IDN Root Zone LGR Generation Panels Workshop

#ICANN50
Agenda

• Welcome – 5 min
• Community updates:
  o Chinese Prospective Generation Panel - 15 min
  o Japanese Prospective Generation Panel - 15 min
  o Korean Prospective Generation Panel - 15 min
  o CJK Coordination Report – 15 min
  o Neo-Brahmi Prospective Generation Panel - 15 min
  o Arabic Generation Panel – 15 min
• LGR project mailing lists – 5 min
• Discussions and Q&A- 20 min
Community Work

Chinese Prospective Generation Panel

#ICANN50
Updates on Chinese Generation Panel

Wang Wei
ICANN 50th, London
25th June 2014
## Evolution of Chinese Character

<table>
<thead>
<tr>
<th>1600 ~ 1100 BC</th>
<th>221 BC</th>
<th>200 AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>oracle bone</td>
<td>greater seal</td>
<td>lesser seal</td>
</tr>
<tr>
<td>jiaguwen</td>
<td>dazhuan</td>
<td>xiaozhuan</td>
</tr>
<tr>
<td>clerkly script</td>
<td>standard script</td>
<td>running script</td>
</tr>
<tr>
<td>lishu</td>
<td>kaishu</td>
<td>xingshu</td>
</tr>
<tr>
<td>cursive script</td>
<td>caoshu</td>
<td>modern simplified</td>
</tr>
<tr>
<td>jiantizi</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>rén (*nin) human</td>
</tr>
<tr>
<td>nǚ (*nra?) woman</td>
</tr>
<tr>
<td>ěr (*nha?) ear</td>
</tr>
<tr>
<td>mǎ (*mrā?) horse</td>
</tr>
<tr>
<td>yú (*nha) fish</td>
</tr>
<tr>
<td>shān (*srān) mountain</td>
</tr>
<tr>
<td>rì (*ni) sun</td>
</tr>
<tr>
<td>yuè (*not) moon</td>
</tr>
<tr>
<td>yǔ (*wha?) rain</td>
</tr>
<tr>
<td>yún (*wan) cloud</td>
</tr>
</tbody>
</table>

Totally 100,000 Chinese Characters
Countries with Significant User Communities for Chinese Script

- hanzi in Chinese
- Traditional Chinese
  - Taiwan
  - Macau
  - Hong Kong
- Simplified Chinese
  - Mainland China
  - Singapore
- TC & SC
  - Malaysia
- kanji in Japanese
- hanja in Korean

Chinese script is the logogram used in the writing of Chinese and some other Asian languages.
Target Script: Hani

ISO 15924 code: Hani
ISO 15924 no.: 500

English Name: Han (Hanzi, Kanji, Hanja)
Target Script and Scope (2/3)

CJK and Han Unification:

*CJK Unified Ideographs: 4E00–9FFF*

*CJK Unified Ideographs Extension A: 3400–4DBF*

*CJK Unified Ideographs Extension B: 20000-2A6DF*

*CJK Unified Ideographs Extension C: 2A700-2B73F*

*CJK Unified Ideographs Extension D: 2B740–2B81F*

➢ MSR-1 includes CJK, CJK-A and a minor part of CJK-B.

➢ The complete character set for the Chinese script in the IANA IDN tables from .CN, .TW, and .JP all fall in the ranges of CJK and CJK-A.
Best Practice: **CDNC Character Set and Guide Rules**

- In 2004, according to RFC 3743 and RFC 4713, CDNC submitted to IANA a unified Chinese Character Set (19320 characters).
- CDNC put forwards its guild rule for SLD registration, applied on CN, TW and etc.
  - The IDL and its Variant Labels SHOULD belong to the same registrant.
  - The Simplified Chinese form and Traditional Chinese form of the applied-for IDL SHOULD be resolvable simultaneously or non-resolvable at all.
- In 2013, among 76 Chinese new gTLD applications, 32 applications are announced to abide by CDNC Character Set and Guide Rules.

