
LONDRES – Programa de TLD con variantes de IDN
Miércoles, 25 de junio de 2014 – 08:30 a 10:00
ICANN – Londres, Inglaterra

PRESENTADOR:siete cadenas y hay consultas respecto de cuándo se van a incluir en el MCR también, en cuanto a que el RGA haga su trabajo.

También hubo una pregunta en el sentido que el MCR sí incluye algunos códigos y no hubo comentarios en el proceso de comentario público, por lo tanto no quedó claro si esos repertorios fueron revisados por la comunidad y por lo tanto sería posible que la comunidad lo revise después.

También hubo necesidad de llegar a otras comunidades para que la gente pudiese involucrarse más y cuáles son los esfuerzos que se iban ocurriendo en ese contexto. Hubo también algunas consultas sobre la inclusión de algunos (code points) incluyendo idiomas hablados por comunidades más pequeñas.

Voy a dejar la respuesta sobre el contenido para la próxima presentación al panel de integración, pero rápidamente en lo que respecta al proceso para abordar los procesos planteados por la comunidad y lo que se está considerando en este momento son libraciones cíclicas, tanto de MCR como de LGR como para que las comunidades de ICANN puedan organizar su trabajo.

Nota: El contenido de este documento es producto resultante de la transcripción de un archivo de audio a un archivo de texto. Si bien la transcripción es fiel al audio en su mayor proporción, en algunos casos puede hallarse incompleta o inexacta por falta de fidelidad del audio, como también puede haber sido corregida gramaticalmente para mejorar la calidad y comprensión del texto. Esta transcripción es proporcionada como material adicional al archivo, pero no debe ser considerada como registro autoritativo.

Y en ese contexto, lo que va a ocurrir es que las comunidades que no pudieron dar su opinión en este momento o no han considerado ciertos (code point) lo van a poder hacer en las próximas emisiones del LGRA y MCR.

Este proceso va a ser adicional es decir que una vez que se vieron se pueden hacer cambios y pueden haber cuestiones de seguridad y estabilidad en el sistema y se pueden agregar más adelante.

Hay algunos esfuerzos de difusión externa que está realizando la ICANN en este contexto. La ICANN le ha dedicado varios años a hacer difusión en la comunidad para poder ayudar a que haya una formación de paneles y se trata de llegar a las comunidades que no están organizadas aún, y tratar de que se organicen.

El MCR 1 ya está emitido, estamos tratando de cubrir varias de las cadenas de caracteres que quedaron afuera y eso va a ocurrir en el primer trimestre del año que viene.

El grupo de trabajo del LGR va a empezar ni bien recibamos las propuestas de los paneles de generación, por eso en base a los aportes de comunidad si recibimos paneles de generación, al principio del año próximo esperamos tenerlo emitido para el primer trimestre del 2015.

Los paneles de generación que se formaron después no van a poder presentar sus propuestas al principio del 2015 y se van a considerar en las próximas emisiones del LGR, por lo menos ese es nuestro plan en este momento y nosotros le pedimos a la comunidad que nos dé su opinión en este sentido.

Nosotros le pedimos a la comunidad que se organice en paneles de generación e inicie este proceso, todavía hay muchas cadenas de caracteres por cumplir y si ustedes quieren ser voluntarios o ayudar en el proceso, yo les puedo ayudar.

También tenemos un taller muy detallado sobre los distintos paneles que organizan los desafíos que están teniendo, lo que estamos discutiendo en esta sesión. Pero vamos a tener un taller más detallado después así que encuentren el tiempo para poder ir.

Estas son algunas de las referencias que vamos a tener para ustedes, a través de este sitio web.

Le voy a pasar ahora la palabra a nuestros colegas, Michele y Nicholas que son miembros del panel de integración y nos van a contar los detalles del informe del repertorio máximo inicial que se está generando.

Tiene la palabra Nicholas.

ORADOR SIN IDENTIFICAR: Buenos días, yo soy (...), soy miembro del panel, soy experto en UNICODE, represento los intereses de UNICODE, en el panel de integración y de hecho también participo en el lado ISO de todo esto y se ha hecho un trabajo especialmente en el Este de Asia que está sincronizado con UNICODE y por eso es importante tener una conexión.

Anteriormente yo implementé AIDN para Microsoft y por eso tengo también experiencia y conocimientos y por eso estoy en el panel de integración.

De todas formas vamos a hablar acerca de MCR esta mañana, MCR 1, se desarrolló el 20 de junio, poco tiempo antes de esta reunión, pueden dirigirse al vínculo para ver el anuncio sobre los distintos documentos que constituyen el MCR 1, eso significa que ahora el trabajo puede proceder con 22 códigos de escritura, ahora vamos a hablar acerca de qué significa tener 22 códigos en MCR 1.

Los paneles de generación recibirán, elegirán el repertorio desde dentro del MCR y veremos cuáles son los caracteres que se pueden utilizar, para agregar a las LGR.

