

ICANN72 | Semana de preparación – Actualización sobre las Reglas para la Generación de Etiquetas para la Zona Raíz
Jueves, 14 de octubre de 2021 – 11:00 a 12:00 PDT

PITINAN KOOARMORNPATANA: Buenos días, buenas tardes a todos y bienvenidos a la sesión de actualización de las reglas de generación de etiquetas de la zona raíz en la semana de preparación de la ICANN72. Para hoy tenemos esta agenda, rápidamente veremos una reseña de cinco minutos del proyecto de reglas para la generación de etiquetas en la zona raíz y luego le daré la palabra a nuestros oradores.

Tendremos a Mats Dufberg para el código de escritura latino, miembro del panel de generación, Hiro Hotta presidente del panel de generación del código de escritura japonés y Michel Suignard del panel de integración, quien se referirá a la próxima versión de la RZLGR, así que, me ocuparé de la primera parte.

Una reseña de las RZLGR, una breve historia primero del proyecto. Disculpas, espero que puedan ver. Alrededor del 2010 la comunidad de la ICANN identificó la necesidad de contar con dominios de nivel superior con variantes, para ello se conformó un grupo de la comunidad que trabajó en el informe de cuestiones integradas.

En esa época se exploró la necesidad de integrar variantes en la zona raíz y esto representaba una cuestión, fue un estudio de cuestiones, no de soluciones. En ese momento había seis códigos de escrituras,

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.

japones, devanagari, griego, latino y otros, eran múltiples tipos de códigos de escrituras que se utilizaban en distintas lenguas.

Se hizo el informe, el cual fue publicado en 2012, en este se identificó también la necesidad de tener un mecanismo que permitiera definir cuáles eran las variantes del código de escritura y para este mecanismo debía existir una única fuente porque eventualmente todo esto iba a ir a la zona raíz.

A partir de eso, el grupo identificó las LGR como ese mecanismo para definir el procedimiento para desarrollar y fue aprobado, el procedimiento, por la Junta Directiva de la ICANN en el 2013, desde ese entonces se constituyeron los paneles para los respectivos códigos de escrituras.

En paralelo, la Junta Directiva formuló recomendaciones para el manejo de los TLD con variantes de IDN. Estas recomendaciones fueron aprobadas por la Junta Directiva en el 2019, se resolvió que la GNSO y la ccNSO debían tener en cuenta estas recomendaciones en sus procesos de desarrollo de política y también en la integración.

En 2020 la Junta Directiva resolvió asimismo que la GNSO y la ccNSO tuvieran en cuenta las recomendaciones para la utilización técnica de las LGR de zona raíz y se resolvió en 2021, la GNSO también publicó su informe para los procedimientos posteriores que incorporó el uso de las RZLGR para la próxima ronda.

¿Cómo funcionan las reglas de generación de etiquetas o RZLGR? Son una fórmula para las etiquetas, son las reglas para la generación de etiquetas. Cuando estas reglas se integran a la zona raíz pueden ser

leídas por la herramienta y ahí se puede identificar mecánicamente si la etiqueta es válida o no.

Si es una etiqueta válida también se puede identificar su variante, que puede ser asignable o estar bloqueada, el mecanismo del procedimiento de RZLGR funciona en dos pasos. Primero es el trabajo que hace el panel de generación, que desarrolla una solución para el código de escritura en particular y una vez que el panel de cada código de escritura ha completado la solución, esta se presenta al panel de integración.

El panel de integración considera la propuesta que formuló el panel de generación o GP y puede haber varias instancias de consultas entre el IP y el GP, y una vez que la solución es aceptada de manera exitosa, esta es integrada a la zona raíz, así puede ser utilizada para validar los TLD.

Este es el resumen del trabajo de los distintos GP, para estos 28 códigos de escrituras que tenemos en pantalla... Hay trabajos desarrollados en 17 ya, comenzamos en el 2014 con distintos años y en este momento muchos de los paneles de generaciones ya han concluido su trabajo y en este momento tenemos los códigos de escrituras japoneses y latino que están en el procedimiento.

El de Myanmar es el que está finalizando la propuesta para el próximo comentario público. Bueno, esta es la línea de tiempo de las reglas de generación de etiquetas para la zona raíz, a partir del año 2013 cuando se hizo la convocatoria a los paneles de generación; que se completaron en el 2014, se fueron desarrollando e integrando las distintas reglas a la zona raíz.

