Academic Curriculum for Universal Acceptance (UA) of Domain Names and Email Addresses

Sarmad Hussain
Senior Director of IDN and UA Programs, ICANN
28 March 2024

Universal Acceptance

Welcome!

* Many people around the world are currently excluded from experiencing the full benefits of the Internet simply because they're unable to use a domain name or email address in their language and script of choice

* Through Universal Acceptance (UA), all those who develop, provide or manage online websites and applications have the opportunity to enable users globally to experience the social and economic power of the Internet

* A strong understanding of UA is the new competitive differentiator every developer should have in their skill set
  * UA is the cornerstone of a more inclusive and multilingual Internet
  * UA is essential for developers who want to be at the forefront of their industry and keep pace with the new, global Internet
  * UA offers a $9.8+ billion opportunity to businesses

Evolution of Domain Name and Email Systems

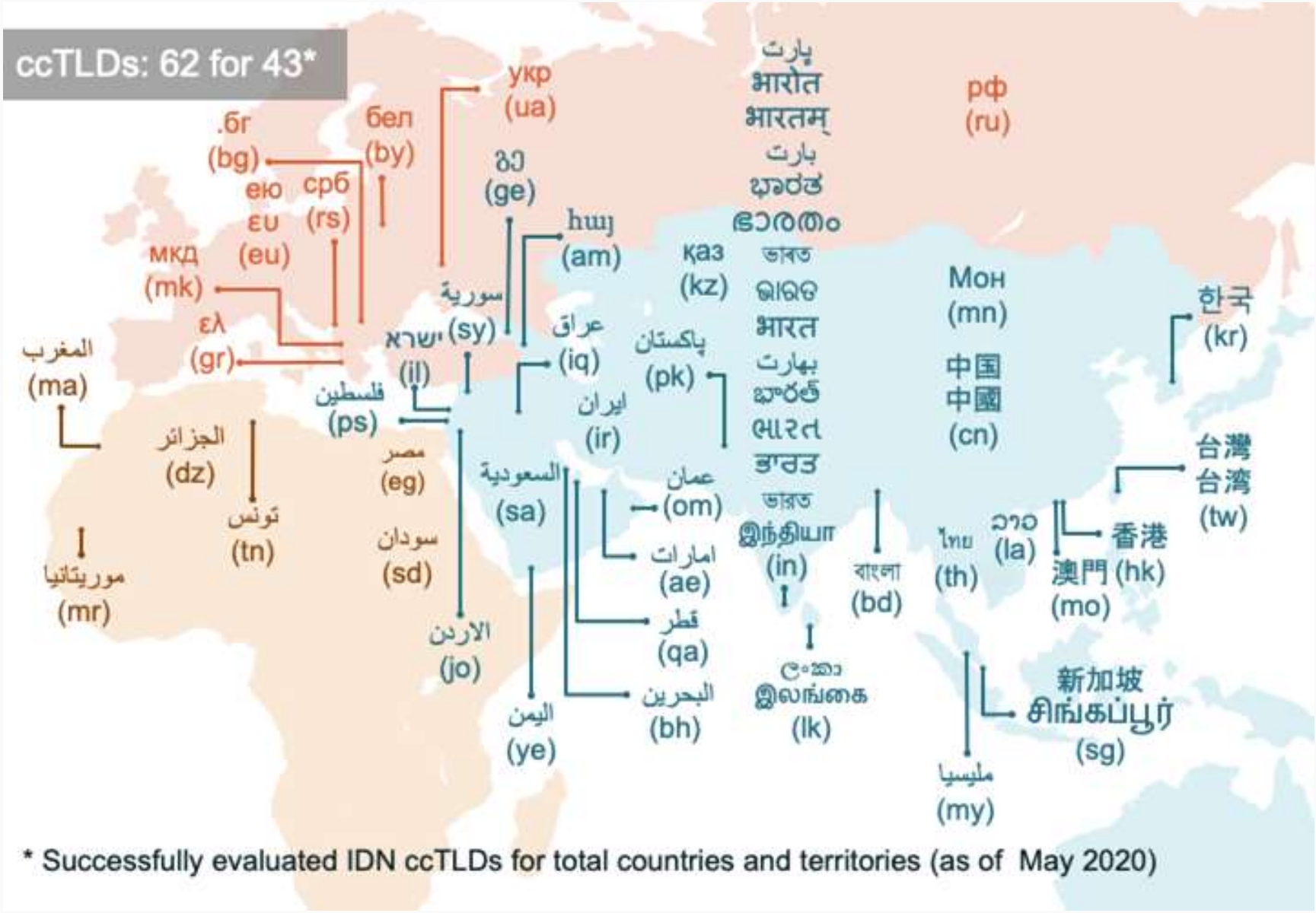
* A domain name is a unique name that forms the basis of the uniform resource locators (URLs) that people use to find resources on the Internet (e.g., web pages, email servers, images, and videos)

