



jPRS Japan Registry Service

## Why IDN?

- Increases of the Internet users who are not familiar with English.
  - Increasing demand for using their own language to access the Internet.
  - Easy to memorize, type in, etc.
- Drastic changes of usage of domain name.
  - Domain name is now used as not only host name but also signboard.

Copyright © 2004 JPRS 7 jP

jPRS Japan Registry Service

## A series of RFCs

- RFC is a document that defines protocols used on the Internet.
  - Standardization organization is IETF.
  - IDN WG did the work.
- IETF published a series of RFCs as IDN standards on March, 2003.
  - RFC3490 (IDNA)
  - RFC3491 (NAMEPREP)
  - RFC3492 (Punycode)

Copyright © 2004 JPRS 10 jP

jPRS Japan Registry Service

## Drawback of IDN

- Loses global acceptability at end-user interface.
  - Hard to type in or display NON-ASCII characters without appropriate I/O devices and / or softwares.
- Cause impact to the operation.
  - Requires software update and / or additional processing.
  - Deployment issue.

Copyright © 2004 JPRS 8 jP

jPRS Japan Registry Service

## Scope and priority of IDN standardization

- Standard track protocol.
  - Not to divide the global connectivity and communication of the Internet.
- Backward compatibility.
  - Compatibility with current DNS and application protocols to work with current Internet infrastructure.
- No localization.
  - Independent from certain regions, countries and / or languages.
  - Refer to existing universal standards.
  - Common framework essential to internationalization.

Copyright © 2004 JPRS 11 jP

jPRS Japan Registry Service

## IDN Standards

Copyright © 2004 JPRS jP

jPRS Japan Registry Service

## IDNA

(Internationalizing Domain Names In Applications)  
RFC3490

- An architecture denotes how to process IDN.
  - Use Unicode which is upper compatible with ASCII as a character codeset.
  - Normalize internal representation of characters which has multiple code points such as upper/lower, full-width/half-width and composing characters, into a single representation to perform matching correctly.
  - Represent NON-ASCII characters which inputted or displayed at user interface as an ASCII Compatible Encoding (ACE) string on the Network.
  - Those processes be performed in application software.

Copyright © 2004 JPRS 12 jP

Japan Registry Service

## Important point of IDNA

- Representation at the user interface layer and the network layer is different.
  - Though the same for ASCII domain names.
  - Internationalized form on display, ASCII form on the network.
  - To keep backward compatibility.
    - Comply with RFC2825 and RFC2826
- Application solution.
  - Least impact to the Internet infrastructure.
- Following domain names are equivalent (1 to 1 mapping) in meaning of IDN.
 

日本語ドメイン名 ↔ xn--eckwd4c7c777u7mw04bc84j

Copyright © 2004 JPRS 13

Japan Registry Service

## Important point of NAMEPREP

- Normalize representation of Internationalized domain name string to perform matching correctly.
  - 'a' vs 'A'
  - 'u'+'' vs 'ü'
  - 'ア' vs 'ア'

Copyright © 2004 JPRS 16

Japan Registry Service

## Image of the IDNA

Copyright © 2004 JPRS 14

Japan Registry Service

## Processes in NAMEPREP

- map
  - Case folding of upper/lower characters (UAX#21)
- normalize
  - Normalize representation of string (UAX#15's NFKC)
- prohibit
  - Check out inappropriate character as domain name.
- Bidi check
  - Check out inappropriate mixing of bidirectional characters.

Copyright © 2004 JPRS 17

Japan Registry Service

## NAMEPREP

(Stringprep Profile for Internationalized Host Names)  
RFC3491

- Profile for STRINGPREP (Preparation of Internationalized Strings)
  - RFC3454
- Some scripts such as alphabet have multiple representation for a character.
  - Domain name is case insensitive.
- Normalization process to unify representation of strings that is the same in meaning or displaying into a single representation.
  - Case (upper / lower)
  - Compatible character (full / half width)
  - Composing character (accent / combination mark)

Copyright © 2004 JPRS 15

Japan Registry Service

## ACE

(ASCII Compatible Encoding)

- Represent NON-ASCII characters by ASCII characters.
  - Easy to apply current DNS.
  - Least impact to current applications.
- Decreases maximum characters in each label.
  - Penalty of using only 5-6bit to represent 8bit data.
  - Requires some sort of compression algorithm.