En este momento estamos en la versión 4 de las RZLGR, habrá 18 códigos de escrituras en total y anticipamos que para el 2022 habremos integrado ocho códigos más, haciendo un total de 26. Aquí es donde estamos hoy, hoy tenemos miembros del GP latino, del GP japonés para mostrarnos la solución y para invitarlos a participar en el período de comentarios públicos.

Así que, le voy a pasar la palabra a Mats Dufberg del panel de generación del código de escritura latino.

MATS DUFBERG:

Muchas gracias, espero puedan oírme bien. Siguiente, por favor.

Estos son los componentes de mi presentación, yo hablaré de la propuesta que hemos presentado para comentarios públicos, estas son las principales áreas a las cuales me referiré. La siguiente.

Tenemos una propuesta para el código de escritura latino para comentarios públicos los cuales serán aceptados hasta el 23 de noviembre de este año, este es el vínculo hacia la propuesta y les solicitamos a todos que ingresen y analicen nuestra propuesta, consideraremos todos los comentarios, tanto los mayores como los menores, todos son bienvenidos. Siguiente.

Esta presentación entonces es una recorrida de esta propuesta y esperamos que no sea un umbral que impida que ustedes ingresen a la propuesta para analizarlo en detalle y darnos comentarios. El foco de la presentación será en el documento principal y sus apéndices, pero el

archivo LGR XML es el documento normativo y no me referiré demasiado a este.

Los capítulos que no discutiremos en esta presentación son capítulos breves con información general y no los veremos aquí. El capítulo dos de la propuesta define la delimitación de los códigos de escrituras procesados por esta propuesta.

La primera delimitación es que, la propuesta no puede incluir ningún carácter que no esté incluido en lo que se llama el repertorio máximo inicial, este MSR es un subconjunto de puntos de códigos válidos, el protocolo IDNA, que es un subconjunto de Unicode, por lo tanto, este MSR es un subconjunto de Unicode y esto es lo que tenemos que rescatar de la propuesta. El MSR ha sido definido por el panel de integración.

Durante el proceso hubo algunos caracteres que nosotros propusimos agregar al MSR y el IP, el panel de integración, añadió estos caracteres a una nueva versión del MSR. El MSR es un subconjunto vivo de Unicode, es decir, esto significa que puede crecer y también disminuir, entonces durante el proceso se propuso añadir algunos caracteres.

El capítulo 4 describe el proceso de nuestro trabajo, teníamos que determinar qué caracteres íbamos a utilizar para nuestra propuesta y decidimos utilizar lo que se llama la escala EGIDS; que explicaremos en la próxima diapositiva, desde el nivel 0 que son las lenguas de nivel internacional al nivel 4, educativo.

Y también añadimos el nivel 5, en desarrollo, cuando tenían al menos un millón de hablantes. La siguiente.

Aquí vemos la escala EGIDS, que es la escala de interrupción intergeneracional ampliada, esa es la que usamos, casi todas las lenguas del mundo han sido incluidas en esta escala en grados, entonces nosotros hemos seleccionado con qué lenguas íbamos a trabajar. De la 0 a la 4 y adicionalmente el nivel 5, que son las lenguas cuyos usuarios comprenden al menos un millón de hablantes, hay otras lenguas que fueron excluidas.

El apéndice 6 es una lista completa de todas las lenguas seleccionadas, no tengo aquí el número, pero son muchas. Para cada lengua incluida tenemos el nombre de la lengua, son 212 lenguas las que incluimos en nuestra selección, entonces en esta lista está el nombre de la lengua, el código de la ISO 6393 y también el nivel de la EGIDS para esa lengua.

Para cada una de las lenguas en la lista, identificamos todos los caracteres que se utilizan para escribir, entonces vimos cuáles eran los caracteres que se usaban para cubrir estas lenguas, la idea era que estos 212 idiomas debieran poder crear nombres de dominios de nivel superior para las palabras de esas lenguas.

El conjunto de caracteres de cada lengua no está documentado en el informe, pero hay referencias a cada lengua en el capítulo 9. También se identificaron posibles candidatos de variantes internamente de un código de escritura y entre códigos de escrituras.