* In the past, top-level domains (TLDs) have mostly been two or three letters long formed with Latin characters a-z (e.g., .com, .org). The evolution of the DNS has allowed for newer and longer TLDs
    * There are now around 1,200 new generic gTLDs that represent brands, communities, and geographies (e.g., .photography, .london)

* These new TLDs also include Internationalized Domain Names (IDNs), which allow for domain names in local languages and scripts such as ".例子" and ".مثال" ("example" written in Chinese and Arabic)
    * Other labels within a domain name can also be internationalized, allowing for a domain name to be in a local language and script completely. IDNs are separate from online content, which can also be multilingual

* Email addresses also available in local languages and scripts (EAI)

IDN Country Code Top-Level Domains (ccTLDs)



* Successfully evaluated IDN ccTLDs for total countries and territories (as of May 2020)

IDN Generic TLDs (gTLDs)

90 IDN gTLDs are delegated.

Examples of IDNs

1. համընդհանուր-ընկալում-թեստ.հայ	Armenian
2. ସୁନିଭରସାଲ-ଏକସେପ୍ଟନ୍ସ-ଟେଷ୍ଟ.ଭାରତ	Oriya
3. უნივერსალური-თავსობადობის-ტესტი.გე	Georgian
4. 다국어도메인이용환경테스트.한국	Korean
5. സാർവത്രിക-സ്വീകാര്യതാ-പരിശോധന.ഭാരതം	Malayalam
6. تجربة-القبول-الشامل. موريتانيا	Arabic

Examples of Internationalized Email

1.	էլփոստ-թեստ@համընդհանուր-ընկալում-թեստ.հայ	Armenian
2.	电子邮件测试@**普遍适用**测试.**我**爱你	Chinese
3.	ईमेल-परीक्षण@सार्वभौमिक-स्वीकृति-परीक्षण.संगठन	Devanagari
4.	موريتانيا.الشامل-القبول-تجربة@الكتروني-بريد-تجربة	Arabic
5.	ηλεκτρονικό-μήνυμα-δοκιμή@καθολική-αποδοχή-δοκιμή.ευ	Greek
6.	மின்னஞ்சல்-சோதனை@பொது-ஏற்பு-சோதனை.சிங்கப்பூர்	Tamil

What is Universal Acceptance (UA)?