Copyright © 2004 JPRS 18

Japan Registry Service

## ACE prefix

- An explicit ACE-identifier.
  - To recognize the string is ACE.
    - Important at reverse conversion.
  - 'XN--' was selected as ACE Prefix.
    - IANA did selection on 14 Feb 2003.
      - <http://www1.ietf.org/mail-archive/ietf-announce/Current/msg22619.html>
  - Denoted in RFC3490 (IDNA).

Jp

Copyright © 2004 JPRS 19

Japan Registry Service

## Compression process of Punycode (simplified for understanding)

- Use "文字列例" as an example.
- Compression.
  - 0:U+6587 1:U+5B57 2:U+5217(3)U+4F8B
  - 3:0x4F8B 2:0x28C 1:0x940 0:0xA30
  - 81455 2610 9473 10432
  - (1FTM) (20H) (7A5) (81R)
  - BASE36 (Sample)

placement (0 origin)

Sort and take diff  
 To integer (diff\*chars+ position)

Jp

Copyright © 2004 JPRS 22

Japan Registry Service

## Criteria of ACE selection

- Simple algorithm.
  - For ease implementation.
  - Interoperability.
- Effective compression mechanism.
  - To accommodate characters as much as possible.
- Bilateral corresponding between encoding and decoding.
  - To avoid existence of alternative encoded representation for one IDN.
  - Security consideration.

Jp

Copyright © 2004 JPRS 20

Japan Registry Service

## Generalized variable-length integers of Punycode

- 12345 in decimal is represented as  $1*10^4+2*10^3+3*10^2+4*10^1+5*10^0$
- Digits in all place are 0-9, so components in sequential 12345 cannot distinguish 123 and 45 or 1234 and 5.
- Furthermore, 012345 and 12345 are the same value with different representation.
- GVLI (Generalized variable-length integers) is an idea to solve this problem.
- Defines threshold for each place, and recognize a number below the threshold is delimiter.
- Threshold is an appropriate number smaller than base number.

Jp

Copyright © 2004 JPRS 23

Japan Registry Service

## Punycode

A Bootstring encoding of Unicode for IDNA  
RFC3492

- ACE algorithm for IDN.
- Two key concepts.
  - Compression.
  - Encoding/Decoding.
- Compression algorithm.
  - Extract characters by ascending order of codepoint.
  - Encode difference of codepoint from previously processed character's and the position into an integer.
- ASCII conversion algorithm.
  - Introduced new concept named 'Generalized variable-length integers'.
  - BASE36 (A-Z, 0-9).

Jp

Copyright © 2004 JPRS 21

Japan Registry Service

## Encoding process of Punycode (simplified for understanding)

- Assign A-Z0-9 to GVLI.
  - Assume 36 for base, 10, 18, 25, 25 for thresholds.

- 81455 2610 9473 10432
- NIUD  $23*1+18*26^{(1*(36-10))}+30*468^{(2*(36-18))}+13*5148^{(3*(36-25))}$
- AS4  $10*1+28*26+4*468$
- ZLJ  $35*1+21*26+19*468$
- WML  $32*1+22*26+21*468$

- 「文字列例」=>"NIUDAS4ZLJWML" (Pseudo Punycode)
- "1FTM20H7A581R" (BASE36)
- "FSQW5D78MBSK" (Real Punycode)

Jp

Copyright © 2004 JPRS 24

Japan Registry Service

## Standardization of IDN is just the start point of utilization

- End users uses IDN through application softwares.
  - Web, Mail, etc.
- IDNA requires application's correspondence.
- Must define how to deal IDNs in application protocols.

Standardization of IDN does not mean ready to use. Just a start point for applications incorporating new features.

Copyright © 2004 JPRS 25

Japan Registry Service

## IDN Technical Operation

Copyright © 2004 JPRS

Japan Registry Service

## Example of Web site browsing

GET http://xn--eckwd4c7c777u.jp/ HTTP/1.1  
Host: xn--eckwd4c7c777u.jp  
Referer: http://xn--eckwd4c7c777u.jp/

XN--ECKWD4C7C777U.JP

Web server's IP address

User

Web

Contents

http://ドメイン名.JP/

Copyright © 2004 JPRS 26

Japan Registry Service

## Overview

- No need to update name servers and / or resolvers.
- Local Encoding  $\leftrightarrow$  Punycode converter is required.
  - Such as idnconv in idnkit
- General procedure:
  - Edit configuration / zone file(s) using editors.
  - Convert encoding using converter.
  - Reload configuration / zone file(s) to name server.
  - Check the settings.

