | NIXI | One month | | |
| | Preparation and finalisation of Detail Project Report | NIXI<br>(Along with current committee and MeitY) | | | |
| | MOU with UASG for resources and reference implementation, tools and technologies | NIXI, MeitY, UASG (ICANN) | | | |
| | Formation of "Central Monitoring and Measurement Committee". | NIXI<br>With participation from<br>FICCI, NASSCOM NIXI, MeitY<br>UA Ambassador<br>UASG<br>Subject experts | | | |
| | Finalisation of Portal features and contents (Basic plus advance portal) | NIXI with committee | | | |
| | Initiation for establishment of the "Centre of Excellence (CoE) in UA" at academia(s) premises as per the MeitY set processes | MeitY, NIXI | Three months | | Preparation of DPR, call for proposals for "Centre for Excellence in "UA" |
| | Design, development and deployment of basic information and dissemination portal including contents | NIXI / C-DAC / UASG | | 25.00 | |

| Activities | Detailed Description of the Activities | Proposed Implementation agency (ies) | Duration | Tentative budget in Rs. lac | Remarks |
|---|---|---|---|---|---|
| | Reach out plan / promotional material - flyer / brochure / video & multimedia kit for UA | NIXI | Six months | 120.00 | |
| | Establishment and Operationalisation of support Team & SOP | NIXI | | 241.50 | Operational and infra costs are not included.<br>Assuming that the support team will be in place by first three months |
| | Establish a mechanism for registering and making functional IDNs under sarkar.bharat for Govt. sites. | NIC and NIXI<br>State Governments & UT | | | Minimum 3 languages IDNs<br>2000 Govt. of India websites<br>Support to Govt. websites authorities for possible IDNs as per policy including variants (if any) |
| | Support to early movers, private players, social media platforms and others | NIXI, FICCI, NASSCOM)<br>(with help of support Team | | | 250 private websites |
| | Equipment's / software | NIXI | Six months | 75.00 | |
| | | | | | |
| | First six months tentative budget | | | 461.50 | |

## MEDIUM TERM (6-18 MONTHS)

| Activities | Detailed Description of the Activities | Proposed Implementation agency (ies) | Duration | Tentative budget in Rs. lac | Remarks |
|---|---|---|---|---|---|
| **IDN / EAI Acceptance**<br>**EAI - L1 level** | Award of and functioning of Centre for excellence in "UA", capacity building | Academic / Research institutions | Six months | 75.00 | • The functioning of the Centre will start from second year of UA Programme.<br>• Budget earmarked for travel<br>• Grant in aid will be allocated from second year of the UA programme and for six months |
| | Design, development and deployment of portal with Advanced features including reference implementation | NIXI / CDAC | Six months | 50.00 | |
| | Awareness, Training and workshops | FICCI, NASSCOM, NIC, NIXI, and industry. | | 120.00 | Assuming one / two workshops every quarter. Minimum 6-8 programmes |
| | Support Team & SOP | | | 966.00 | |
| | **Next one-year tentative budget** | | | **1211.00** | |

## LONG TERM (12-36 Months)

| Activities | Detailed Description of the Activities | Proposed Implementation agency (ies) | Duration | Tentative budget in Rs. lac | Remarks |
|---|---|---|---|---|---|
| **EAI – L2 level**<br>**CoE :**<br>• SPAM filter in Indian languages<br>• Accessing DNS security challenges and possible remediation | EAI level one plus creation of EAI mail boxes | NIXI, FICCI, NASSCOM, State Governments and UTs | One year | | |
| | SPAM filter in Indian languages | CoE | One and half year | 375.00 | 15 Major Indian languages |
| | Accessing, research and remediation regarding DNS security challenges | CoE, CERT-in, NIXI | | 1449.00 | |
| | Support Team & SOP | NIXI, FICCI, NASSCOM | | | |
| | Awareness, reach out material | NIXI, FICCI, NASSCOM, | | 180.00 | |
| | **Next one and half year tentative budget** | | | **2004.00** | |
| | | | | | |
| | **Cumulative tentative budget for 3 years** | | | **3676.50** | |

## A. PROPOSED FUNDING MODEL:

1. MeitY / NIXI–Seed funds for initiation of the activities, more specifically
   a. Planning / management
   b. Portal development
   c. Contents
   d. Support Team
   e. Central Measurement and Monitoring committee
   f. Establishment of Centre for Excellence in "UA"
2. State Government may be approached to provide seed funds for activities in their region. Also State Government(s) additionally undertake Reach out / reach out material, campaigns, events and "support members" for their respective region / languages.
3. Encourage Industry bodies / private players to sponsor certain activities such as
   a. Reach out / reach out material
   b. UA campaigns / events
   c. "Support Team(s)" manpower / manpower costs
4. Industry bodies / private players may sponsor / donate (Platinum, Gold, Silver for a duration of time) the overall activity in cash from CSR funds. Due acknowledgement mechanism for contributions to the cause needed to be worked out.

## B. TENTATIVE BUDGET REQUIREMENTS (FOR FIRST YEAR OF OPERATION):

| Sr No | Description | Details | Tentative budget in Rs. Lacs for first year |
|---|---|---|---|
| 1 | Manpower | • 1 x Programme Head<br>• 2 x Domain experts<br>• 2 x Program manager – Government interface (State and Central), private interfacing and handholding and reporting to Central Measurement and monitoring committee<br>• 2 x Technical Heads (Industry / Government interface)<br>• 1 x General Admin<br>• 2 x technical content writers<br>• 22 x support team (1 member per language covering 22 scheduled languages)<br>• 1 x Techno / Managerial<br>• 3 x compliance testing members | 966.00 |
| 2 | Travel | Internal | 25.00 |

| Sr No | Description | Details | Tentative budget in Rs. Lacs for first year |
|---|---|---|---|
| 3 | Infrastructure / space | Not assumed, since current infra of NIXI (if available) may be used | 0.00 |
| 4 | Equipment's / Software | Systems, Printer, scanner, software, hosting platform, etc | 75.00 |
| | Material | Reach out material, media campaign, etc | 100.00 |
| 5 | Sub total | | **1166.00** |
| 6 | Overheads | @ 20 % | 233.20 |
| | Total | Rounded off | **1400.00** |

*Following recommendations are largely for Private sector to consider.*

For manpower, following may be pursued

- Parttime engagement with UASG ambassadors for overall support in implementation
- Engagement of volunteers (from industry, academia) specifically for support
- Engagement of language experts as volunteers / parttime / as per the need.