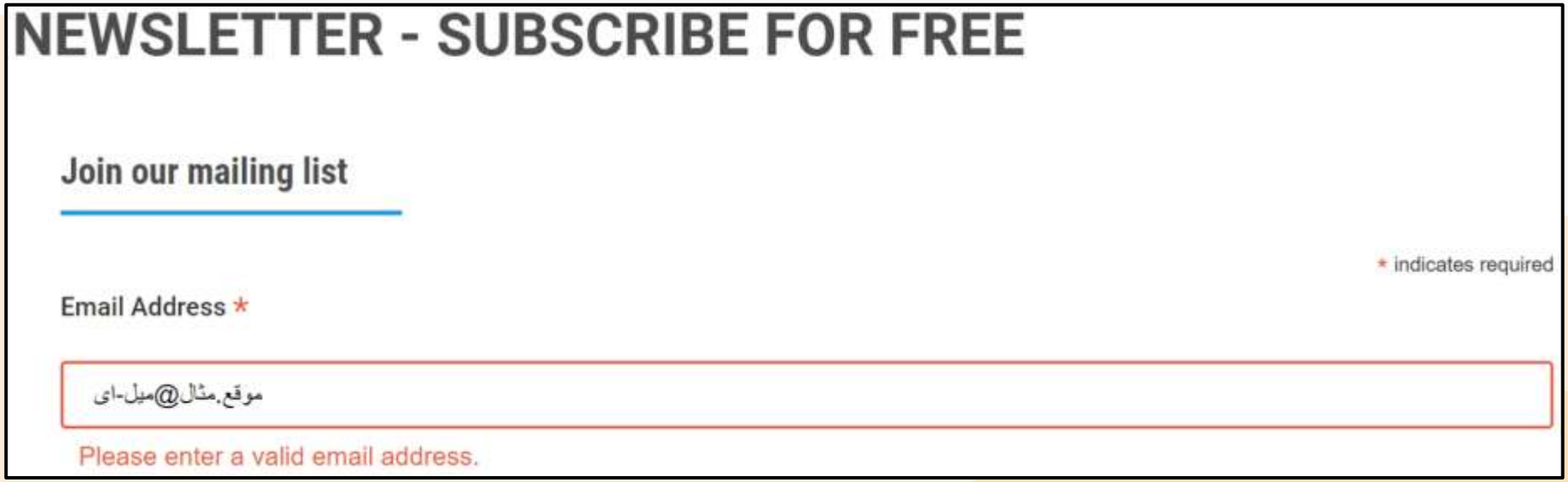
* While the Domain Name System (DNS) has evolved, the checks used by many software applications to validate domain names and email addresses remain outdated

* In addition, not all online portals are primed for the opening of a user account with a related email address, leaving many people unable to navigate the Internet using their language and online identity of choice

* Considered a technical compliance best practice, UA solves these issues by ensuring all valid domain names and email addresses, regardless of script, language or character length, can be used equally by all Internet-enabled applications, devices, and systems

Example of a UA Issue

A valid email is rejected by a form used on a website and is incorrectly displayed from left-to-right instead of right-to-left:



Examples of Categories Affected by UA

* Domain Names that may not work in applications:
  * ASCII: example.sky
  * ASCII: example.engineering
  * Unicode: คน.ไทย
* Internationalized email addresses (EAI) that may not work in applications:
  * Unicode: marc@société.org
  * Unicode: 测试@example.com
  * Unicode: ईमेल@उदाहरण.भारत
  * Unicode: ای-میل@ مثال. موقع(written right-to-left, RTL)

ASCII is based on letters A-Z, a-z, digits 0-9 and hyphen.
Unicode supports characters of global languages and scripts.

UA Goal and Impact

Goal

All valid domain names and email addresses work in all software applications.

Impact

Promote consumer choice, improve competition, and provide broader access to end users.

Making Applications UA-Ready

* Support all valid domain names and email addresses:

* Accept: The user can input characters from their local script into a text field
* Validate: The software accepts the characters and recognizes them as valid
* Process: The system performs operations with the characters
* Store: The database can store the text without breaking or corrupting
* Display: When fetched from the database, the information is correctly shown

Key Objectives for Academic Curricula for UA

* Introduce basic internationalization, including Unicode, IDNs and EAI
* Operationalize UA use cases, i.e., accept, store, process, validate and display
* Learn to use built-in libraries for processing Unicode and IDNs
* Provide technical knowledge and skills on standards or RFCs for IDNs and EAI
* Produce test cases and procedures to verify UA support

Curricular Design Considerations

* Cover theoretical knowledge, standards and best practices
* Develop practical programming competence
* Develop short (micro-learning) modules to integrate key concepts into the existing coursework:
    * Internationalization using Unicode
    * Internationalized Domain Names
    * Email Address Internationalization
* Distribute new concepts across relevant courses
* Introduce minimal new courses, only focused on more advanced topics