También es necesario (...) variantes, esto no se aplica a todos los códigos, es más bien una preocupación relacionada con CF... pero mucho de los códigos pueden utilizar esto. Una de las variantes que tenemos que decidir es si van a ser asignables o bloqueadas. Esa decisión excede nuestro espectro, nosotros ahí no tenemos decisión simplemente lo que decimos es que el panel de generación tendrá que decir esto.

Tenemos que ver cuáles son asignables y cuáles no pero eso excede nuestro ámbito, tenemos que generar una propuesta de LGR, el panel de generación creará el panel de LGR que serán presentadas para revisión pública y luego serán integradas por el panel de integración.

En números, 22 códigos, acá tenemos la lista. El punto principal es que esto no significa que sean todos los sistemas, hay una diferencia entre códigos y sistema de escritura, por ejemplo, acá tenemos (han), (han) utiliza no solamente caracteres chinos sino también japoneses, y quizás también se utilizan en Corea.

Dependiendo de la raíz.

Muchos de los elementos en la lista se utilizan en India, en el Sudeste asiático.

Esto incluye a la gran mayoría de códigos de escritura que se utilizan hoy. ¿Y por qué hicimos esta lista?

Está basada en los gTLDs propuestos por lo tanto muestra todos los códigos por los cuales había interés en los nuevos gTLDs.

Obviamente lo ampliamos un poco agregando algunos códigos relacionados. Podrán ver que algunos de los códigos de India están incluidos acá y no teníamos gTLDs para ellos, porque luego se agregó de Devenagari y por lo tanto agregamos los demás códigos de la India, lo mismo con Thai y Lao.

Tailandia pidió que se agregaran gTLD por lo tanto agregamos Lao, porque Lao y Thai son códigos relacionados, así es como determinamos esta lista de 22.

Los códigos comunes y heredados. Esto tiene que ver con puntos de código que necesitan múltiples códigos de escritura.

En total tenemos 32.790 puntos de código que es una gran cantidad, de hecho, desde una cantidad aún más grande de P validos, fíjense que tenemos casi 100 mil caracteres.

Lo vamos a reducir en cierta forma utilizando la sigla (han gul) y también (idiograma) (han).

Si hacemos los cálculos veremos que no necesitamos tanto, podríamos decir que es un tributo a la eficiencia, lo que observamos aquí.

En la gran mayoría de los casos vemos una reducción de 97 mil a 20 mil y esto se debe a que redujimos significativamente todo lo que es P “valid”, lo pasamos a un conjunto que es utilizado por los gTLDs.

Yo soy el representante simbólico de los comentarios públicos; se presentó el MCR 1 y está disponible para comentarios públicos. El período de comentario se extendió por lo menos una vez, y terminó hacia fines de mayo.

Recibimos dos clases de comentarios, en realidad tres, uno o dos comentarios decían “bueno, este es muy buen trabajo, muchas gracias”, que es lo que necesitábamos.

Sin embargo, había más –no diría crítica constructiva – sino comentarios acerca de poder cambiar el MCR 1 y había dos clases de comentarios, por un lado se decía que habíamos admitido algunos puntos de códigos que a los que hicieron los comentarios les gustaría que estuvieran incluidos.

Analizamos las sugerencias y vimos que estas sugerencias no eran P válidas o en algunos casos estaban dirigidas a áreas no estables del campo del punto de código, en el amplio campo de los caracteres (han) ocurre esto. Sin embargo nos pidieron en esta revisión y en ese análisis de los puntos de código que introdujéramos algunos caracteres nuevos y que compartiéramos la flexibilidad de criterio y pudimos así incluir 7 puntos de códigos adicionales.

Pero esto es poco en comparación con las decenas de miles que ustedes vieron recién y que están incluidos en MCR 1.

Otro punto que se señaló es que corríamos en riesgo de poner en peligro más idiomas al no incluir los códigos de escritura de los grupos minoritarios.

Este problema lo resolvimos de una forma que espero que sea objetiva, y que también, esté sometida a revisión en caso de que cambie la situación.

Lo que hicimos fue basar nuestros criterios en una escala independiente que surge del Instituto de Lingüísticas y que fue desarrollado específicamente por George Fisher, que si alguna vez trabajaron en el área de la lingüística social lo conocerán.

Es lo que llamamos la escala (EGUIX) que se expandió y es una escala inter-generacional, inter-generacional porque se ocupa de la perspectiva del idioma que puede ser propagado de una generación a la siguiente.

Tiene distintas escalas, distintos niveles que muestran en qué medida un idioma es seguro o corre riesgo en cuanto a la posibilidad de que se vea alterada la transmisión generacional.

Y hay una segunda versión de esta escala que se llama “la escala expandida”.

Es importante saber que la consideración básica no es el tamaño de la población sino lo que podríamos llamar la vitalidad establecida de un idioma.