Following recommendations are largely for Private sector to consider.

## BEST PRACTICES:

Recommendations to catalyse promote of UA and multilingual Internet

- Different language website(s) to map and support language domain name(s) and EAI
- Indian language advertisement to mention the relevant IDNs and EAIs
- Use EAI and IDN on their letterheads / visiting cards.
- Upgrade / update various software's, services for UA readiness
- Government to procure UA Ready Solutions and Softwares
- Standardised Indian language (s) engraved keyboard to be made available by the hardware manufacturers.
- Default browsers in laptops and mobiles to be encouraged to have UA readiness.
- Mobile handset manufacturers must adhere to IS 16333 (Part 3) for Indian language support in Mobile Phones, as mandated by the MeitY.

# ANNEXURE - I

## List of scheduled Indian languages and major scripts used

| Sr. No | Language | Recognition in state | Major script used | Also uses |
|---|---|---|---|---|
| 1 | Assamese | Assam, Arunachal Pradesh | Assamese | This script is similar to the Bengali-Assamese script |
| 2 | Bengali | West Bengal, Tripura | Bengali | This script is similar to the Bengali-Assamese script |
| 3 | Bodo | Assam | Devanagari | |
| 4 | Dogri | Official language of Jammu and Kashmir | Devanagari | |
| 5 | Gujarati | Dadra and Nagar Haveli and Daman and Diu, Gujarat | Gujarati | |
| 6 | Hindi | Andaman and Nicobar Islands, Bihar, Dadra and Nagar Haveli and Daman andDiu, Chhattisgarh, Delhi, Gujarat, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Jammu and Kashmir, Mizoram, Rajasthan, Uttar Pradesh, Uttarakhand, and West Bengal | Devanagari | |
| 7 | Kannada | Karnataka | Kannada | |
| 8 | Kashmiri | Jammu and Kashmir | Perso-Arabic | |
| 9 | Konkani | Dadra and Nagar Haveli and Daman and Diu, Maharashtra, Goa, Karnataka, and Kerala (The Konkan Coast) | Devanagari | |
| 10 | Maithili | Bihar, Jharkhand | Devanagari | Mithilakshar or Tirhuta. This script is similar to the Bengali-Assamese script. |
| 11 | Malayalam | Kerala, Lakshadweep, Puducherry | Malayalam | |

| Sr. No | Language | Recognition in state | Major script used | Also uses |
|---|---|---|---|---|
| 12 | Manipuri | Manipur | Bengali | MeeteiMayek |
| 13 | Marathi | Maharashtra, Goa, Dadra and Nagar Haveli and Daman and Diu | Devanagari | |
| 14 | Nepali | Sikkim and West Bengal | Devanagari | |
| 15 | Odia | Official language of Orissa | Odia | |
| 16 | Punjabi | Official language of Punjab and Chandigarh, 2nd official language of Delhi and Haryana | Gurmukhi | |
| 17 | Sanskrit | Himachal Pradesh, Uttarakhand | Devanagari | |
| 18 | Santali | Spoken by Santhal people mainly in the state of Jharkhand as well as in the states of Assam, Bihar, Chhattisgarh, Mizoram, Odisha, Tripura, West Bengal | Devanagari | Ol-Chiki |
| 19 | Sindhi | Gujarat and Maharashtra, especially Ulhasnagar | Perso-Arabic | Devanagari |
| 20 | Tamil | Tamil Nadu, Puducherry | Tamil | |
| 21 | Telugu | Andhra Pradesh, Telangana, and Puducherry | Telugu | |
| 22 | Urdu | Jammu and Kashmir, Telangana, Jharkhand, Delhi, Bihar, Uttar Pradesh, and West Bengal | Perso-Arabic | |

# ANNEXURE - II

## Readiness Status and Challenges of various services in implementation of UA Support

## A. HARDWARE

Hardware includes the physical parts of a computer, such as a case, central processing unit (CPU), monitor, mouse, keyboard, computer data storage, graphics card, sound card, speakers, and motherboard. Hardware includes both "traditional" computers (such as laptops and desktops) as well as mobile devices (such as phones and tablets). Software is the set of instructions that can be stored and run by hardware. An operating system is the main software that sits between Hardware and User.

### Scope of work in Hardware space

- Personal Computer- Desktop
- Personal Computer- Laptop
- Smartphones with touchscreen
- Phones without touchscreen
- External hardware keyboards
- Adoption and popularization

### Readiness States:

For achieving Universal Acceptance and Multilingual Internet for Hardware including traditional computers such as desktops, laptops, etc., and also mobile devices such as phones and tablets, a variety of different software platforms should be included. These hardware devices come with Operating Systems that fall either in open source or proprietary platforms. Some popular platforms are Windows, MacOS, Linux (Ubuntu, Red Hat, SUSE, Mandrake, BOSS etc.) and BSD on personal computers and iOS and Android on smartphones. Hardware is UA ready, if the technology stack of the Operating systems is capable to accept / input, validate, process, store, and display all domain names and email addresses including those in Indian languages.

**Advisable:**

- The government may motivate and encourage key manufacturers to launch at least one Desktop/Laptop/External Keyboard model with Indic characters inscribed on their keyboards. This will help reduce the entry barrier for UA.
- Review of existing support for Indic keyboard layouts in major PC/mobile devices and keyboards
- Establishment of a support system for companies that need know-how to implement the below recommendations
- General advice to manufacturers and developers to support IDNs and IEAs in IOT devices, wherever applicable
- The government should support schemes for free distribution of InScript/default keyboard stickers for Indian Languages.