Owing to the above factors, CGP would like to propose that CDNC Character Set be treated as the **initial character set** of CGP.
Initial Composition of the Panel

* The CGP gathers 13 experts from a variety of backgrounds (bringing varied linguistic and technical perspectives) from 7 different countries/regions, including:
  * national and regional policy makers.
  * members from technical community directly working with the DNS (e.g. registries and registrars), security, academia (technical and linguistic).
  * members of community based organizations.
  * members with experience of local language studying.
<table>
<thead>
<tr>
<th>Name</th>
<th>Designation</th>
<th>Organization</th>
<th>Country/Region</th>
<th>Language Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chris DILLON</td>
<td>Mr</td>
<td>University College London</td>
<td>UK</td>
<td>Chinese/Japanese/Korean</td>
</tr>
<tr>
<td>Chao QI</td>
<td>Mr</td>
<td>CNNIC</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Di MA</td>
<td>Dr</td>
<td>ZDNS</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Guoying LI</td>
<td>Dr</td>
<td>Beijing Normal University</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Jiagui XIE</td>
<td>Mr.</td>
<td>CONAC</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Joe ZHANG</td>
<td>Dr.</td>
<td>Beijing UniHan Digital Technology Co., Ltd.</td>
<td>China</td>
<td>Chinese Japanese Korean</td>
</tr>
<tr>
<td>Jonathan SHEA</td>
<td>Mr</td>
<td>HKIRC</td>
<td>Hong Kong</td>
<td>Chinese</td>
</tr>
<tr>
<td>Joseph YEE</td>
<td>Mr</td>
<td>Afilias</td>
<td>Canada</td>
<td>Chinese, (Familiar with Japanese)</td>
</tr>
<tr>
<td>Kenny HUANG</td>
<td>Dr</td>
<td>TWNIC</td>
<td>Taiwan</td>
<td>Chinese</td>
</tr>
<tr>
<td>Linlin ZHOU</td>
<td>Ms</td>
<td>CNNIC</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Nai-Wen HSU</td>
<td>Dr.</td>
<td>TWNIC</td>
<td>Taiwan</td>
<td>Chinese</td>
</tr>
<tr>
<td>Ryan TAN</td>
<td>Mr.</td>
<td>SGNIC</td>
<td>Singapore</td>
<td>Chinese</td>
</tr>
<tr>
<td>Wei WANG</td>
<td>Dr.</td>
<td>CNIC</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Xiaodong LEE</td>
<td>Dr.</td>
<td>CNNIC</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Zheng WANG</td>
<td>Dr.</td>
<td>CONAC</td>
<td>China</td>
<td>Chinese</td>
</tr>
<tr>
<td>Zhiwei YAN</td>
<td>Dr.</td>
<td>CNNIC</td>
<td>China</td>
<td>Chinese, (Familiar with Japanese)</td>
</tr>
</tbody>
</table>
Relationship with Past Work

* From JET to CDNC to VIP to CGP
  * JET: April, 2004, produced RFC 3743
    * It includes various concepts for variant handling, such as bundling, atomic IDL Package, and reserved variants.
    * It also defines a standard table as well as an algorithm to generate the preferred variant and reserved variants.
  * CDNC: TC-SC Equivalence
  * VIP: Chinese Case Study Report
  * CGP: variants and label rules for Root Zone
Work Plan

* Four stages (< ICANN51 L.A. in October 2014):
  * 1. Finalization of Code Points
  * 2. Finalization of Variants
  * 3. Finalization of Whole Label Rules
  * 4. Finalization of LGR Documents for Chinese Script and Submission to ICANN

  * 14th Mar, 1st joint meeting of CGP and CDNC
  * 2nd Apr, proposal submitted to ICANN (comments feedbacked on 3rd May)
  * 13th May, principle of character set submitted to ICANN
  * 29th May, 2nd joint meeting of CGP and CDNC
Challenges

* Character Reduction and Extension
* Coordination between C J K
  * Character, variant, rule
Reduction & Extension

* preliminary question:
  * Will the requirement of generation rules for root zone be different from rules applied in 2nd layers (e.g., RFC3743 for CJK registration)?
  * Will future character extension be allowed after LGR finished the current work?
Some Suggestions

* The obsolete Korean Lidu be removed from the CGP repertoire.
* Remove the Han Radicals /Strokes that are not regarded as independent Hanzi.
* Remove some Kanji only used in J script context.
* Include more local Hongkong character into CGP repertoire.
Coordination
Coordination
Coordination

* Different language panels (C, J and K) have different views about the Chinese characters and variants:
  * Chinese: Equivalent mapping between SC and TC and generic variants. (*reflected in the CDNC character table*)
  * Japanese: It is appropriate to distinguish new and old forms (of Kanji) as different and independent characters instead of handling them as variants.
  * Korean: Its IDN policy does not allow Hanja as reflected by the language table it submitted to IANA (*Hanja are no longer widely used in ROK*)