El repertorio del código de escritura latino en la propuesta está basado en el punto de código de Unicode porque esa es la limitación, en el caso más sencillo, un punto de código es solo un carácter como la letra A, un

punto de código también puede ser un signo de modificación y ahí es cuando se usan con otro punto de código para formar un carácter.

En la diapositiva hay un ejemplo de la letra G con la virgulilla arriba que forma esta combinación de dos puntos de códigos. En muchos casos hay puntos de códigos pre compuestos con un carácter de base y un acento, por ejemplo, la letra A con el acento agudo, es lo que se ve en pantalla, ese es entonces un único punto de código.

Los principios, para incluir o no incluir un carácter identificado en una lengua, están detallados en la introducción del capítulo 5. Siguiendo diapositiva, por favor.

Bien, la propuesta para el código de escritura latino enumera 218 caracteres, 197 de ellos son de un solo código o punto de código y 21 caracteres están compuestos por una secuencia de dos o más puntos de códigos, es decir, esto incluye un símbolo o quizás dos.

Para cada carácter existe un código de punto de código o códigos, ya sea uno o una secuencia de los mismos. Los idiomas o el idioma que utiliza ese carácter para la escritura, no necesariamente en todos los idiomas, pero sí los suficientes para poder asegurar que este carácter se utiliza para estos idiomas seleccionados y también en referencias a los alfabetos de esos idiomas que utilizan ese carácter se incluyen en esta lista.

Si volvemos a la diapositiva anterior para los caracteres de la A a la Z, no tenemos ningún idioma o ninguno en la lista porque estos están ya incluidos por defecto. Siguiendo diapositiva.

Aquí el repertorio también es uno de los componentes principales del archivo LGR XML, el repertorio en el capítulo 5 está clasificado numéricamente por código de punto y ustedes van a poder ver los caracteres agrupados, por ejemplo, la forma de glifo, entonces ven la A agrupada con diferentes tipos de acentos y esto lo pueden encontrar en el apéndice C, es decir, los caracteres están clasificados de manera distinta.

En la sección 4 en el capítulo 5 enumeramos aquellos caracteres que se encuentran excluidos, es decir, caracteres que están enumerados y están al menos en un idioma seleccionado, pero que no se pueden incluir porque no pertenecen al MSR.

Entonces el MSR es el proceso previo que establece el límite superior de lo que se puede incluir. En el caso del GP latino y sus procedimientos, se incluyen caracteres seleccionados en el MSR, si no son parte del MSR bueno, no hay nada que el grupo o el panel de generación latino pueda cambiar, primero tiene que estar esto incluido en el MSR. Siguiente diapositiva, por favor.

Ahora voy a hablarles sobre el repertorio, en esta parte se encuentran las reglas sobre los conjuntos de variantes, que es una parte importante de la propuesta, en el capítulo 6 se abarca este tema.

Entonces, ¿qué es una variante? Una variante o un conjunto de variantes consiste en dos o más caracteres que de alguna manera se perciben como similares, como si fueran los mismos y son similares de alguna manera, pueden tener, por ejemplo, la misma forma o se pueden

usar indistintamente para alguna parte o toda la parte del código de escritura de una comunidad.

Cuando hay un conjunto de variantes no es posible utilizar ambos caracteres de manera independiente, en la mayoría de los casos, si se utiliza uno de los caracteres el otro queda bloqueado dentro del mismo contexto. Entonces para el código de escritura latino estas reglas de variantes llevan a varias variantes y bloqueos.

Dentro de los conjuntos de variantes tienen miembros del mismo código de escritura y en los conjuntos de variantes entre códigos de escrituras tienen miembros de diferentes códigos de escrituras, por ejemplo, el código de escritura latino, cirílico y griego.

El capítulo 6 junto con los apéndices D1 a D9 contienen los principios de los conjuntos de variantes, también datos y análisis sobre los conjuntos de variantes y los conjuntos de variantes que pueden ser candidatos. Esto se hace para poder ver por qué llegamos a la conclusión de que son variantes.

Hay dos conjuntos de variantes que son especiales y que están también relacionadas con la versión del IDNA, versión 2003, que también incluyen reglas que permiten asignar otras variantes y estos dos conjuntos se relacionan con la S sharp y la SS, en el otro caso tenemos la I con punto y la I sin punto.