**Recommendations**

- Hard and/or soft keyboards (in desktop computers, laptop computers, tablets, hybrid devices, smartphones and any other touchscreen-enabled productivity/communication devices, and external keyboards) should be able to support the operating system's mechanism for Indic text input for IDNs and IEAs, in the form of:
  - » Unicode support
  - » Support for standard keyboard layouts in Indian languages (Inscript across India, and additionally, Tamil99 (Tamil Nadu) and KPG (Karnataka)
- Mobile handset manufacturers must adhere to IS 16333 (Part 3) for Indian language support in Mobile Phones, as mandated by the MeitY.

## B. EMAIL

**Readiness State:**

- As the e-mail service hosted to a particular domain name is configured and operated by the domain name owner, it is up to the domain owner whether to comply with the Internationalized Email updated RFCs.
- Not every email service provider uses its in-house software implementation.
- Most of the e-mail services use some of the Standard Email Protocol implementing softwares e.g. Dovecot, Postfix, Open Exchange, Sendmail.
- As far as standard implementations of Email Protocols are concerned, some have upgraded their current versions to comply with the Internationalized E-mail protocols. However, to upgrade to the next version of this softwares is a call that rests still with the Email service provider.
- An e-mail is a two point communication, unless both the sending and receiving e-mail services comply with the Internationalized E-mail protocols, it could seem as a case of non-acceptance/non-functioning.
- This holds true even if one of them would be fully complying with the IMAP protocols.

- Overall ecosystem issue which cannot be resolved till the time "most" of the Email service providers start complying with the same.

### Current scenario

- Standards, guidelines and protocols (IDNA 2008–RFC 6530, RFC 6531, RFC 6532, and RFC 6533, define mechanisms and protocol extensions needed to fully support internationalized email addresses.)
- Various components / subsystems–Free and open source / commercially available and its status

### Security concerns / Challenges

EAI support to be extended by the Email service providers, there seems to be a bit of reluctance to the same. Mostly because of the following factors.:

- A fairly small share of users is demanding such a service
- These companies have their own upgradation plans and Internationalized Email support is not an immediate priority.
- Technical challenges (specifically Indian languages complexities) – Mostly taken care of, can be augmented with regional requirements by way of policy.

### Recommendations

- **Implementation plan**
  - » Organisation level–L1 & L2 together
    - — (mailbox rules – less user to manage, can be stringent and easily manageable)
- Free email services–L1 and then L2
  - » (mailbox rules – large user base, generic, can be confusing at times with linguistic variations)
- **Mail box recommendations:**
  - » Generic recommendations from ICANN
  - » Specific regional recommendations (NIXI / Govt. of India)

## C. SECURITY SYSTEMS

### Introduction to Security aspects

Cyber security is an important element in ensuring safe and secure access to applications and services. Since the advent of the Internet, cyber security has evolved to a robust level through multilevel protections like perimeter security, end system security, application specific security, mechanisms for data protection, communication security, event monitoring & management, incident handling and cyber forensics.

However, in the context of Universal Acceptance, it becomes even more important as the notion of largely ASCII based internet and applications get extended to Unicode that would encompass support for various languages beyond English. This has a direct impact on the

standard protocols, systems and applications to gear up for necessary support. There are solutions already available supporting such transformations through PunyCode which are implemented in many of the standard applications and protocols for Web services, DNS and Email. Global level security concerns surrounding IDN and EAI related implementations include phishing, spam, semantic attacks, typo-squatting, malicious browser extensions, etc.

Email Address Internationalization (EAI) is a process that enables email addresses with either the domain name or mailbox name in different languages and scripts to work properly when sending and receiving emails. These may include email addresses represented in Hindi, Chinese, Japanese or any other scripts. For instance, if an individual tries to send email (sender) to other individual (recipient) using an email address like that of मेलएडमनि@सीडैक.भारत, the message would never be received by the recipient if the email box is not EAI ready (UTF-8 enabled) and they would not be able to communicate with each other. On the other hand if recipient's mail servers are UTF-8 enabled, then they may communicate in native language.

Similarly, Internationalized Domain Names is the ability of registering domain names using characters available in Unicode. Domain names were until some years back possible using Latin Script only. However, now domain names like that of सीडैक.भारत are possible to register.

Use of email addresses and domain names that are in local scripts poses great security issues. For example, DoS/DDOS attack, spamming, phishing attack and other malicious attacks. In order to prevent the attacks while using EAI email and IDN services, it is required to ensure all the security measures are in place.

## Scope

The scope includes capturing the security issues related to Internationalized Domain Names (IDN) and Email Address Internationalization (EAI) with respect to Indian languages. This also covers the readiness state of various components associated with IDN and EAI and recommendations for successful implementation of the same.

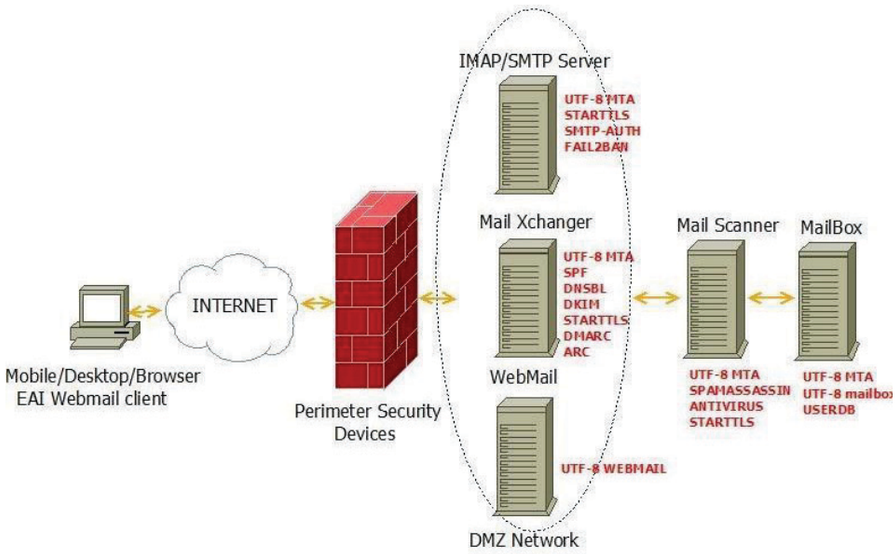



**Figure 1** – Security components for IDN and email

## Security concerns and challenges

### 1. Phishing

**Homograph attack:** Homograph attack is a way a malicious party may deceive computer users about what remote system they are communicating with, by exploiting the fact that many different characters look alike. Homograph attack is based on creating URLs from different language character sets, although technically homoglyph is the more accurate term for different characters that look alike.