Lo que estamos haciendo en este caso es tratar de encontrar una forma de caracterizar la demanda eficaz de un sistema de escritura. Que es algo que no se hace muy ampliamente en el mundo intelectual, pero es

un problema en lo que a nosotros respecta. Y hay ciertas probabilidades, por supuesto, de que lo logremos porque algunos idiomas son las entidades fundamentales acá, que están en más de un código de escritura y a veces la forma en que están incluidos en ese código de escritura utilizan algunos de los caracteres de forma impredecibles. Sin embargo hay documentación de todo esto, los idiomas escritos son cosas escritas y lo que decidimos fue que en esta escala intergeneracional hay 9 indicadores, 9 grados, podríamos decir, y el mejor punto en que deberíamos caracterizar el uso diario, es decir, los idiomas y los códigos de escritura que nos preocupan, deberían utilizarse diariamente.

Eso es lo que heredamos como criterio de procedimiento.

Ese punto de corte, el mejor punto de corte, nos parecía que estaba entre el punto 4, que es el uso educativo, es decir en donde el idioma se utiliza en forma vigorosa con la estandarización y con su aparición en la bibliografía a través de un sistema difundido de educación institucional.

Y también tenemos el nivel en desarrollo, en donde el idioma tiene bibliografía escrita.

Pero, esta literatura y bibliografía no está propagado por la comunidad, el sistema de escritura tampoco, y en muchos casos el uso preciso de los caracteres tampoco está y por eso necesitamos algo que lo indique así en la raíz.

Esta es la razón por la cual nos pareció que era lo correcto. Nos parece que esta escala es objetiva, nos parece que es un sistema que se actualiza cada tres o cuatro años, esto lo hace el SIL, por lo tanto es

posible revisar el estatuto del idioma así como su relación con el sistema de escritura. Por lo tanto pensamos que tenemos un indicador determinado en forma objetiva y una escala que no va a seguir siendo la misma para siempre; que puede ser revisada pero que nos permite contar con un objetivo fijo para agregar algo en el MCR.

Por supuesto hablé acerca de agregados del repertorio máximo inicial.

Y este primer repertorio máximo inicial es en principio lo que figura acá, el MCR 1. Hay otros códigos que se utilizan mucho pero por distintas razones no fueron incluidos en el MCR 1.

Y a veces se debe a que el código es pequeño y al parecer no estaba representado por las personas que estaban involucradas en la raíz en ese momento.

Y un ejemplo es el código armenio. Y hay otros como algunos de los códigos de sudeste asiático, como el (inaudible), el (MER) que tienen el problema de que la codificación del código no es estable y por eso nos pareció que crearía dificultades innecesarias en la primera etapa.

Sin embargo estamos pensando en incluirlos en la siguiente etapa en la que agregaremos códigos al repertorio máximo inicial. Por lo tanto es probable que haya más cosas, no se va retirar nada, sino que se van a incluir probablemente algunos códigos adicionales y posiblemente también algunos caracteres más dentro de los códigos de escritura que ya tratamos.

Estamos trabajando en una línea de tiempo que es la que pueden ver acá, y quizás resulte útil en este momento hablar acerca de cómo se

incorporan las distintas propuestas de LGR y cómo están relacionadas con los repertorios máximos iniciales.

Lo que muestra este diagrama es que la gente podría estar trabajando en un panel de generación que podría haber empezado, y de hecho empezó, especialmente el de CJK, chino, japonés y coreano y los grupos arábigos, comenzaron antes de que saliera MCR 1 y posible que los grupos comiencen ahora una vez que se terminó el MCR 1, y esperamos que haya una serie de códigos de escritura que tengan propuesta y esperamos poder validarlos, y esto marcaría la etapa de LGR 1.

Hay algunas cosas que podrían comenzar y que quizás no estén listas para ser validadas en LGC 1, esas van a continuar y quizás haya una superposición en el período en el que se continúe con MCR 2.

Y si no se validan en la fase de LGR1 tendrán que esperar a la siguiente validación en LGR 2.

Tal como tratamos de mostrar en este diagrama, estos paneles de generación podrían comenzar en cualquier momento y hay suficiente flexibilidad incorporada al sistema de forma tal que es una ventaja comenzar lo antes posible pero no se van a ver interrumpido.

Creo que esos son los puntos básicos de este diagrama.

¿Quisiera continuar usted?

ORADOR SIN IDENTIFICAR: Los paneles tienen que asegurarse de que las propuestas sean exitosas, su propio interés común apunta a que el (...) sean aceptadas cuando las proponen porque si hay un rechazo eso demora todo el proceso.

Entonces obviamente el panel de integración está disponible para ayudar al panel de generación a asegurar que se cuenten con todas las guías que indiquen qué es lo que se considera aceptable.

Cuando hacemos el trabajo sabemos que no existe el blanco o el negro en cuanto a que algo sea aceptado o no.

Peor tenemos que ayudar con nuestros conocimientos y experiencias para que las LGR tengan éxito.