* Then a coordination scheme between different panels is needed (especially between CGP and JGP).*
<table>
<thead>
<tr>
<th>Code point</th>
<th>Allocatable variant</th>
<th>Blocked variant</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>恋 U+604B</td>
<td>恋 U+6200</td>
<td>-</td>
<td>und-hani-TBD-root</td>
</tr>
<tr>
<td>恋 U+604B</td>
<td>-</td>
<td>恋 U+6200</td>
<td>und-jpan-TBD-root</td>
</tr>
<tr>
<td>恋 U+6200</td>
<td>恋 U+604B</td>
<td>-</td>
<td>und-hani-TBD-root</td>
</tr>
<tr>
<td>恋 U+6200</td>
<td>-</td>
<td>恋 U+604B</td>
<td>und-jpan-TBD-root</td>
</tr>
<tr>
<td>愛 U+611B</td>
<td>爱 U+7231</td>
<td>-</td>
<td>und-hani-TBD-root</td>
</tr>
<tr>
<td>愛 U+611B</td>
<td>-</td>
<td>爱 U+7231</td>
<td>und-jpan-TBD-root</td>
</tr>
<tr>
<td>爱 U+7231</td>
<td>爱 U+611B</td>
<td>-</td>
<td>und-hani-TBD-root</td>
</tr>
</tbody>
</table>
Coordination

* In the last CDNC meeting in Shanghai, a coordination scheme was proposed by the ICANN expert:
  * Each CJK panel creates an LGR and each LGR includes a repertoire and variants.
  * If an LGR includes Han characters, the variant mappings must agree for all three panels.
  * The variant types may be different (blocked or allocatable), the variant types do not have to agree across LGRs.

* C and J (and K) need to work closely on the above principle.
Thanks

wangwei@cnic.cn
About CDNC

- Evaluates all Chinese domain name compliance with international criterion.
- Establish the standards for Chinese domain name and corresponding regulation on Chinese domain name registration.
- Coordinates the operations and management of Chinese domains in various countries or regions,
- Collaborates with all corresponding international organizations on Chinese domain names
More 10 years Experience On Variant Issues

- CNNIC began its testing process for CDN registration
- .CN/.TW CDN Second Level Domain open for registration
- IANA publishes the Chinese Character Table
- Delegation of Synchronized IDN ccTLDs 中国/中國, .台湾/台灣
- CDNC reaches agreement on synthesis character table format for the root
- Approval of RFC 3743
- Approval of RFC 4713
- Publish IDN Variant issue - Chinese Case Study Report
- New gTLDs launches
- .CN/.TW CDN Second Level Domain open for registration
Community Work

Japanese Prospective Generation Panel

#ICANN50
# JGP Schedule (draft)

<table>
<thead>
<tr>
<th>Milestone</th>
<th>5月</th>
<th>6月</th>
<th>7月</th>
<th>8月</th>
<th>9月</th>
<th>10月</th>
<th>11月</th>
<th>12月</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese Generation Panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ICANN50</td>
<td></td>
<td>ICANN51</td>
</tr>
<tr>
<td>Japanese Community</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CJK Joint Panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action</th>
<th>5月</th>
<th>6月</th>
<th>7月</th>
<th>8月</th>
<th>9月</th>
<th>10月</th>
<th>11月</th>
<th>12月</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact to candidates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Mtg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mtg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mtg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mtg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mtg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mtg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly meeting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explain draft Japanese Rule, call for comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirm each CJK Rules, Discussion, Consensus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▲Approval of JGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation of CJK Unified Rule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Community Work
Korean Prospective
Generation Panel

#ICANN50
The Status of KLGP
(Korean Label Generation Panel)


gimgs@pnu.kr, PNU
choimh@kisa.or.kr, KISA
The Status of KLGP
I. The Status of KLGP

- Members of Korean Label Gen. Panel

<table>
<thead>
<tr>
<th>Classification</th>
<th>Name</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair (Linguistic Expert)</td>
<td>Kyongsok Kim</td>
<td></td>
</tr>
<tr>
<td>Policy Expert</td>
<td>Youngeum Lee</td>
<td></td>
</tr>
<tr>
<td>Policy Expert</td>
<td>Dongman Lee</td>
<td></td>
</tr>
<tr>
<td>Linguistic Expert</td>
<td>Kyungran Kang</td>
<td></td>
</tr>
<tr>
<td>Language Expert</td>
<td>Yongju Hwang</td>
<td>New Member</td>
</tr>
<tr>
<td>Community</td>
<td>Eunghwi Jeon</td>
<td>New Member</td>
</tr>
<tr>
<td>Community</td>
<td>Yangwoo Ko</td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>YJ Park</td>
<td></td>
</tr>
<tr>
<td>Registry</td>
<td>MinJung Park</td>
<td></td>
</tr>
<tr>
<td>Registry</td>
<td>Myunghee Choi</td>
<td></td>
</tr>
</tbody>
</table>