For example, Homographs in Indian Languages.ईश्वर and ईश्वर, क्रमशः and क्रमशः In first example, ई can be written in two ways. Similarly, in second example, शः is written in two different ways. It is difficult for users to understand the difference.

**Typosquatting:** Look-alike website URL that appears similar to the genuine URL of an established organization but actually contains a typo. Users may type by mistake, having been lured there by a phishing scam.

**Semantic Attack:** Semantic attack involves IDNs created for malicious motives by translating English brand names to other languages or the adversaries combine brand domain names with keywords from another language to create IDNs.

**Alternative writing styles in Indian languages:** Some of the Indian languages have alternate writing options which may lead to phishing attacks.

e.g. हिन्दी हिंदी, कम्पन = कंपन

**Zero Width Joiner (ZWJ) and Zero Width Non Joiner (ZWNJ) issue for Unicode:** ZWJ and ZWNJ leads to phishing attacks / (receiving from and responding to wrong email address)

The zero-width non-joiner (ZWNJ) is a non-printing character used in the computerization of writing systems that make use of ligatures. When placed between two characters that would otherwise be connected into a ligature, a ZWNJ causes them to be printed in their final and initial forms, respectively. It is represented by the Unicode code point U+200C.

The zero-width joiner (ZWJ) is a non-printing character used in the computerized typesetting of some complex scripts such as the Arabic script or any Indic script. When placed between two characters that would otherwise not be connected, a ZWJ causes them to be printed in their connected forms. It is represented by the Unicode code point U+200D.

Halant is used to form all “consonant clusters” by inserting the character “virama” (or “halant”) between the two relevant consonant letters. It is represented by Unicode code point U+094D.

For typing Halant, ZWJ and ZWNJ we have to use the following sequence.

Halant = ् (Placed over D key), ZWJ = CTRL + SHIFT + 1, ZWNJ = CTRL + SHIFT + 2

| Normal Character Sequence | Character SequenceWith ZWJ | Character SequenceWith ZWNJ |
|---|---|---|
| प + ् + र = प्र | प + ् + ZWJ + र = प्‌र | प + ् + ZWNJ + र = प्‌र |
| ष + ् + ट = ष्ट | ष + ् + ZWJ+ट = ष्‌ट | ष + ् + ZWNJ + ट = ष्‌ट |
| त + ् + र = त्र | त + ् + ZWJ+ र = त्‌र | त + ् + ZWNJ + र = त्‌र |

### 2. Spam and Scam emails–

It is observed that spam filters do not identify Indian language-based spam/scam emails due to lack of resources viz. spam email corpus, regular expressions, spam word dictionaries, etc. in Indian languages. The spam/scam emails may be in the form of images also.

### 3. Policy aspects–

Currently there exists a policy for "Internationalized domain names in Indian languages". With respect to the local part of email addresses, for example मेलएडमनि in मेलएडमनि@सीडैक.भारत, it is expected that the respective email administrators have their appropriate policies.

### 4. Implementation of User Agent–

mail UA UI rendering is a significant part that affects the user's perception of an email's authenticity. Most of the webmail and email clients only show From header without any more authentication details. Therefore, it is difficult for ordinary users to judge the authenticity of emails.

## State of readiness

Here we bring out the existing solutions and mechanisms available pertaining to security for IDN and EAI.

### Overall Security

- Domain name service based blackhole lists (DNSBLs): to see if your hostname or IP addresses are listed on major anti-spam DNS blacklist databases.
- OS patches and upgrades on a regular basis and constantly trained for encountering new threats and updating antivirus signatures.
- To protect a non-English language website/ web application, some of the perimeter security solution (for eg. Web Application Firewall) has a feature called utf8toUnicode that helps to normalize data for inspection.

### Email specific security

SMTP- UTF8 facility is already available as part of the email server stack and following to be taken care of.

- Enablement of support for UTF-8.
- Policy for EAI.
- Mechanisms and processes to create new signatures for full proof security.

- Put in place perimeter security to handle security issues related to IDN and EAI to a larger extent.

Following are specifically available security solutions that can be leveraged

- SMTP-AUTH with secure SMTP.
- STARTTLS should be enabled to ensures encryption of the communication between two email servers if both sides support STARTTLS
- Sender policy framework (SPF) : To authenticate the sender of an email
- Domain Keys Identified mail (DKIM): is a method of email authentication that helps prevent spammers and other malicious parties from impersonating a legitimate domain
- Domain-based Message Authentication Reporting and Conformance (DMARC) to reduce spam and spoof emails.
- Mail Scanner is an anti-spam/anti-virus server that scans mail before delivering it to the receiver's mailbox. It protects against viruses, spam, malware, and phishing.
- Continuously need to train UTF-8 based spamming content to anti-spam to avoid flooding of EAI based Spam mails on user's INBOX
- Email Aliases – Accept both Unicode and ASCII domain names in e-mail addresses and offers ASCII mailbox aliases along with EAI mailbox names (Multilingual identities mapped to single inbox)

### Recommendations

- Email Service Providers should make sure that MTA and MUA should be UTF-8 enabled along with security components mentioned in this report. (Universal Acceptance)
- UTF-8 Support for SPAM Filters (UA)
- To develop an eco-system for development of plugins for existing spam filters for handling Indian language spam/scam emails taking care of necessary resources for: (Multilingual Internet)
  - » Spam word dictionary for Indian languages, regular expressions, spam email corpus for training the AI/ML models for filtering.
  - » Develop filters for detecting Malicious Domains involving IDN characters
  - » Create DGA (domain generator algorithms) for deriving possible malicious domain names involving IDN characters for a given domain name.
- Agencies like CERT-In (Computer Emergency Response Team of India) to develop and host incident reporting portal for Indian language-based spam/scam emails and publishing attacks and advisories. (Policy and Multilingual Internet)
- The scope of Network & Security Audit assessment including VAPT (Vulnerability Assessment & Penetration Testing) of Application and hardware system should factor UA (IDN and EAI) related security aspects. (Policy and UA)

## D. SOFTWARE/OPERATING SYSTEM

### Application Software and Operating Systems:

Application Software is one type of software that runs or executes as per user request. High-level languages such as java, c, c++, etc. are used to develop the application software. Application software is specific purpose software intended

to perform some tasks grouped. Without an operating system, application software can not be installed. Its examples are Photoshop, VLC media player, PowerPoint, Google-Docs, Safari, Acrobat, custom applications, etc.