Otro punto es que cuando los códigos están relacionados es muy importante la coordinación de los paneles de generación. Por ejemplo hay una preocupación acerca de las variantes y es necesario que haya una coordinación entre las distintas comunidades que trabajan en el sistema de escritura basado en (han).

Esto también se aplica a los caracteres latinos, griegos, en donde hay una diferencia en los caracteres cirílicos, griegos, etc. Por lo tanto es necesario que haya una cierta coordinación.

Tenemos una lista de "mailing" que es la forma en la que nos pueden contactar, si tienen alguna duda o pregunta pueden escribirnos y vamos a responder, a darle cualquier aclaración que necesiten acerca al proceso.

Y esta lista está archivada y es acceso público.

Un punto rápido con respecto al rol del XML en este proceso, los LGRs están representados en XML, esto puede parecer intimidatorio, la IANA está reformulando los gTLD, especialmente los nuevos gTLDs, entonces usamos la misma herramienta. Esta también es una forma de ser

congruente y de tener una forma sencilla de contar con herramientas que funcionan en estas representaciones de XML para asegurarnos de que se siga el proceso adecuado cuando alguien quiere verificar un dominio o una etiqueta, y verificar que una cadena de caracteres específica esté ahí.

Bueno, no voy a continuar, nos quedamos sin tiempo porque hoy comenzamos más tarde, pero si quieren alguna explicación pueden contactarnos.

Gracias.

ORADOR SIN IDENTIFICAR:

Gracias Michel y Nicholas, vamos a pasar directamente a la próxima presentación. Este fue el trabajo que hizo el panel de integración, y el trabajo que sigue haciendo el panel de integración que ahora va a pasar al trabajo basado en la comunidad. Y les vamos a dar una visión general de lo que está sucediendo y vamos luego a entrar más en detalle en las presentaciones de la comunidad en el taller que vamos a tener más adelante en esta sala.

Vamos a empezar entonces ahora con el panel de generación árabe, que se presenta en forma remota por el señor (Mohamed Masud). Que nos llama desde Pakistán en este momento y le vamos a dar ahora la palabra a él para que hable sobre la actualización del panel de generación de árabe.

Sí, lo escuchamos, adelante.

ORADOR SIN IDENTIFICAR: En representación del IDN y el (...) nosotros somos los miembros del IDN que tiene que llenar ese espacio vacío en el futuro.

Estoy representando algunas de las palabras, algunos de los trabajos, el (FIDN) es una fuerza de trabajo que trabaja bajo el grupo de trabajo de Medio Oriente que fue creado con el objetivo de trabajar en el código de árabe para la zona de la raíz y los datos de registración internacional del script árabe, el IDN árabe, los desafíos técnicos en la registración de los script de árabe, el software operativo para las operaciones de registro y registradores. Los asuntos de seguridad con respecto al script arábigo y los materiales técnicos relevantes para la comunidad árabe que están siendo llevados a cabo por el (TFIDN) en forma regular y que empezaron a partir del CUI y la comunidad tiene que seguir trabajando en eso.

Vamos a la siguiente diapositiva.

El TFIDN en este momento tiene 26 miembros de 15 países, no solamente de las regiones definidas del Medio Oriente, sino también de otros países como Australia, Inglaterra, Etiopía, Alemania y Malasia.

En estos lugares se hablan nueve idiomas como árabe, malayo, (...) persa, (...), y también se cubren muchos otros idiomas. Estos miembros también cubren idiomas del sudeste asiático, el sur de Asia, Medio Oriente, África del Norte y África en general.

Estas personas tienen contextos muy distintos como el contexto académico, registros, registradores, organismos de políticas regionales y nacionales y organizaciones comunitarias.

La membresía de este grupo está abierta para todas las personas que puedan contribuir para la comunidad, esencialmente tienen experiencia en el script arábigo y en el DNS, y la información respecto del grupo se actualiza en forma regular en el sitio web del grupo, también en la comunidad de ICANN en el espacio wiki. Y en el sitio web central de ICANN.

Además, nuestros archivos de mails también están disponibles en el sitio web del grupo de trabajo de Medio Oriente.

Y les agradecemos por actualizar la información eficientemente sobre todos estos asuntos.

Siguiente diapositiva.

Estos son los scripts arábigos de TLDs que han sido asignados o delegados.

Para los que no lo pueden leer, uno es Argelia, dos Omán, el tres es Irán, cuatro UAE, cinco Pakistán, seis Jordania, siete (...), ocho Marruecos, nueve Arabia Saudita, diez Sudán, once Malasia, trece (...), catorce Siria, quince Túnez, diez y seis Egipto, diez y siete Country, diez y ocho Palestina y diez y nueve (...).

Siguiente diapositiva por favor.

Los IDNs han sido discutidos uno por uno en cada uno de los espacios con gran deliberación, se discutieron 339 caracteres a pesar de que hubo algunos que no fueron permitidos por el MCR, y el (...) no permitió cubrir los caracteres alfa.