Still recruiting members who have expertise in culture.
The Result of KLGP Meeting
II. The result of KLGP Meeting (just once)

1. The Scope of KLGR?

- Hangul Syllables
  - seems to be No Variant issues
  - Range: AC00-D7A3 (11,172 syllables)

- Hanja (Hanzi, CJK Ideographs)?
  - probably need to have a public hearing
II. The result of KLGP Meeting (just once)

2. The Review of MSR-1 (1)

• Hangul: **Hangul syllables**(11,172)

5.14 Korean Jamo and Hangul

Modern Korean is written with 11,172 Hangul syllables, which, in turn are combinations of elements, called Jamo. The MSR contains all modern Hangul but none of the Jamo characters, which are only needed for non-modern Hangul.

The modern Hangul for the MSR are listed in a separate PDF file "Maximal Starting Repertoire – MSR-1: Annotated Repertoire Tables, Hangul".
II. The result of KLGP Meeting (just once)

2. The Review of MSR-1 (2)

- Han Script (CJK Unified Ideographs, CJKU main, CJKU ExtA, and CJKU ExtB; IICORE)

5.13 Han Ideographs

In creating the MSR, the Integration Panel reviewed existing IDN tables for CJK domains and compared them to various subsets, including IICORE, defined in the Unicode Consortium's UniHan database [UAX38]. From this analysis it appears that a superset of certain IDN tables plus the IICORE is most likely to produce a starting set that satisfies the requirement of being larger than the expected final LGR, while at the same time not being overly inclusive.

Chinese Characters (Han ideographs) for the MSR are listed in a separate PDF file “Maximal Starting Repertoire – MSR-1: Annotated Repertoire Tables, Han”. This file uses a different convention for
II. The result of KLGP Meeting (just once)

2. The Review of MSR-1 (3)
   • [Question] How was the list of Han characters in MSR-1 created?
CJK Meeting in Shanghai
III. CJK Meeting In Shanghai

1. The Overview of CJK LGP Meeting

- Meeting : CGP (Chinese Generation Panel)
- Date : 2014. 5. 29.
- Place: Shanghai Branch of Chinese Academy of Science Building No.1, 2nd Floor Meeting Room
- Participants : about 20 people (Prof. Kyongsok Kim from Rep. of Korea attended the meeting)
  - Japan and USA by teleconferencing
III. CJK Meeting In Shanghai

2. The detail of CJK Meeting

• MSR-1 Review

• The Status of Hanja Domain in Korea
  – In Korea, we don’t use Chinese (Han script) under ‘.한국(.xn--3e0b707e, Korean Top-level IDN)’:
    – e.g., “漢字.한국” not allowed currently.

• KLGP will continuously participate in the CJK LGP meeting.
3. Idu (89/91 chars)

- Some Idu chars are included in IICORE.
- For most Idu chars, K (South Korea), KP (North Korea), and G (China) columns are filled in. For some chars, T (19) and H (14), J (5), and V (2) columns are also filled in.
- Need to keep IICORE chars

- Suggest NOT to exclude Idu from MSR-1
IV Issues
IV. Issues

1. KLGR Scope?

• Need more time to consider the KLGR scope.
  – KLGR: Hangul only?
    & participate in CJK LGP
  – KLGR: (Hangul + Hanja)?
    & participate in CJK LGP
IV. Issues

2. Multi-script IDN TLD ?

• **Multi-script IDN TLD allowed?**
  – E.g., “.한국”, “.한국”, etc (Hangul + Hanja)

• **[Ques] Can KLGR limit the Hangul TLD ?**
  – Allowed chars in Hangul TLD must be the same as those in the second level domain (2LD) under “.kr”/”.한국”?
  – 2LD under “.kr”/”.한국”: composed of complete Hangul syllables (11,172), English letters (a~z), digits (0~9), and hyphen (-).
  – No other chars allowed in TLD having Hangul ?
IV. Issues

3. Members of KLGP

– How to make North Koreans involved in KLGP?