Entonces estos dos conjuntos se consideran especiales dentro de la propuesta, pasemos a la siguiente diapositiva, por favor. Finalmente, en el apéndice E contamos con una lista de conjuntos que son visualmente confusos o que se pueden confundir visualmente, estas son variantes

que no son parte, en realidad, del archivo XML LGR formal y el apéndice C incluye referencias para quienes quieran hacer un análisis visual de la similitud entre estas dos cadenas.

Aquí vemos el proceso, nos encontramos en un proceso de comentario público en curso, esperamos enviar la versión final en el mes de enero del próximo año, así que, los invito a que lean y que hagan sus comentarios. No sé cómo vamos con las preguntas.

PITINAN KOOARMORNPATANA: ¿Hay alguna pregunta en el chat? O bien pueden levantar la mano y tomar la palabra. Bueno, viendo que no hay preguntas... Si luego surgen preguntas para el panel de generación latino, podemos volver.

Ahora quiero darle la palabra al siguiente orador, Hiro Hotta, quien es el presidente del panel de generación del código de escritura japonés. Adelante, Hiro.

HIRO HOTTA: Hola a todos, espero que me escuchen bien.

PITINAN KOOARMORNPATANA: Sí, lo escuchamos bien. Adelante, por favor.

HIRO HOTTA: Bueno, mi conexión no está muy estable el día de hoy, les pido disculpas, en algunos minutos quizás tenga que reiniciar mi sistema. Siguiendo diapositiva, por favor.

Bueno, en este caso, vemos los miembros del panel de generación del código de escritura japonés, una situación similar a lo que ocurre con otros paneles de generación, tenemos miembros de diferentes campos. Siguiendo diapositiva.

Esta es una descripción general del código de escritura japonés y el idioma japonés, creo que es bastante único en su naturaleza. El idioma japonés tiene tres códigos de escrituras, incluso con ASCII y con los números, pero estos son tres códigos de escrituras que se conocen como kanji, hiragana y katakana.

Estos caracteres se pueden mezclar en cualquier orden en una palabra y los caracteres del LGR definidos en el JIS, que es el estándar industrial japonés nivel 1 y nivel 2, se utilizan diariamente, componen el repertorio del LGR japonés.

El kanji tiene varios caracteres, tenemos aproximadamente unos 6000 y, por supuesto, esto está en el MSR, tenemos estos tres tipos de códigos de escrituras. Y el kanji también se utiliza en los idiomas chinos y coreanos, hay algunas diferencias en la definición de estas variantes que luego les voy a contar.

Además de hiragana, en el país contamos con este código de escritura y con katakana. Hiragana se utiliza para los sufijos o en kanji para poder completar una lectura completa de una palabra, por ejemplo, las conjunciones y para reescribir diferentes formas de kanji a nivel escrito o en la lectura. Por ejemplo, para decir: “Un niño de cinco años” o expresiones similares. En el caso del código katakana se utiliza para otro tipo de palabras, katakana se utiliza principalmente para

representar palabras extensas de idiomas extranjeros o lo que se denominan palabras onomatopéyicas.

Con respecto a las variantes, básicamente todos los caracteres japoneses se consideran independientes, en los idiomas coreanos, chinos y japoneses comparten algunos conjuntos de caracteres de kanji que se consideran variantes cuando dos o más caracteres tienen el mismo significado y pronunciación.

Algunas personas en la comunidad de idioma japonés creen que algunas variantes coreanas o chinas se consideran variantes también en el idioma japonés. Siguiendo diapositiva, por favor.

Bien, este es un resumen generalizado del LGR japonés, como dije, el repertorio cuenta con más de 6000 caracteres, no hay variantes intrínsecas para el idioma japonés o el código de escritura japonés para el mismo significado y pronunciación, pero para lo que es la coincidencia visual hay 12 conjuntos de variantes que se identifican.

Y este también es un punto único, es que acomodamos las variantes de kanji definidas en el LGR chino y coreano, creo que hay unos cientos de estas que hemos importado del LGR chino y coreano. Aquí hay otro punto importante, que no hay ninguna etiqueta que se pueda asignar o ninguna etiqueta de variante que se pueda asignar.