Después de una deliberación el TFIDN ha recomendado otros 172 caracteres MCR y tenemos en TFIDN estamos trabajando para reducir estos caracteres en el RGR.

Todavía está el trabajo en proceso, pero lo vamos a tener listo dentro de algunos días.

Además TFIDN está deliberando sobre otros asuntos teniendo en cuenta la seguridad y la estabilidad, y también las reglas de “phishing” que puede haber. Tenemos que definir reglas para la inclusión, exclusión y diferimiento del (code point) y de la variante de IDN.

Hasta ahora ha habido más de 120 casos similares que han sido identificados por los caracteres del script árabe y que han sido identificados por el caso de estudio. Las variantes no deben ser asignadas independientemente y durante el proceso de registración también se discute lo que sucede en el grupo. El grupo está discutiendo la usabilidad en aplicaciones como emails, navegadores y otras cuestiones relacionadas.

Mientras hablamos del progreso el TFIDN ha finalizado el script árabe para el panel de generación, se ha recibido la respuesta el MCR y esto tiene que ver con la adición, el diferimiento que se mencionaron en la diapositiva anterior y la variante principal para la adicional y diferimiento ya también han sido finalizados.

Hay distintas actividades de difusión externa que se realicen como una reunión en At-Large, esto ocurrió durante el IGF en Bali, también en el foro de DNS. Hubo una presentación para la comunidad en Singapur, y

una presentación para la comunidad en la reunión de APTLD, y ahora en la reunión número 50 de ICANN en Londres.

Hay plazos definidos por el grupo, que han sido decididos mutuamente y durante la reunión en persona y a través también de listas de correo y en forma remota se tomaron decisiones.

Nosotros estamos esperando finalizar esta regla de variante de generaciones para finales de este año, para la fuerza de trabajo, el grupo de trabajo, la fecha límite es diciembre de este año donde el grupo va a entregar todos los documentos a ICANN para comentario público.

Muchas gracias. Esto es todo de mi parte.

ORADOR SIN IDENTIFICAR: Gracias (Ahmed). Vamos a continuar. Tenemos ahora el informe sobre la coordinación CJK – SJK. Le vamos a pasar ahora la palabra a otro orador.

KENNY HUANG: Buenos días damas y caballeros, es honor hablar del este Comité de Coordinación. Soy Kenny Huang, primero, todos vemos que el (...) hasta acá es muy complicado. Y poner una etiqueta CJK en la zona raíz es incluso más complicado, de hecho no es una pregunta sino que es una realidad.

Cuando nos encontramos con este tipo de situación, ¿cómo podemos poner un CJK en la zona raíz?

El primer paso para resolver el problema es una estructura institucionalizada, la estructura propuesta por ICANN o por el panel de

integración, en este panel de integración nosotros tenemos oficialmente el panel de generación C, el K y el J, y cada tipo de coordinación debe estar bajo cada generación individual.

Como dije, el comité de generación CJK o el comité de otro tipo, nos da un potencial con los TLDs.

El panel de generación de CJK no está aprobado oficialmente hasta ahora pero vamos a en ese sentido hacia una aprobación. El problema que enfrentamos no tiene que ver con que nosotros tenemos un sistema sin ningún sitio de control. El panel genera el LGR, incluye un repertorio y las variantes, incluye permisos de etiquetas de variantes, etiquetas de variantes definidas y también las disposiciones asignables y las bloqueadas, pero incluso en este caso tenemos un gran esfuerzo de coordinación.

Por ejemplo si un LGR incluye un carácter JAN, el mapeo de variante tiene que estar de acuerdo con todos los paneles y los tipos de variantes podrían ser diferentes, los repertorios también podrían ser diferentes.

El nivel de coordinación que se nos dio de ICANN –estoy tratando de ir rápido porque no tenemos mucho tiempo – y voy a tratar de ir más lento.

Mucha gente viene aquí a hacer preguntas, nos preguntan qué tipos de problemas vamos a tener. Les voy a dar un caso muy simple que es superposición entre la comunidad china y la japonesa y probablemente también la coreana.

Las reglas de generación de chino, el (code point) es uno, y también este (code point) es X en el uso japonés. Nosotros tenemos tres variantes en

este (code point), tenemos la variante 1, 2, 3, y ellos tienen distintas disposiciones, por ejemplo, la primera variante definida en la RSC 43 tiene que ser asignable y nosotros la llamamos variante específica. El resto de las dos variantes son variantes generales y la situación en la que nosotros entramos en la operación es de dos variantes.

Ahora volvemos al japonés y qué ocurre ahí. El japonés 1 no tiene ningún tipo de variante, utilizando el mismo (...) pero con distintas variantes y con distintas disposiciones entre el chino y el japonés, si nosotros proponemos este tipo de variantes y este conjunto de reglas al panel de integración, eso va a tener un problema de (swuch) donde la propiedad no se va a integrar. Este tipo de propuesta va a ser rechazada y va a volver al panel de generación. Ese es el problema con C, y que vamos a tratar de resolver.