Several levels/layers of technologies make the Internet system ready for UA. Let's take one example to understand this in detail: if we can book railway tickets using the IDN@IDN. IDN email id. Still, if we cannot receive emails to verify the email id used for registration, it will not yield the solutions per the users' expectations. Hence, all the layers are somehow interconnected, we need to work on each layer, and one of the essential layers is the Software/Operating Systems layer.

An operating system is a computer program that works as an interface between user and hardware and provides standard services for computer programs. A computer system's entire process or functionality depends on the operating system. It is developed by using c++, c, assembly languages. An operating system performs various tasks like managing files and directory creation and deletion, process creation, deletion, synchronization, memory allocation, and deallocation. An operating system also prevents the computer system from unauthorized access and secures resources, information, and data. Its examples are Microsoft Windows, Linux, Unix, DOS.

#### Examples:

Modern operating systems have various command-line tools used in system management and program development. Many of these tools work on Domain Names and a few on email addresses. We look at these tools to see how well they support Universal Acceptance.

#### Readiness state

The tools generally accept domain names as arguments from the command line, and then use them in the tools' operation, which includes looking them up in the DNS. They all return some sort of report to the console, sometimes including the domain name, sometimes not. We say a tool can Accept and Validate a name if it receives a name from the command line and correctly recognizes it as an ASCII or IDN domain name. It can Process the name if it does something useful with it, typically a DNS lookup. Some tools put domain names in their output, so if they do so correctly, they can Display names.

#### Linux/BSD tools

Many non-Windows systems are derived from Unix and Linux. These include linux distributions such as Ubuntu and Centos, BSD systems including FreeBSD, OpenBSD, and NetBSD, and Apple's MacOS. (The Android system used on phones and tablets is derived from linux but does not generally include the command line tools described here so we don't consider it further.)

Many of the tools come from software packages that are bundled with the systems. This means that the version of those tools in any particular system depends on the version of the package they come from, so there's no need to test each tool individually, once we know which versions of the packages come with each system.

### DNS software packages

**ISC BIND** from the Internet Software Consortium is the most widely used DNS server package. Along with the server program, it includes the lookup tools dig and nslookup. Versions since at least 9.6 can handle IDN U-labels as a build-time option, although versions older than 9.11 don't usually have the option enabled.

**LDNS** is a package of DNS tools from NLNet Labs, who provide the popular nsd and unbound DNS servers. It includes a drill which is intended to be an alternative to dig. Currently drill does not support IDN U-labels.

### Cryptography packages

Cryptographic packages can have IDN issues in two areas. One is certificate management, managing TLS and S/MIME certificates. They may handle U-labels in the domains they include and EAI addresses in optional contact addresses and, for S/MIME certificates, the address the certificate is issued for. The other is connection debugging, tools that connect to web and other servers and check the certificates they use. The user provides the hostname of the server, which may include U-labels.

**OpenSSL** is the most widely used cryptography package. It has an openssl tool that has both options for certificate management and for connection debugging. The current versions as of February 2019 are 1.0.2r and 1.1.1b, although older versions remain in wide use. LibreSSL was forked from OpenSSL in 2014 and is separately maintained by the people who maintain OpenBSD. Its current versions are 2.9.0, 2.8.3, and 2.7.5 although older versions remain in wide use. MacOS includes LibreSSL 2.6.5.

**GnuTLS** is another popular package from the Free Software Foundation. It has gnutls-certtool for managing certificates and gnutls-cli for debugging TLS connections to other systems

UA Readiness Status:

Programming Language Libraries and Frameworks. The report tested three kinds of libraries for:

1. Unicode strings
2. Domain names, including IDNs
3. Email, to test for EAI

These findings are summarized in Table 1 below, with further details in the report and the test results. Red signifies not being UA-ready; pink signifies being UA-ready but some details need to be managed; and green signifies being UA-ready. The results show that programming languages Java, JavaScript, and Python3 have support for processing IDNs and EAI. Some additional platforms support IDNs but EAI is not supported by the remaining libraries.

| Language | Library Name | Type of Test |
|---|---|---|
| C | libcurl | Email Syntax |
| C | libidn2 | ASCII to/from Unicode |
| C# | mailkit | Email Syntax |
| C# | microsoft | ASCII to/from Unicode |
| Go | idna | ASCII to/from Unicode |
| Go | mail | Email Syntax |
| Go | smtp | Email Syntax |
| Java | commons-validator | Email Syntax, Domain Name Syntax |
| Java | guava | Domain Name Syntax |
| Java | icu | ASCII to/from Unicode |
| Java | jakartamail | Email Syntax |
| Java | jre | ASCII to/from Unicode |
| JavaScript | idna-uts46 | ASCII to/from Unicode |
| JavaScript | nodemailer | Email Syntax |
| JavaScript | validator | Email Syntax, Domain Name Syntax |
| Python3 | django_auth | Email Syntax, Unicode ID |
| Python3 | email_validator | Email Syntax |
| Python3 | encodings_idna | ASCII to/from Unicode |
| Python3 | idna | ASCII to/from Unicode |
| Python3 | smtplib | Email Syntax |
| Rust | idna | ASCII to/from Unicode |
| Rust | lettre | Email Syntax |