– Making North Korean refugees in South Korea involved in KLGP?
Future Plan
V. Future Plan

• To hold KLGP meetings
• To hold a public hearing
• To draft KLGR
Thanks.
Community Work

CJK Coordination Report

#ICANN50
Ruling The Root

CJK Rules For The Root Zone

Kenny Huang, Ph.D.  黃勝雄博士

Member, CDNC / CGP
Co-author, RFC3743 IETF
Member, Executive Council, APNIC
Member, Board of Directors, TWNIC
huangksh@gmail.com
2014 Jun
Problem: CJK Is Complicated

Putting CJK labels in the root zone is even more complicated.
Institutionalized Problem Solving: Structure

Unified LGR for the Root Zone

Merge

Integration Panel

C (Chinese) Generation Panel

K (Korean) Generation Panel

J (Japanese) Generation Panel

CJK Coordination Committee
## Constraints for CJK LGR

### Independent Tasks

- Each CJK Panel creates an LGR
- Each LGR includes a repertoire and variants
  - Define labels permission
  - Define variants labels
  - Assign dispositions
    - Allocatable
    - Block

### Coordination Tasks

- If an LGR includes Han characters:
  - The variant *mappings* must agree for all the panels
  - The variant *types* may be different
  - The repertoires may be different

*Presented by Lee Han Chuan & IP, Shanghai 2014 May 29*
Overlap Case Illustration

Chinese LGR

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variant</td>
<td>1 壹 2 弐 3 壹</td>
</tr>
<tr>
<td>Unicode</td>
<td>U58F9 U5F0C U58F1</td>
</tr>
<tr>
<td>Disposition</td>
<td>allocate block block</td>
</tr>
</tbody>
</table>

Japanese LGR

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Variant</td>
<td></td>
</tr>
<tr>
<td>Unicode</td>
<td></td>
</tr>
<tr>
<td>Disposition</td>
<td></td>
</tr>
</tbody>
</table>

Integrate?

- T: Integrated Root Zone Label Generation Rules
- F: Rejected
- Generation Panel
# High Level Conflict Strategies

<table>
<thead>
<tr>
<th>ID</th>
<th>Strategy</th>
<th>Pros</th>
<th>Cons</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adopt X and Abandon $R_{cjk}$</td>
<td>Permit $X$</td>
<td>No label rule</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Adopt X $\cap (R_{cjk})$</td>
<td>Permit $X$</td>
<td>Rules changed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Permit $\cap (variants/disp)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Adopt X $\cup (R_{cjk})$</td>
<td>Permit $X$</td>
<td>Rules changed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Permit $\cup (variants/disp)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Abandon X and $R_{cjk}$</td>
<td>No conflict</td>
<td>Label not available</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Adopt rules based on frequency of use</td>
<td>Fair &amp; scientific approach</td>
<td>Rules changed; fairness doesn’t mean appropriate</td>
<td></td>
</tr>
</tbody>
</table>

$C$: rule $R_c$  
$J$: rule $R_j$  
$K$: rule $R_k$  

**CJK overlap**
Unified CJK LGR Illustration

Chinese LGR

- U4E00

<table>
<thead>
<tr>
<th>Variant</th>
<th>➱</th>
<th>➲</th>
<th>➳</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicode</td>
<td>U58F9</td>
<td>U5F0C</td>
<td>U58F1</td>
</tr>
<tr>
<td>Disposition</td>
<td>allocate</td>
<td>block</td>
<td>block</td>
</tr>
</tbody>
</table>

Japanese LGR

- U4E00

<table>
<thead>
<tr>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicode</td>
</tr>
<tr>
<td>Disposition</td>
</tr>
</tbody>
</table>

Integrated LGR

- U4E00

<table>
<thead>
<tr>
<th>Variant</th>
<th>➱</th>
<th>➲</th>
<th>➳</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicode</td>
<td>U58F9</td>
<td>U5F0C</td>
<td>U58F1</td>
</tr>
<tr>
<td>Disposition</td>
<td>allocate</td>
<td>block</td>
<td>block</td>
</tr>
</tbody>
</table>

Union

Intersection
CJK Integration Methodology

Divide & Conquer (D&C)

Diversified CJK Demands

- C Demands
- J Demands
- K Demands

Plan and Define

Strategic Direction

Unified CJK Rules

- Root Zone Admin
- CJK Rules
- CK Repertoire

Evaluation Method

- LGR
- Constrains
- MSR

Split

Variant Dispositions

- Minimal Viable Solution
- Repertoire

CJ Overlap

- CK Overlap
- CJK Overlap

CJ Usage Pattern

- CK Usage Pattern

Merge
Splitting Non-overlapping Code Points From Repertories

CJK Han-overlap in IANA IDN Repository

- J-Han : 6356 (JPRS)
- K-Han : 0 (KRNIC)
- C-Han : 19520 (CNNIC/TWNIC)