Nosotros tenemos conflictos potenciales en la comunidad CJK, inicialmente teníamos un nivel muy alto de conflicto. Y no abordábamos cada (code) por separado, teníamos una idea general de cómo resolver este tipo de situación conflictiva.

Esto podría hacer muchas combinaciones. Por ejemplo en el (...) original hablamos de X que existe en un repertorio CJK, durante el repertorio CJK hubo distintas reglas, por ejemplo, en chino tenemos la regla C, para japonés tenemos la regla R, RJ, y para coreano tenemos la regla K. Sin embargo, cómo integramos los tres tipos de reglas bajo un (code point) registrado X.

La combinación podría ser un (code point) original y también un conjunto donde nosotros podemos abandonar las reglas anteriores bajo el CJK.

Esa es una situación. Y también podemos adoptar el (code point) X y elegir la intercesión entre las reglas CJK. Podemos también adaptar X o eligiendo la unión de las reglas CJK. Podría ser (ABANDON X, ABANDON R) o podría ser elegir bajo la frecuencia de uso.

Hay muchos tipos de cosas a favor y en contra que podemos aplicar, esta es una estrategia con un muy alto nivel de conflicto que no se aplica a todos los (...) individuales.

Vamos a reutilizar la misma regla para elegir cuáles son los tipos de estrategias que vamos a adoptar.

Por ejemplo en el (inaudible), como mencioné antes, esto es para el chino que tenemos tres tipos de variante. Para el japonés, el japonés no tiene ninguna variante.

Si nosotros elegimos la unión o UNICODE y unimos las tres variantes, también si elegimos la intersección de las tres variantes esto quedaría pendiente. Por lo tanto, de cualquier forma, ya sea el uso de japonés o de chino de (hun) no parece muy justo para las comunidades, pero a pesar de que hace mucho tiempo que tenemos la operación de IDN con respecto a los ccTLDs, cuando tratamos de aplicar este tipo de regla a la zona raíz la situación sería totalmente diferente.

¿Qué clase de metodología potencial podemos utilizar si tratamos de resolver un problema? Es una forma potencial de dividir y conquistar. Tratamos de dividir todos los componentes en subcomponente, y cada pequeño componente con una metodología diferente.

Inicialmente, usemos nuestra imaginación porque todavía no empezamos con la coordinación real, aunque no tenemos un panel de generación oficial de CJK.

Inicialmente podemos hablar de una superposición en el (...) de repertorio, una vez más, no tenemos el repertorio todavía, entonces, yo elegí el repertorio del repositorio de IDN. Y en este depósito de IDN de la IANA en chino y en japonés, tenemos 6181 (code points) superpuestos, no encontré ninguna superposición entre chino y coreano o – es decir que a menos que el panel de generación de coreano haya propuesto nuevo repertorio sobre la base del depósito o repositorio de IDN, no hay superposición entre ambas comunidades.

Entonces, sobre la base de este supuesto podemos (...) en la superposición entre chino y japonés y tenemos que ver cuáles son las estrategias de conflicto.

En esta situación redujimos los (code points) a 6 mil.

Ahora tratamos de hacer una especie de diseño de ingeniería o informático, por ejemplo, podemos capturar la fuente de la noticia, por ejemplo de un medio de noticias tradicional chino y a través del (code) tratamos de tomar el beta del (...) y en japonés del (inaudible), tratamos de basarnos en el uso y la frecuencia de uso, y este es el resultado computacional de esa primera etapa.

Acá podemos dividir los puntos de código no utilizados desde la superposición. Acá tenemos el chino tradicional y el japonés, y vemos que hay puntos de código no utilizados. Puntos de códigos no utilizados, y esto es lo que ocurre en el uso diario en las noticias. En total la

superposición es de 1312, bueno, finalmente tenemos que calcular la frecuencia de uso en estos puntos de códigos. Tenemos 1312 conjuntos de datos, este es el resultado de las diez palabras más populares en el chino tradicional simplificado y japonés. Y estas son las diez palabras más utilizadas.

Acá tenemos la frecuencia de uso chino y vemos que la frecuencia del uso del chino es mayor que la frecuencia de uso de japonés, elegimos los 20 principales a los fines de la ilustración y es lo que ven en esta diapositiva y en total el conjunto de puntos de códigos es el que observan acá, 939. Y también encontramos muchos puntos de código en japonés y vemos que la frecuencia de uso es mayor que la de chino y tenemos 363 conjuntos de datos generados.

Esta es la frecuencia de uso chino que es mayor que la frecuencia del uso japonés.

También vemos que coincide la frecuencia de uso chino con la frecuencia de uso japonés, Vemos que el conjunto de datos generados es 10, esto significa que hay dos conjuntos de datos, y por lo tanto la frecuencia de uso de chino y japonés estarían parejas.