***Table 3:*** *Level of UA Support by Programming Language Libraries*

## UA-Readiness of Networking Command Line Tools

The UA-related support for the tools reviewed is summarized in Table 2 below. See the report for further details (https://uasg.tech/download/uasg-024-ua-readiness-of-command-line-networking-tools-en/). Yes means that the command at least accepts, validates, and processes. Yes(D) means it also displays domain names and shows IDNs as U-labels. Yes means that the command accepts IDNs but processes them with IDNA2003 rather than IDNA2008. An empty box means that the system is not distributed with that tool. No means IDNs as U-labels are not supported. An empty box means that the system is not distributed with that tool.

| Tool | MacOS 10.14 (BSD/Mach) | FreeBSD 12 (BSD) | Ubuntu 18 (linux) | Centos 7 (linux) | Windows 10 |
|---|---|---|---|---|---|
| Host | No | No | No | Yes*(D) | |
| Ping | Yes* | No | Yes*(D) | Yes*(D) | Yes |
| ping6 | Yes* | No | Yes*(D) | Yes(D) | |
| Traceroute | Yes* | No | Yes*(D) | Yes(D) | |
| traceroute6 | Yes* | No | Yes*(D) | Yes(D) | |
| Dig | No | | No | Yes*(D) | |
| nslookup | No | | No | Yes*(D) | No |
| telnet | Yes* | No | No | | |
| openssl | Yes* | No | Yes* | No | |
| gnutls-cli | | Yes | Yes | | |
| tracert | | | | | Yes |

***Table 4:*** *Level of UA Support by Some Networking Tools*

## Recommendations

The committee needs to work on the stack of technologies at different levels to make the whole internet system Universal Acceptance (UA) ready. The following are the two most basic layers/levels of the stack are as follows:

1. **Applications–**PowerPoint, Google-Docs, Safari, Acrobat, custom apps, etc
2. **Platforms, Operating Systems and System Tools**- iOS, Windows, Linux, Android, App Stores, Active Directory, OpenLDAP, OpenSSL, Ping, Telnet, etc

## Universal Acceptance

- Identification of IDNs and EAIs in command line tools such as host ping, ping6, traceroute, traceroute6, dig, nslookup, telnet, openssl, gnutls-cli, tracer.
- Productivity software should be able to identify an IDN as a domain name and EAI as an email address (by automatic hyperlinking).
- Major software to be reviewed: Microsoft Office, Google Docs, iWork and Open Office.
- Development software should accept IDNs and EAIs in the code and databases
- Development software should also enable the applications to take IDNs and EAI as form values
- Utility software such as FTP should accept IDNs as web/server addresses
- Acceptance of IDNs and EAIs while
- Registering the products (OS, Other software)
- Creating user accounts

### Multilingual Internet

- At least one standard Indic keyboard layout to support all Indic languages-
- Standard Indic Keyboard + One additional transliteration keyboard in major languages
- Additional mechanism (i.e. Character Map) in the OS to input characters not included in the keyboard

The committee may devise a testing framework to identify gaps and test the aforementioned layers by a stepwise mechanism. Therefore, the whole technology stack would need to be reviewed and upgraded, where required, to accept, validate, process, store, and display all domain names and email addresses.

## E. BROWSERS

A browser is an application program that provides a way to look at and interact with all the information on the World Wide Web or a website. With a focus on Universal acceptance of IDNs and IEAs in the browser, the key considerations, readiness and recommendations are listed below:

### Key Considerations

Accessibility: Is the browser and its features usable in the local language?

Input: Are we able to input IDNs and Unicode URLs in the browser and access the correct website / page?

Multilingual websites: Are we able to view and navigate local language websites using the browser?

Mobile: What are the unique considerations with respect to Mobile browsers?

### Readiness state

- UASG016: Universal Acceptance of Popular Browsers
  - » The UASG conducted a prior UA readiness evaluation for browsers that was performed in Q2 of 2017, the report is available here on the UASG.tech website: https://uasg.tech/wpcontent/uploads/documents/UASG016-en-digital.pdf
- UASG036: Universal Acceptance of Popular Browsers
  - » An updated version of the study on UA readiness on browsers was published in Jan 2022. This report is available here on the UASG.tech website: https://uasg.tech/download/uasg-036-ua-readiness-of-browsers-en/

### UA-Readiness of Browsers

Light green represents the most successful browsers in terms of UA-readiness.

| | 360 | Amigo Mail | Atom Mail | Chrome | Edge | Epic Privacy Bro | Firefox | Internet Explorer |
|---|---|---|---|---|---|---|---|---|
| Windows | | | 1st | 3rd | 7th | 3rd | 6th | |
| Mac OS | | | | 2nd | 5th | 6th | 3rd | |
| Linux | | | | 2nd | | | 1st | |
| Android | | | | 3rd | 6th | 2nd | 4th | |
| iOS | | | | 2nd | 6th | 4th | 3rd | |

| | Opera | Safari | Samsung Browser | Sogo | UC Browser | Yandex |
|---|---|---|---|---|---|---|
| Windows | 4th | | | 5th | | 2nd |
| Mac OS | 4th | 4th | | | | 1st |
| Linux | 2nd | | | | | |
| Android | 2nd | | 5th | | 7th | 1st |
| iOS | 5th | 2nd | | | | 1st |
| | | | | | | |

*Yandex and Chrome score highly across all platforms.

**UA-Readiness of Browsers:** Desktop vs. Mobile Device

| | Passed all tests | Failed tests |
|---|---|---|
| Windows | 86 | 66 |
| Mac OS | 42 | 91 |
| Linux | 9 | 48 |
| Android | 48 | 104 |
| IOS | 31 | 102 |

*More Universal Acceptance difficulties with mobile device environments.

*Most failures were due to the following tests

*Confirm that the URL display is in the correct format as added, and

*Confirm that the URL is displayed correctly in the bar.