Copyright © 2004 JPRS 29

Japan Registry Service

## Unresolved Issues

- The same characters, different characters.
  - Similar looking but different one.
    - A and □ and □ and □
  - Different looking but the same one (SC/TC).
    - 机 and 機, 叶 and 葉, 国 and 國
  - Language dependent issue.
- IDN-compliant application.
  - Vendors never be interested in without demand.
  - Recently, many of brand-new browsers are IDN-aware.
- Another identifiers.
  - Mail address, URI, and so on.
  - draft-duerst-iri-10.txt

Copyright © 2004 JPRS 27

Japan Registry Service

## DNS Settings

(Example of named.conf of BIND)

```

named.conf
zone "xn--eckwd4c7c777u.jp" {
  type master;
  file "JDN-test.zone";
};

named.conf
zone "xn--eckwd4c7c777u7mwo4bov4jioau09j.jp" {
  type master;
  file "JDN-test.zone";
};

```

Copyright © 2004 JPRS 30

Japan Registry Service

## DNS Settings (cont.)

(Example of zone file of BIND)

```

JDN-test.zone
@ IN SOA ...
  閲覧 IN A 10.20.30.40
  開け CNAME 閲覧
:
:

JDN-test.zone
@ IN SOA ...
xn--gx2asoz3zgme IN A 10.20.30.40
xn--08j2i6ia IN CNAME xn--gx2asoz3zgme
:

```

Copyright © 2004 JPRS 31 Jp

Japan Registry Service

## How to operate other servers

- IDNs can be used with existing servers.
  - Web, mail, etc.
- Principle
  - IDNs MUST be represented with ACE format (Punycode) in configuration file.
    - Ex. Virtual host name in httpd.conf

Copyright © 2004 JPRS 34 Jp

Japan Registry Service

## DNS settings (cont.)

(Example of BIND)

- Using Makefile makes maintenance ease.
- Divide editing file and configuration file, then generate latter one from former one.

```

.SUFFIXES: .conf .conf-j .zone .zone-j
.conf-j.conf:
    idnconv $< > $@
.zone-j.zone:
    idnconv $< > $@
all:
    named.conf JDN-test.zone

```

Copyright © 2004 JPRS 32 Jp

Japan Registry Service

## IDN Compliant clients & implementations

- Netscape7.1 / Mozilla-1.4 or higher
  - <http://channels.netscape.com/ns/browsers/download.jsp>
  - <http://www.mozilla.org/products/mozilla1.x/>
- Mozilla Firefox-0.8 or higher
  - <http://www.mozilla.org/products/firefox/>
- Opera7.2 or higher
  - <http://www.opera.com/>
- i-Nav (A plug-in) for Internet Explorer 5 or higher
  - <http://www.idnnow.com/>
- JPNIC idnkit
  - <http://www.nic.ad.jp/ja/idn/mdnkit/download/>
- GNU libidn
  - <http://www.gnu.org/software/libidn/>
- VeriSign SDK
  - [http://www.verisign.com/nds/naming/idn/sdk\\_form.html](http://www.verisign.com/nds/naming/idn/sdk_form.html)

Copyright © 2004 JPRS 35 Jp

Japan Registry Service

## How to check DNS settings

- Using DNS lookup tool and code converter:
  - Using shell alias or function is convenient
    - csh: alias idig 'dig `echo ¥!\*` | idnconv`'
    - bsh: idig () { dig `echo "\$@" | idnconv` }
  - Output in ACE format
    - idig 日本語ドメイン名試験.jp
  - Output in IDN format
    - idig 日本語ドメイン名試験.jp | idnconv -r

Copyright © 2004 JPRS 33 Jp

Japan Registry Service

## IDN Service Implementation

Copyright © 2004 JPRS 36 Jp

jPRS Japan Registry Service

## Association with Language

- IDN itself has no language information.
- At registration service layer, IDNs MUST be associated with certain Language(s).
  - To reduce possible confusion due to [unresolved issues](#).
  - Define a Language table that includes:
    - Language Name
    - List of acceptable codepoints
    - Variants (if any)

Copyright © 2004 JPRS 37 jPp

jPRS Japan Registry Service

## Case study of JP

- JPRS defined “Japanese Domain Name” and started its registration on Feb 2001.
- The first “Language Domain Name” registration service in the world.
- More than 40,000 registration.