Proposed Modules

* Module-1: Unicode Programming Fundamentals
* Module-2: Unicode Advanced Programming
* Module-3: Unicode in Data Structures and Algorithms
* Module-4: Unicode in Database Management Systems
* Module-5: Introducing Internationalized domain names (IDNs)
* Module-6: Programming with Internationalized domain names (IDNs)
* Module-7: Advanced Topics for Internationalized domain names (IDNs)
* Module-8: Email address internationalization (EAI)
* Module-9: Programming for Email address internationalization (EAI)
* Module-10: Processing IDNs and EAI in Mobile Apps
* Module-11: IDN Security
* Module-12: Unicode Support in Operating Systems

* Detailed module descriptions and contents available at: https://community.icann.org/display/TUA/Draft+UA+Curriculum

Module 1: Unicode Programming Fundamentals

Pre-requisites: Data types and data representations, ASCII encoding, ASCII input and output, ASCII files,

1. Introduction to encoding schemes (ASCII, ISO 8859,Unicode)
2. Why do we need Unicode?
3. Character data type for Unicode code points
4. Introduction to Unicode code charts
5. Processing Unicode Strings - creation, input and output, concatenation.
6. Store Unicode data in UTF8 format files

Course for integration: Fundamentals of Programming or Programming I

Module 2: Unicode Advanced Programming

Pre-requisites: Module 1

1. The character-glyph model - difference between processing textual information and displaying text
2. Displaying text - use of text engines, fonts, glyph shapers
3. Normalization of Unicode strings - NFC and NFD
4. Comparing Unicode strings
5. Introducing bidirectional and shaped scripts
    1. Display format
    2. File storage in key-press order
    3. Glyph shapers
6. Interface/Class/Object
    1. Variables using Unicode data
    2. Methods using Unicode data
7. Unicode in Other file formats and their handling - JSON File Unicode handling
8. Using Unicode libraries - ICU and others

Course for integration: Advanced Programming or Object Oriented Programming

Module 3: Unicode in Data Structures and Algorithms

Pre-requisites: Module 2, Arrays, Lists, Sets, Queues, Trees, Dictionaries, etc., Sorting and Searching.

1. Data structures using local language strings in Unicode
2. Unicode collation algorithm for sorting of Unicode strings
3. Searching through Unicode data (Normalization + Collation)

Course for integration: Data Structures and Algorithms

Module 4 : Unicode in Database Management Systems

Pre-requisites: Module 3

1. Indexing on columns with Unicode characters
2. Integrity constraints based on columns with Unicode characters (e.g. for primary key and foreign keys)
3. Queries- regular expression with Unicode characters
4. Storing Unicode data in databases (create, insert and update records)

Course for integration: Database Management Systems

Module 5: Introducing Internationalized Domain Names (IDNs)

Pre-requisites: Module 2, Networks and Protocols, IP, Domain Name System (in ASCII)

1. Introducing the root zone: TLD, gTLD, ccTLD
2. Why do we need IDNs
3. Unicode based domain names: U-label
4. Need for A-labels
5. Punycode algorithm (RFC 3492)
6. IDNA2003 and its limitations
7. IDNA2008
8. Limitations of protocols on IDNs: FTP, HTTP, HTTPS
9. Network troubleshooting commands with U-Labels- dig, traceroute, nslookup, etc.

Course for integration: Data Communication and Computer Networks

Module 6: Programming with Internationalized Domain Names (IDNs)

Pre-requisites: Module 5

1. IDNA2008 compatible libraries and how to use them
   a. Python, Java, and/or other platform(s) of choice
   b. Pitfalls of Using the IDNA2003 Library: Avoiding IDNA2003 libraries
   c. Parsing active IDN TLDs from https://data.iana.org/TLD/tlds-alpha-by-domain.txt

Course for integration: Data Communication and Computer Networks

Module 7: Advanced Topics for Internationalized Domain Names (IDNs)

Pre-requisites: Module 6

1. Limitations of IDNA2008
2. Introducing Label Generation Rules (LGRs)
3. Exploring XML based Label Generation Rules format (RFC7940; RFC 4690, and RFC 3743 as earlier text based formats)
4. Using available LGRs in ICANN to create valid labels
5. Understanding a few LGRs
6. Programming exercise to use LGRs to verify a label and identify its variant labels

Course for integration: Advanced Topics in IDNs and Universal Acceptance

Module 8: Email Address Internationalization (EAI)