Top Web-Browser market share in India[5]

| Chrome | Firefox | Edge | Safari | Opera | UC Browser | IE | Edge Legacy | Chromium | Mozilla | Other |
|---|---|---|---|---|---|---|---|---|---|---|
| 86.36 | 5.93 | 3.55 | 1.78 | 1.45 | 0.34 | 0.29 | 0.19 | 0.05 | 0.02 | 0.03 |

Other browsers used in India:

| | |
|---|---|
| JioPages | https://www.jio.com/en-in/apps/jio-pages |
| Epic Privacy Browser | https://www.epicbrowser.com/ |

5 https://gs.statcounter.com/browser-market-share/all/india

## Top Web-Browser market share in India[6]

### Recommendations:

UA readiness of browsers (both web browsers and mobile browsers) essentially necessitates the browser to be able to accept and recognise IDNs and navigate to the appropriate website. From the user's perspective, the URLs should be displayed in unicode for understandability.

### Browser Input

- Ability to accept typing in local language in the address bar.
- Accept IDNs and unicode encoded URLs in address bar
- Recognise the IDNs and validate the domain names
- Load the correct page for the URL and display the page Title in the browser tab
- Display the URLs in unicode in address bar for readability eg.
- Should not display the URL as puny-code which will be confusing for the user e.g. https://xn—h2brbk1b0bdb.xn—h2brj9c/

### Browser accessibility and usability

- Support for bookmarking these URLs and bookmarked URLs should be displayed properly in unicode (along with the Title where appropriate)
- Creating bookmark folder names in local languages
- Support for URLs in browsing history, and other storage
- Popup blocker / exception list should also support IDNs
- Browsers come with a default search engine. This should recognise IDNs in the search term

### Rendering multilingual websites

- HTML the primary language in websites, browsers render the html files. They should support the Unicode characters in the page content and display them uniformly.
- Browser should support Unicode in Links, URLs, anchors used for navigation and should handle all IDNs and Unicode URLs in a standard manner
- Support for IDNs and IEA in developer-tools / save-download page / view source etc.

### Mobile browser considerations

- Many mobiles come with pre-installed browsers / default browsers; these should be UA ready
- Mobiles to support typing in local language (on-screen keyboards, voice input, etc)

6 https://gs.statcounter.com/browser-market-share/all/india

- Some fonts and scripts may not be rendered well in the mobile screens / low resolution, which needs alternate way of handling in mobile browsers

**Final Recommendations**

- Testing current level of support in top browsers in the Indian context i.e., Browsers that are popular in India and their support for the scheduled Indian languages in the following:
  - » Mobile browsers
  - » Web browsers
  - » Defining compliance standards and testing framework based on the above discussed inputs
  - » Default browsers in laptops and mobiles to be encouraged to have UA readiness

## F. WEBSITES

**Challenges:**

- The non English internet today has very little engagement in comparison to the number of literate people having access and using apps like whatsapp and youtube. Discovery of content is key to driving consumption of whatever content that is available.
- Consumption fuels content creation
  - » Discovery is poor because
    - — Content on the internet is noisy (a few screenshots or data needed).
    - — Most IL content is in PDF forms and PDFs are non searchable and non quotable (pointers to popular sources)
    - — Display is ambiguous and inconsistent (same pages in different browsers/ platforms vary in display)
    - — Browsers architecture provides languages support only for those languages that are released with it. External enablement/addition isn't available. This restricts potential growth entirely.
- Engagement and content in nearly all UGC platforms have increasing amount of non English content in English script.
- Spellings are non-standard and widely varied. That makes this content unusable for search or analytics.
- ML for Indian languages will depend largely on user data. All user data based ML and analysis will be inefficient at best.
- Ambiguous and inconsistent character set and display open serious security loopholes.

**Recommendations:**

- Implementing unambiguous script and language display support within the browsers
- Implement a significantly simplified and standardised font format so that more options can be available
- Standardised keyboard aligned with language as taught in schools “only”.
- Open architecture of browsers to add support and tools for languages for display and editing (spellcheck and grammar check)
- Uniform display experience necessitates use of embedded fonts.
- Usage of modern CSS3 attributes to conditionally adjust and render text elements for each locale
- Use of UTF-8 encoding for charset until better standards are available. Lang attributes are critical (especially to help SEO, Analytics etc.)
- RTL languages (Urdu) suffer. Use CSS layout properties to allow the same.
- Avoid concatenating Text templates; For example :
  - » “To access the ” + {section_name} + “ section, click here”
  - » Rather use built-in templatization :
  - » To access the ${section_name} section, click here
  - » So that the dynamic section can be moved around during translations
- Avoid using whitespace characters for text-alignment and line-breaks, instead use text-align properties and line break tags.
- Avoid using block level tags like “div” inside a non block level tag like “span” that may contain text
- Allow text input for different locales

## G. APPS/SOCIAL MEDIA APPS

**Readiness State / Challenges:**

- Social Media currently does not accept IDN / EAI the way they should.
- Only KOO accepts EAI as an email address to sign up and in profile.
- Govt. Apps do not accept EAI for signup.
- Auto hyperlink IDN in messages. (visiting website)
- Air tickets / Aadhar / Mail OTP do not work on EAI.
- Auto MAILTO EAI in messages. (launching mail app)
- Mail app in android and ios is EAI ready (L2). ( configurable )
- Google / Apple do not allow us to sign upon the phone with EAI.

**Recommendations:**

- All govt apps must accept EAI as a signup / newsletter.
- IDN in content must be dealt like a valid domain name.
- Social media companies must allow signup with EAI.

## H. INDUSTRY BODIES IN INDIA:

Association of System Integrators & Resellers in Technology(ASIRT)
Chairman – Jiten Mehta
Tel : +91 022-61542333/22-66122000
Website: www.asirt.in

Association of Unified Telecom Service Providers of India
President – Amit Mathur
Tel: 011–23358585
Fax: 011–23327397
Website: www.auspi.in

The Associated Chambers of Commerce and Industry of India (ASSOCHAM)
President – Niranjan Hiranandani
Tel: 011–46550555
Fax: 011- 23017008
Website: www.assocham.org

Association Of MSMEs in IT ( AIM-IT)
Chairman – Vinit Goenka
Tel : 011-40574900
Website: www.aimitindia.com

Automotive Component Manufacturers Association Of India (ACMA)
Director General -Vinnie Mehta
Tel: (91 11) 2616 0315
Fax: (91 11) 2616 0317
Website: www.acmainfo.com