Esto es algo importante e interesante en lo que es el código de japonés, decidimos, finalmente, que no habría etiquetas de variantes. Y también hay una regla específica simple en el código de escritura japonés que es, no se debe comenzar una etiqueta con una cana pequeña o algún otro carácter. Siguiendo diapositiva.

Se debe prestar especial atención a los siguientes puntos, lo que tiene que ver con la coincidencia o similitud visual en los códigos de escrituras japoneses, en este caso, tenemos varias cuestiones a tener en cuenta. En el primer caso vemos una especie de guion o signo menos, son dos marcas.

El consorcio de Unicode enumera una lista de caracteres que se pueden confundir entre diferentes códigos de escrituras, aquí en Japón tenemos... Bueno, en el código de escritura japonés tenemos 10 pares aproximadamente, que se pueden considerar como candidatos para ser clasificados como visualmente idénticos.

Aquí vemos cuáles son estos caracteres que se muestran a la derecha en ejemplos. Pasemos a la siguiente diapositiva, por favor.

Y en cuanto a esta similitud visual es necesario confirmar si todos son visualmente idénticos o no, entonces hemos examinado aproximadamente unos 40, entre ellos, por ejemplo, japoneses, y hemos también clasificado cada uno de esos pares, tal como se indica en la diapositiva.

Y el resultado fue que todos los pares tienen un puntaje que es menor a 3.2, esto quiere decir que son lo suficientemente idénticos para ser confundidos, por lo tanto, deberían ser variantes. El siguiente campo de investigación fue determinar si hay pares idénticos visualmente adicionales, hicimos esta investigación para aproximadamente unos 176 receptores diferentes, respondieron unos 73 y el resultado fue que no hay pares confusos para el más del 3% de quienes respondieron.

No hay más de 10 pares que sean confusamente similares. Siguiendo diapositiva.

Este es un segundo tema especial, básicamente cualquier combinación de caracteres está permitida en etiquetas japonesas, con esta característica tendríamos cadenas de caracteres de variantes muy largas, en especial, las importadas de los LGR chino y coreano. Por ejemplo, este es un carácter de país para acondicionar, tiene tres cadenas de caracteres variantes.

La universidad de Keio registra y utiliza todos los SLD con variantes bajo .JP, entonces si la universidad de Keio quisiera utilizarlos todos, los cuatro, y se le permitiera, habría simultáneamente en la zona raíz cuatro TLD. No es tanto cuatro, pero una regla de este tipo a la larga haría explotar la zona raíz, entonces se necesitaba reducir el número de etiquetas de variantes asignables para impedir que explotara el tamaño de la zona raíz. Siguiendo, por favor.

Entonces el panel de generación japonesa intentó diseñar distintos métodos para reducir el número de etiquetas asignables, puede utilizarse cualquier método para etiquetas asignables, había unos cinco. Y, finalmente, decidimos que íbamos a permitir solo las etiquetas de nombres solicitados válidos y se iban a bloquear todas las etiquetas de variantes para esa etiqueta solicitada. Siguiendo.

El paso en el que nos encontramos es el procedimiento de comentarios públicos que cerrará el 16 de noviembre, así que, agradecemos vuestros comentarios. Gracias, eso es todo.

PITINAN KOOARMORNPATANA: Muchas gracias, Hotta-san, ¿alguna consulta o preguntas en el chat?

SARMAD HUSSAIN: Tenemos una pregunta, pero no está dirigida al idioma japonés, es más general, así que la vamos a retener y quizás después de la presentación de Michel incorporamos esta pregunta.

PITINAN KOOARMORNPATANA: Bueno, gracias, Sarmad, gracias, Hotta-san y pasamos ahora al panel de integración, Michel Suignard, quien se referirá a la versión 5 de las reglas de generación de etiquetas de la zona raíz. Adelante, Michel.

MICHEL SUIGNARD: Siguiendo diapositiva, por favor. En mi presentación cubriré cuatro temas, el primero es la integración, ¿cómo hicimos para hacer la integración de los códigos de escrituras chino, japonés y coreano, luego los pasos para procesar las etiquetas, cómo trabajaremos con los códigos de escrituras armenio, cirílico, griego y latino, y el alcance de la versión 5 de las RZLGR. Siguiendo, por favor.