Develop Conflict Strategy

No conflict

Problem Domain (Unsolved Overlap) : 6181
Engineering Design

Computation for Word Usage and Frequency

TC : Apple News
SC : Sina News
JP : Mainichi News
Sample size is statistical significant

Matching

C/J overlap code points

usage

Split unused code points

frequency of use

Split code points of low frequency of use
Splitting Unused Code Points from The Overlap

C / J Overlap Data Set : 6181

J only : 203
C only : 1927
R_c C only : 1927
R_j J only : 203

total unused : 2739

C / J usage overlap : 1312

unified code points:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2739</td>
<td>203</td>
<td>1927</td>
</tr>
<tr>
<td>+</td>
<td></td>
<td>1312</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4869</td>
</tr>
</tbody>
</table>

Problem Domain (Unsolved Overlap) : 1312
Computing Frequency of Use of Code Points

Initial Data Set : 1312
Top 10 Most Popular Words

TC
- 台, 718
- 的, 2774
- 人, 1005
- 在, 975
- 一, 964
- 是, 960
- 大, 776
- 有, 883
- 中, 896
- 不, 951

SC
- 日, 20942
- 月, 20315
- 国, 3754
- 人, 4430
- 中, 3521
- 被, 2791
- 称, 2340
- 南, 2152
- 生, 2027

JP
- 日, 822
- 年, 496
- 国, 393
- 会, 345
- 月, 325
- 人, 325
- 大, 319
- 市, 253
- 本, 251
Top 20: Chinese Frequency of Use > Japanese Frequency of Use

Generated Data Set: 939
Top 20 : Chinese Frequency of Use < Japanese Frequency of Use

Generated Data Set : 363
Chinese Frequency of Use = Japanese Frequency of Use

Generated Data Set : 10
Frequency of Use Reassembly

C / J Usage Overlap Data Set: 1312

Freq J > C: 363

Freq C > J: 939

Problem Domain (Unsolved Overlap): 10

<table>
<thead>
<tr>
<th>unified code points</th>
</tr>
</thead>
<tbody>
<tr>
<td>363</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>939</td>
</tr>
<tr>
<td>=</td>
</tr>
<tr>
<td>1302</td>
</tr>
</tbody>
</table>
Data Processing & Computation Recap

>20K Han Code Points
Splitting Non-overlapping

6181 CJK Overlap
Splitting Unused

1312 Usage Overlap
*Frequency of Use Computation*

Problem domain was effectively reduced
Future Work

Overlap range redefine
Expand (?) Std Dev.

Require intensive
CKJ coordination &
deliberation

Mean = 0.034465
S.D. = 0.158477

Chinese Frequency of Use Minus Japanese Frequency of Use
Re-consider Language Tag

Policy

Language tag support

Sources of Language Tag

- RFC 2860: The name space of language tags is administered by IANA
- ISO Standard 639:
  - when a language has both an IANA-registered tag and a tag derived from an ISO registered code, one MUST use the ISO tag.
- Maintenance Agency: International Information Centre for Terminology (Austria)
Perfection Syndrome

“Engineering isn't about perfect solutions; it's about doing the best you can with limited resources.” Randy Pausch

Thank You Question?

Kenny Huang, Ph.D.
huangksh@gmail.com
Community Work

Neo-Brahmi Prospective Generation Panel

#ICANN50
Neo-Brahmi Generation Panel
What is Brahmi?

• An ancient script

• Most of the modern scripts in Indian subcontinent have been derived from Brahmi

• Geographically the scripts being used in Central Asia, South Asia and South-East Asia

• These scripts are used by multiple language families: Largely by Indo-Aryan and Dravidian
Why Brahmi?

- All the scripts derived from Brahmi are Abugida, also known as alphasyllabary.
- The lineage, example of “Letter Ka”
Why Brahmi?

• Despite their variations in the visual forms, the basic philosophy in their usage is common

• They all are “akshar” driven, and follow a specific syntax
  
  – Analogical reference can be made to Indian National standard, IS 13194:1991 – Section 8

• This syntax being the implicit foundation in representation of these scripts in the digital medium, adherence to the structure acts as a obligatory security consideration even in the case of Internationalized Domain Names.

• Here we are dealing with “root”
Why Neo-Brahmi?