Este es el resultado final, la frecuencia de uso, japonés igual a chino es 10, por lo tanto vemos que tenemos diez puntos de código.

Después del procesamiento de los datos y después de los resultados hechos a través de los cálculos informáticos vemos cuál es la cantidad de punto de código, vemos que tenemos una superposición mayor a 6 mil y tratamos de llegar a 1300, y al hacer los cálculos de frecuencia de uso

vemos que la superposición son 10 puntos de código y así se redujo el problema de los dominios.

En el futuro potencialmente, podemos tratar de redefinir el rango general de superposición, y si tenemos tiempo y recursos vamos a tratar de extender ese desvío, por ejemplo un desvío estándar podrían ser seis días, pero para eso necesitamos una mayor coordinación de CKJ y también una mayor deliberación.

Y acá tenemos la media que es 0,34465 y el desvío estándar de 0,15.

Ahora también trabajamos desde el punto de vista de las aplicaciones, podemos utilizar también un TAG y para este tipo de apoyo podemos reducir y validar requerimientos del mercado y acá vemos cuáles son los códigos que utilizaríamos.

Este es el fin de mi informe, muchas gracias.

ORADOR SIN IDENTIFICAR: Muchas gracias, ahora vamos a pasar a la última presentación de esta sesión. (Gupta) va a hacer una presentación acerca del trabajo relacionado con los códigos para Neo-Brahmi.

(MEHA GUPTA): Soy (Meha Gupta), del panel de generación de Neo-Brahmi, estamos en la fase de formación en realidad.

Voy a comenzar mi presentación con una introducción del Brahmi y el Neo-Brahmi, luego les voy a dar una actualización acerca de lo que está ocurriendo en este momento.

El brahmi es un código antiguo, la mayor parte de los códigos modernos en el subcontinente indio provienen del brahmi. Por ejemplo en India tenemos 22 idiomas oficiales, de los 22, 21 provienen del brahmi y geográficamente los códigos se utilizan en el Centro de Asia, Sur de Asia, y Sudeste asiático. Y estos códigos son utilizados por múltiples familias de idiomas, especialmente (indoariano). ¿Qué es Neo-Brahmi?

La mayor parte de los códigos provienen del brahmi, algunos están reconocidos, otros no, algunos tienen uso histórico, por lo tanto no podemos cubrir todos los idiomas, todos los códigos y acá consideramos solamente aquellos códigos que se utilizan en la actualidad.

Vamos a la composición del panel de generación de Neo-Brahmi. Actualmente el grupo está formado por diez miembros que provienen de áreas como lingüísticas, UNICODE y el sector académico.

Necesitamos una mayor participación, más miembros para cubrir toda la diversidad que necesita el grupo y vamos a tratar de cubrir todos los códigos de escritura y los idiomas de la familia brahmi, pero todo esto siempre depende de la disponibilidad de expertos en idiomas y en la comunidad, entonces quizás en la propuesta final algunos códigos de idioma queden afuera si no logramos los expertos para esos códigos.

El grupo actualmente está tratando de lograr la mayor participación dentro y fuera de India, estamos tratando de incorporar expertos. Queremos – las personas que están interesadas pueden enviar su interés a las direcciones de correo que figuran aquí.

Esperamos una mayor participación en el taller, después de la reunión de Singapur hicimos un llamado a la participación en el sitio web de la

ICANN, aquellos que quieren participar en el panel pueden enviar un mail expresando su interés a estas direcciones de correo electrónico.

Ya comenzamos nuestro trabajo, la primera actividad fue una revisión y comentario sobre repertorio máximo inicial. Se desarrolló un taller, tenemos planificado llevar a cabo un taller, en AP IGF para llevar las comunidades con diversidad lingüística a unirse para tener un conjunto de reglas IDN unificados.

Esta es la composición interna del panel de generación de Neo-Brahmi, acá tenemos los idiomas, pueden ver el panel de generación de Neo-Brahmi que tendrá múltiples paneles de código de escritura y si el código incluye más de un idioma quizás tengamos sub-paneles.

Acá vemos dos desafíos, por un lado, la comunicación entre los sub-paneles, el panel es muy grande, va a ser difícil comunicarse y coordinarse con los sub-paneles y en segundo lugar, vamos a necesitar tener una gran cantidad de especialistas en los distintos idiomas para cubrir todos los idiomas y todos los códigos de escritura.

Como ya dije, si no tenemos cubiertos todos los idiomas quizás tengamos que dejar algún idioma y código de escritura afuera en la propuesta final.

Quizás tengamos en el futuro también una división de Neo-Brahmi 1 y Neo-Brahmi 2.

Este es el fin de mi presentación, muchas gracias.