BICSI
India District Chair – Ninad Desai
Tel: 91 22 27902441
Website: www.bicsi.org

Communications Multimedia and Infrastructure Association of India
President – Prof. N K Goel
Tel: 011-26266411
Website: www.cmai.asia

Computer Association Of Eastern India
President: Asif Khan
Tel:(91 33) 22813609
Website: www.compassindia.com

Confederation of All India Traders -CAIT
National Secretary-Praveen Khandelwal
Tel:011-45032664
Website: www.cait.in

Communications Multimedia and Infrastructure Association of India
President – Prof. N K Goel
Tel: 011-26266411
Website: www.cmai.asia

Confederation of Indian Industry (CII)
President – Mr. Uday Kotak
Tel: 91 11 45771000 / 24629994-7 *
Fax: 91 11 24626149
Website: www.cii.in

Digital India Foundation (DIF)
Co-Founder – Arvind Gupta
Tel: 011 -0120-4568926
Website: www.digitalindiafoundation.org

Digital Communication (India) Ltd
President- Vipin Tyagi
Tel: 9560192600
E-mail: vipinsemail@yahoo.com

Electronic Component Industries Association (ELCINA)
President: Amrit Manwani
Tel: (91 11) 26928053
Fax: (91 11) 26923440
Website: www.elcina.com

Electronic & Computer Software Export Promotion Council (ESC)
Executive Director – Gurmeet Singh
Tel: (91 11) 2696 5103
Fax: (91 11) 2685 3412
Website: www.escindia.in

Federation of All India Information Technology Associations (FAIITA)
President – Kaushik Pandya
Tel: +91-9845353629
Website: www.faiita.org
Federation of Indian Chambers of Commerce and Industry(FICCI)
Secretary General–Dr. DilipChenoy
Tel: 91-11-23738760-70
Fax: 91-11-23320714, 23721504
Website: www.ficci.com

Fire and Security Association of India
National President –Dominic K.P.
Contact: 022-63742 1214/22785640
Website: www.fsai.in

India Electronics and Semiconductor Association
Chairman –Dr. Satya Gupta
Phone: +91 80 4147 3250/51
Website: www.iesaonline.org

India Mobile Congress
President–Shri. P Ramakrishna
Phone: 011- 23440236
Website:www.indiamobilecongress.com

Internet and Mobile Association of India
Chairman – Amit Agarwal
Website: www.iamai.in

Internet Service Providers Association of India
President–Rajesh Chharia
Tel: 011–26424001
Fax: 011–41608472
Website: www.ispai.in

India Trade Promotion Organisation (ITPO)
Chairman & MD- L C Goyal
Tel: 91-11-23371540
Fax: 91-11- 23371492
Website: www.indiatradefair.com

Infotech Software Dealers Association (ISODA)
Chairman – Amarnath Shetty
Tel.: +91-9821243881
Website: www.isoda.in
Manufacturers' Association for IT (MAIT)
President–Nitin Kunkolienker
Tel: +91-11-2685 5487
Fax: +91-11-2685 1321
Website: www.mait.com

NASSCOM
President – Debjani Ghosh
Tel: +91-120-4990111
Fax: +91-120-4990111
Website: www.nasscom.org

Other Service Providers Association of India (OSPAI)
President- S.V. Ramana
Phone: 0120 4319797
Website: www.ospai.in

ProgressiveChannels Association of Information Technology (PCAIT)
President: Alok Gupta
Phone: +91- 9810198881
Website: www.pcait.in

Trade Association of Information Technology (TAIT)
President- Samir Parekh
Phone: 022-23861101
Website: www.tait.in

# ACKNOWLEDGEMENTS

We would like to express our gratitude and appreciation to all those who have contributed to the initiative and shared the input and feedback. Sincere thanks are expressed to the committee members who dedicated their time and energy to this effort from the beginning, and to all the members of each sub committees who have contributed their insights, inputs, and direction for the respective reports and for specific services in the Roadmap for India.

We are grateful for the feedback and inputs of stakeholder's experts (KOO, BIF, Rediff, Facebook, C-DAC, Paytm) who participated in the various meetings and gave their valuable inputs. We are thankful to co-opted members in various subgroups for their valued contribution.

**Committee Members:**

| | | |
|---|---|---|
| • Shri Anil Kumar Jain (Chair)<br>NIXI | • Dr. Ajay Data<br>Xgen Plus | • Shri Balendu Sharma<br>Microsoft |
| • Ms. Asna Siddiqui<br>NASSCOM | • Ms. Karen Ravindranath<br>ZOHO | • Shri Mahesh D. Kulkarni<br>Formerly, Sr. Director (Corporate R&D) & HoD GIST, CDAC |
| • Shri T. Santhosh<br>MeitY | • Shri AjitBalakrishnan<br>Rediff | • Shri Aman Jain<br>Google |
| • Shri R.S. Mani<br>NIC | • Shri Uday Narayan Singh<br>AMITY University | • Ms. Neha Gupta (Member Convener)<br>C-DAC |

## Sub-Committee Members

| | |
|---|---|
| Hardware | Ms. Neha Gupta (C-DAC), Shri Balendu Sharma (Microsoft) |
| E-MAIL | Shri Mahesh Kulkarni (Ex C-DAC), Shri AjitBalakrishnan (Rediff), Dr. Ajay Data (Xgen Plus) |
| Websites | Ms. Asna Siddiqui (NASSCOM), Shri VivekanandPani (Reverie) |
| Softwares / OS | Shri T. Santhosh (MeitY), Shri Balendu Sharma (Microsoft) |
| Security devices | Dr. N. Subramanian(C-DAC), Shri Vinod Kumar M (C-DAC), Shri Annarao Kulkarni (C-DAC), Dr. R. Balaji (C-DAC), Shri ChandrakantDhutadmal (C-DAC), Shri Arun Kumar Gupta (C-DAC) |
| Browser | Ms. Karen Ravindranath (Zoho), Shri Vainateya Koratkar (C-DAC) |
| Apps / Social Media Apps | Dr. AjayData (Xgen Plus), Shri AjitBalakrishnan (Rediff) |

Designed by YQAP