• Of all the scripts derived from “Brahmi”, not all are in modern usage

• Approach is in consonance with the conservatism principle of the LGR procedure
Past Experience

• Experience of working for IDN ccTLDs for 22 Official Indian Languages

• Each language is having its own
  • Code point set
  • Variant
  • Complex Whole label Evaluation Rules
Current Members List

• Currently the group is of 10 members
  • Mixed bag expertise like linguistic, Unicode
  • Need more members to cover the diversity within the group

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Udaya Narayana Singh</td>
<td>Raiomond Doctor</td>
</tr>
<tr>
<td>Mahesh D. Kulkarni</td>
<td>Anupam Agrawal</td>
</tr>
<tr>
<td>Akshat S. Joshi</td>
<td>Abhijit Dutta</td>
</tr>
<tr>
<td>N. Deiva Sundaram</td>
<td>Neha Gupta</td>
</tr>
<tr>
<td>Nishit Jain</td>
<td>Prabhakar Pandey</td>
</tr>
</tbody>
</table>
Neo-Brahmi GP
Internal Composition

Integration Panel

Neo-Brahmi GP

Tamil Sub-panel

Telugu SP

Devanagari SP

Gujarati SP

Gurmukhi SP

Bengali SP

Hindi

Marathi

Konkani

Nepali

Bodo

Dogri

Maithili

Santhali

Bangla

Assamese

Manipuri

SCRIPT SECTION

LANGUAGE SECTION
Get Involved

• The group is currently working on gaining more participation within and outside India.

• For participation, please refer “Call for participation - Neo Brahmi Generation Panel”
  • Published at ICANN community page
    https://community.icann.org/display/croscomlgrprocedure/Root+Zone+LGR+Project

• Interested individuals can send their expression of interest to
  neobrahmiGP@icann.org and idntlds@icann.org
Progress so far…

• Reviewed and commented on Maximal Starting Repertoire for the Root

• A workshop is planned in AprIGF on “Bringing diverse linguistic communities together for a unified IDN ruleset” for reaching out to the community for the wider participation in the panel

• Working on the Neo-Brahmi Generation panel proposal – May submit to ICANN by end of August/early September
THANK YOU

धन्यवाद
Community Work

Arabic Generation Panel

#ICANN50
Task Force on Arabic Script IDNs: Overview and Progress

@ ICANN London Meeting (June ‘14)

Task Force on Arabic Script IDNs (TF-AIDN)
Middle East Strategy Working Group (MESWG)

tf-aidn@meswg.org
Community Driven Way forward: Task Force on Arabic Script IDNs

- Creation and oversight by community based Middle East Strategy Working Group (MESWG; https://community.icann.org/display/MES/MESWG+Members)

- TF-AIDN Objectives: a holistic approach
  - Arabic Script Label Generation Ruleset (LGR) for the Root Zone
  - Second level LGRs for the Arabic script
  - Arabic script Internationalized Registration Data
  - Universal acceptability of Arabic script IDNs
  - Technical challenges around registration of Arabic script IDNs
  - Operational software for registry and registrar operations
  - DNS security matters specifically related to Arabic script IDNs
  - Technical training material around Arabic script IDNs
Membership

• Currently **26 members** – applications still being received
• From **15 countries** – Australia, Egypt, England, Ethiopia, Germany, Iran, Jordan, Lebanon, Malaysia, Morocco, Pakistan, Palestine, Saudi Arabia, Sudan, and UAE
• Speaking than **nine languages** – Arabic, Malay, Saraiki, Sindhi, Pashto, Persian, Punjabi, Torwali, Urdu, with **expertise** in use of Arabic script from East Asia, South Asia, Middle East, North Africa and Africa
• Coming from **diverse disciplines** – academia (linguistics and technical), registries, registrars, national and regional policy bodies, community based organizations, technical community
Task Force on Arabic Script IDNs

• Membership open, community based
• Details and interests of members posted by MESWG
• Discussions publicly archived
• Details at http://lists.meswg.org/mailman/listinfo/tf-aidn
• Background and Introduction to TF-AIDN
  – https://community.icann.org/display/MES/Task+Force+on+Arabic+Script+IDNs
• Workspace, news and document archive
  – https://community.icann.org/display/MES/TF-AIDN+Work+Space
• Email Archive
  – http://lists.meswg.org/pipermail/tf-aidn/
# Arabic Script TLDs Assigned or Delegated