ORADOR SIN IDENTIFICAR: Muchas gracias. Acá tenemos algunos vínculos que pueden utilizar para conectarse con los paneles de generación y con los equipos de proyectos en la ICANN y también los paneles de integración, y entonces podemos abrir el espacio para preguntas y respuestas. Nos quedan unos diez minutos.

JASON POLIS: Soy Jason y me estaba preguntando acerca de cómo se trabajó anteriormente, por ejemplo el código griego, al parecer no se está trabajando con esto. Entonces, ¿se está trabajando o no?

ORADOR SIN IDENTIFICAR: Hubo una iniciativa en donde se identificaron los problemas relacionados con las variantes, para los códigos de escritura griego, este fue uno de los equipos que trabajó en la primera fase del proyecto, pero una vez más, como dije antes, en la primera fase se centraron en determinar cuáles eran los problemas posibles con estos códigos de escritura, con algunos códigos. Ahora, los miembros de la comunidad deben volver a reunirse y formar un panel de generación para proponer una solución para resolver los problemas. Por lo tanto esta es una segunda fase de la misma clase de trabaja.

Los problemas para el griego se identificaron pero ahora es necesario formar un panel de generación para proponer una solución final que luego será tratada por el panel de generación. Es decir, estos son los pasos que aún se deben recorrer.

JEAN JACQUES SUBRENAT: Buenos días, habla Jean Jacques Subrenat, miembro de ALAC.

Durante una de las reuniones de ALAC, Edmon Chung y Rinalia Abdul Rahim pidieron candidatos para prestar sus servicios en los distintos niveles, especialmente en los paneles de integración. Y yo hice eso esta mañana, envié mi nombre como candidato para esto. Pero después de haber escuchado estas explicaciones completas me pregunto si aquellas personas que no son especialistas en computación, ni tampoco lingüistas podrían participar acá. Yo estoy en algún punto ahí de la nube y tengo conocimiento muy vago, estudié chino, japonés y otros idiomas, pero no soy un experto en informática.

Entonces, antes de empezar a participar, les sirve que participemos personas como yo que trabajamos con idiomas en un contexto internacional pero no trabajamos como especialistas en sistemas ni tampoco como lingüistas.

Esa es mi primera pregunta. Mi segunda pregunta, es ¿qué opinan cuál sería el tiempo promedio necesario para que se ofrezcan los voluntarios para los paneles de integración? Un mes, seis meses, un año. Esa es una pregunta especial para Kenny, pero yo también me ofrecí como voluntario para los códigos latinos.

ORADOR SIN IDENTIFICAR: Para aclarar la terminología un poco, el panel de integración para nosotros es el panel de integración central, no solamente el panel de integración de CJK.

Y ese panel ya está formado. Ya tiene sus miembros y ya está trabajando. En cuanto a los paneles de generación, nosotros pedimos

aportes no solamente de especialistas técnicos, lingüistas y expertos en UNICODE, sino también miembros de la comunidad. Y es crítico y crucial contar con el aporte de la comunidad, en la formación de estos paneles, porque en última instancia este trabajo tiene implicancias o tendrá un impacto directo sobre la forma en que se van a utilizar los IDNs.

Entonces, esta es una perspectiva que resulta muy útil incorporar.

Por favor participen, si son miembros de la comunidad y si no sienten que tienen los conocimientos técnicos y específicos necesarios pueden aportar la perspectiva del usuario final y eso también es útil.

El proceso tiene múltiples etapas que se pueden ver como por ejemplo revisar un comentario, cada etapa o cada presentación se va a publicar y va a estar sujeta a revisión.

Obviamente todo el mundo tiene la libertad para hacer sus propios comentarios respecto de cuáles son los documentos que se proponen por cada uno de los paneles, incluso el IGI en sí. Va a haber muchas oportunidades para que todos contribuyan en el proceso.

KENNY HUANG:

El primer punto, en cuanto a la composición de cada uno de los paneles. El panel de CJK pidió paneles para que haya miembros y como nosotros tenemos un panel de generación de chino, de coreano y de japonés en conjunto, creo que es bueno que tengamos una combinación de otros panelistas.

Y en cuanto al tiempo de implementación, nosotros, si no recuerdo mal, pedimos que sea el 24 de octubre, pero eso tiene que ver con el CGP y tenemos que integrarnos con el CJP también por el potencial o la fecha potencial, creo que tenemos que redefinir ese plazo también.

¿Hay alguna otra pregunta?

Muy bien, si encuentran el tiempo únanse a nosotros, en el taller más detallado que está ocurriendo, vamos a tener distintas presentaciones en chino, japonés y coreano, además del árabe y el nuevo brahmi.

El taller también va a permitir un debate más detallado sobre algunos de estos puntos que están siendo tenidos en cuenta hoy.

Les agradecemos mucho por haber escuchado estas presentaciones, también agradezco a todos los panelistas por hacer estas presentaciones.

Vamos entonces a cerrar esta sesión.

[FIN DE LA TRANSCRIPCIÓN]