Copyright © 2004 JPRS 40 jPp

jPRS Japan Registry Service

## IDN Registration Guideline

- An algorithm to be applied at an IDN registration.
  - Administration guideline for zone managers.
- Originally developed by Joint Engineering Team of CN, JP, KR, TW and a few Experts.
  - JET Guidelines for IDN Registration and Administration for CJK (RFC3743)
- And other related guidelines.
  - draft-klensin-reg-guidelines-05.txt

Copyright © 2004 JPRS 38 jPp

jPRS Japan Registry Service

## Japanese Domain Name’s registration technical rules

- 1st edition published on 06 Nov 2000.
  - Prior work to any IDN registration guidelines.
- Regulates:
  - Definition of Japanese Domain Name.
  - Maximum number of characters in a label.
  - Normalization rule at the registration.
  - Encoding for the resolution.
  - Reserved names and strings including ACE prefixes.
  - Acceptable characters’ table for registration.
  - Referral standards.

Copyright © 2004 JPRS 41 jPp

jPRS Japan Registry Service

## ICANN Guideline

- Conditions that ICANN permits IDN registration.
  - Targets are TLD registries with ICANN contract.
  - Following IDN Registration Guidelines.
  - Prohibits using marks.
  - Version 1.0 was published on 20 June.
    - <http://www.icann.org/general/idn-guidelines-20jun03.htm>

Copyright © 2004 JPRS 39 jPp

jPRS Japan Registry Service

## ja-JP

- Derived work of the JDN technical rules.
  - According to IDN-Admin format.
- Registered as IDN Language Table at IANA.
  - <http://www.iana.org/assignments/idn/jp-japanese.html>

Copyright © 2004 JPRS 42 jPp

jPRS Japan Registry Service

## Reserved words

- Over 8000 of:
  - Government-and-municipal-offices names.
  - Names of Internet-related organization.
  - Suffixes indicating schools.
  - Prefecture names, big city names.
  - Normal words.
  - ... and so on

Copyright © 2004 JPRS 43 jP

jPRS Japan Registry Service

## References

- Unicode Consortium
  - <http://www.unicode.org/>
- Terminology Used in Internationalization in the IETF (rfc3536)
  - <http://www.ietf.org/rfc/rfc3536.txt>
- IDN and related standards
  - <http://www.ietf.org/rfc/rfc3454.txt>
  - <http://www.ietf.org/rfc/rfc3490.txt>
  - <http://www.ietf.org/rfc/rfc3491.txt>
  - <http://www.ietf.org/rfc/rfc3492.txt>
- IANA IDN Language Table Registry
  - <http://www.iana.org/assignments/idn/>
- JET Guideline for IDN
  - <http://www.ietf.org/rfc/rfc3743.txt>

Copyright © 2004 JPRS 46 jP

jPRS Japan Registry Service

## Duplication exclusion in sunrise and concurrent registration periods

- Sunrise period was to protect IPR.
  - One month period started at 22 Feb 2001.
- Concurrent registration period was to avoid rushing.
  - Three weeks period started at 2 Apr 2001.
- Lottery system for duplication exclusion.

Copyright © 2004 JPRS 44 jP

jPRS Japan Registry Service

## Live IDN examples

- Vietnamese
  - <http://tênmiên tiếngviệ t.vn/>
- Thai
  - <http://ทีเอชที. th/>
- Taiwanese (Traditional Chinese)
  - <http://台網中心. tw/>
- Chinese (Simplified Chinese)
  - <http://中国互联网络信息中心. cn/>
- Korean (Hangeul)
  - <http://. kr/>
- Japanese (Kanji)
  - <http://日本語. jp/>

Copyright © 2004 JPRS 47 jP

jPRS Japan Registry Service

## Dispute Resolution Policy

- Preparation of DRP applicable to JDN.
  - JP DRP is defined by JPNIC.

Copyright © 2004 JPRS 45 jP