1. الجزائر
2. عمان
3. ايران
4. امارات
5. بازار
6. پاکستان
7. الاردن
8. بھارت
9. المغرب
10. السعودية
11. السودان
12. ملیسیا
13. شبكة
14. سورية
15. تونس
16. مصر
17. قطر
18. فلسطين
19. موقع
### SUMMARY OF THE CODE POINTS

<table>
<thead>
<tr>
<th>Color</th>
<th>Description</th>
<th>No. Of codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISALLOWED by MSR</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>ALLOWED by MSR</td>
<td></td>
<td>227</td>
</tr>
<tr>
<td>Not Allowed by IDN2008</td>
<td></td>
<td>64</td>
</tr>
</tbody>
</table>

**Total**: 339

- Submitted to MSR: 172
- Discussion on codes to be handed over to LGR: Work is under process
# IDN Variants Needs and Challenges

## Security and Stability Needs

<table>
<thead>
<tr>
<th>Urdu</th>
<th>Arabic</th>
<th>Unicode Codepoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>پاکستان</td>
<td>و.poly</td>
<td>U+0643 U+06A9</td>
</tr>
</tbody>
</table>

*xn--mgbai9a5eva00b*  *xn--mgbai9azgqp6j*

- 120+ cases of visually same or similar Arabic script characters identified by case study team
  - Variants must not be allocated independently
  - Variants may need activation to allow user access (w/ different KB)
- 16 IDN ccTLD applications with 4 applications with variants

## Security and Stability Challenges

- Consistency and innumerability
  - Consistent across and within TLDs
  - Minimal activation for manageability
- Management tools
  - Registration
  - Configuration and Maintenance
  - Security and Monitoring
- Usability in applications
  - Browsing, emailing, etc.
  - Searching, privacy, etc.
Progress

Work Accomplished

– Arabic Script Generation Panel
– Principles for Inclusion, Exclusion, and Deferral of Arabic
  Script Variants
– MSR Analysis and Feedback
– Principles on Variants
– Code Points for LGR

Outreach to the Community

– Launch at the Arab IGF Meeting in Algiers
– Presentation during the IGF in Bali
– Outreach during the ME DNS Forum
– Presentation to the community at ICANN Singapore
– Presentation to the community at the APTLD Meeting
Current Work and Next Steps

• XML Manual [June 30, 2014]
• Finalize the discussions on Code Points [August 28, 2014]
• Finalize the discussions on Variants [September 30, 2014]
• Whole Label Rules – Aug – Oct 14
  – Document principles for whole label variants
  – Define whole label variants
  – Release for Public Comments
• Finalization – Nov –Dec 14
  – Finalize LGR for Arabic script
  – Submit to ICANN/IP
  – Release for Public Comments
Thank You
LGR project mailing lists

- **LGR@icann.org**: Communicate with LGR community members and the Integration Panel on matters related to LGR work

- **IntegrationPanel@icann.org**: Contact directly the Integration Panel members on all matters related to LGR work

- **ArabicGP@icann.org**, **ChineseGP@icann.org**, **CyrillicGP@icann.org**, **KoreanGP@icann.org**, **NeoBrahmiGP@icann.org**: script community dedicated mailing lists

- **idntlds@icann.org**: Contact ICANN to submit Generation Panel proposals, individual statement of interests, work reports, updates, etc.

- Discuss issues related to the IDN Variant TLDs Program by subscribing to **vip@icann.org** here: [https://mm.icann.org/mailman/listinfo/vip](https://mm.icann.org/mailman/listinfo/vip)

#ICANN50
ICANN IDN Team: Thank You

USEFUL LINKS:

• MSR-1 Public Comment: [https://www.icann.org/public-comments/msr-2014-03-03-en](https://www.icann.org/public-comments/msr-2014-03-03-en)
• Setting up and running a Generation Panel: [https://community.icann.org/display/croscomlgrprocedure/Generation+Panels](https://community.icann.org/display/croscomlgrprocedure/Generation+Panels)
• Community Wiki LGR Project website: [https://community.icann.org/display/croscomlgrprocedure/Root+Zone+LGR+Project](https://community.icann.org/display/croscomlgrprocedure/Root+Zone+LGR+Project)
• For more info on the IDN Variant related pages, please visit: [https://www.icann.org/resources/pages/variant-tlds-2012-05-08-en](https://www.icann.org/resources/pages/variant-tlds-2012-05-08-en)
• To submit expressions of interest, or if you have additional questions, please contact ICANN at: [idntlds@icann.org](mailto:idntlds@icann.org)