Pre-requisites: Module 6

1. Why do we need EAI?
2. Storing EAI mail
3. Email protocol changes for EAI
    1. SMTP - signaling flag (SMTPUTF8)
    2. POP/IMAP
4. EAI and Mail message formats
5. EAI and MIME standard
6. Adoption challenges in a partially-compatible world:
    a. Technical Compatibility Issues -challenges associated with the partial compatibility of EAI across email systems and clients.
        i. Possible technical solution: translation/transliteration feature in receiving email clients, e.g. a reported translation feature of GMail.
    b. Email and Anti-spam measures- anti-spam measures in a partially compatible EAI environment.
7. EAI and cross-cultural email communication skills- Communication Etiquette and Best Practices

Course for integration: Data Communication and Computer Networks

Module 9: Programming for Email address internationalization (EAI)

Pre-requisites: Module 7

1. Handling UTF-8/non-ASCII characters:
   a. Subject line
   b. Header
   c. Email body text
   d. Attachments
2. Libraries or APIs that handle SMTP interactions with EAI support
3. Validating internationalized email addresses using confirmation link workflow
4. Email address parsing for Internationalized emails addresses
5. Considerations for setting up email mailbox names, e.g. as in https://uasg.tech/download/uasg-028-considerations-for-naming-internationalized-email-mailboxes-en/
6. Defining scope of EAI features, e.g. using EAI Self-Certification Guide, e.g. https://uasg.tech/eai-certification/

Course for integration: Advanced Topics in IDNs and Universal Acceptance

Module 10: Processing IDNs and EAI in Mobile Apps

Pre-requisites: Module 2, Module 6, Module 9

1. Unicode Character set support
2. Unicode Text layout and rendering
3. Unicode Support in Mobile Apps Development Frameworks
4. iconv/ICU Converter APIs
5. Processing IDNs on mobile devices - relevant libraries
6. Processing EAI on mobile devices - relevant libraries

Course for integration: Mobile Application Development

Module 11: IDN Security

Pre-requisites: Module 2, Module 6

1. Unicode Security
    1. Unicode Security Mechanisms
    2. Context: there are many other security risks, and DNS, IDN, and Unicode-related risks are small compared to risks from e.g. phishing, user inattention to security, lack of user training about security.
    3. Mitigating EAI addresses blocking.
    4. DNS Security
    5. ASCII Homograph
    6. Phishing, Spamming, …
2. IDN Security
    1. Strings not covered by normalization - managing visually same but distinct strings using LGRs
    2. IDN homographs
    3. LGRs and variants
    4. String similarity

Course for integration: Computer Security

.

Module 12: Unicode Support in Operating Systems

Pre-requisites: Module 2

1. File systems support for Unicode
2. Working with Unicode: Case-sensitive vs case-insensitive vs case-insensitive but case-preserving filename handling
3. Unicode APIs for OS

Course for integration: Operating System

.

Coursework Impacted

Updates in existing courses:

* Fundamentals of Programming or Programming I
* Advanced Programming or Object Oriented Programming
* Data Structures and Algorithms
* Database Management Systems
* Data Communication and Computer Networks
* Mobile Application Development
* Computer Security
* Operating System

Suggested new course:

* Advanced Topics in IDNs and Universal Acceptance

Resources to Support Curricular Update

* Draft module student guide: https://community.icann.org/display/TUA/Draft+UA+Curriculum.
* Draft module instructor guide
* Draft assignments
* References for additional materials

* Faculty training program (on request and availability) by ICANN – please email UAProgram@icann.org

Get Involved with UA!

- For more information on UA, email info@uasg.tech or UAProgram@icann.org
- Access all UASG documents and presentations at: https://uasg.tech
- Access details of ongoing work from ICANN community wiki pages: https://community.icann.org/display/TUA
- Subscribe to the UA discussion list at: https://uasg.tech/subscribe
- Register to participate in UA working groups here
- Follow the UASG on social media and use the hashtag #Internet4All:

  X (formerly Twitter): @UASGTech

  LinkedIn: https://www.linkedin.com/company/uasgtech/

  Facebook: https://www.facebook.com/uasgtech/

Universal Acceptance Steering Group info & recent developments: www.uasg.tech