Esto es similar a lo que nos contaba Hotta-san, una de las principales tareas que enfrentamos ahora es la integración de los códigos de escrituras chino, japonés y coreano con sus propias LGR, estos códigos de escrituras ya están completos, están los LGR. Para los LGR japonés en este momento está en período de comentario público, para coreano y chino hace un tiempo, o sea que es una tarea completada.

Son LGR muy grandes que contienen miles y miles de puntos de códigos, hay cuestiones muy complicadas que veremos ahora o cuando veamos las preguntas, hemos trabajado entonces en la integración de los conjuntos y hemos trabajado con los distintos paneles de generación para ver cómo hacerlo.

Primero nos interesa ver qué ocurre con las variantes de estos LGR que se generan de manera independiente o también se fusionan, hay algunos que se heredan, hay por lo menos dos reglas en estos códigos que han sido afectados, coreano y japonés, porque del chino son las que tienen más variantes.

Quizás lo sepan, pero es porque tenemos este sistema dual del chino tradicional y el chino simplificado, hay muchos caracteres prácticos que en el chino simplificado son reconocidos. Esto tiene que ver con razones históricas, en algunos caracteres simplificados también existen en japonés y coreano con la misma forma, y también hay variantes entre los caracteres tradicionales, hay muchos caracteres chinos que son similares a los que se utilizaban antes.

Entonces las reglas para el código de escritura chino es la que tiene más variantes porque muchos de estos conjuntos de variantes tienen por lo menos dos miembros, en el chino y en el coreano, que también tienen que definirse en las reglas. Por ejemplo, se puede pedir a las reglas de chino y de coreano que incluyan aquellas que tienen al menos dos puntos de código, que también existen como parte del conjunto de variantes de las LGR chinas. Esta es la primera parte.

Entonces es nuestro trabajo garantizar todos los conjuntos de variantes que están presentes en las variantes del chino, también estuvieron definidos en las LGR de japonés y coreano. En estos dos, japonés y coreano, ya se definieron, están todas las variantes bloqueadas y en el caso chino puede haber algunas asignables, además, las LGR chinas tienen algunas variantes no listadas en relación con las LGR heredadas, además de kana que puede generar confusión porque cuando las LGR japonesas, como vimos antes, crean conjuntos de variantes entre los caracteres kanji y kana.

Estas fueron importadas del chino porque el idioma chino muchas veces incluye caracteres que en este momento están en el kanji y algo parecido pasa con el coreano, es un poquito diferente, así que hubo algunas confusiones con los caracteres chino que tuvieron que ajustarse en las LGR chinas.

Para coreano y japonés es más simple porque importamos las que tenían, como decía, dos puntos de código, coreano y japonés tendrán que incluir variantes que tengan al menos un punto de código asociado con un caracter chino que no esté incluido en su propio conjunto.

Los resultados esperados, la colisión de etiquetas, se determinará a través del archivo de LGR fusionado, ahí vamos a determinar lo que son los elementos, es decir, la integración de las LGR propuestas por los paneles y de ahí surgirá un archivo común o fusionado que es básicamente la manera en que detectamos las colisiones entre una etiqueta solicitada o una etiqueta existente o delegada.

Necesitamos un mecanismo para detectar que la etiqueta que se está solicitando está todavía disponible para su delegación, luego las etiquetas asignables están solo en las LGR chinas, que son las que tienen variantes. En los demás están bloqueadas.

En el código de escritura japonesa están todas bloqueadas, estamos a la espera de recibir las LGR de japonés para comentar en el comentario público, pero no esperamos cuestiones significativas. Siguiendo diapositiva, por favor.

Esta diapositiva, este diagrama también es parte de la revisión de las LGR de la zona raíz, aquí lo importante es que, hay dos elementos en la presentación de una etiqueta para su delegación a la zona raíz. Tenemos la etiqueta original y la propuesta de variantes, la primera en el camino es cuando se solicita una etiqueta, ese es el foco principal, digamos, de la solicitud cuando el solicitante utiliza las LGR del elemento para solicitar una etiqueta que es válida, es decir, que está en el repertorio y que cumple con los requerimientos expresados en las LGR.

Ahí luego determinamos si es válida y luego se genera la variante de índice, que es el mecanismo que permite crear una representación única de esa etiqueta que puede compararse con etiquetas existentes que ya están delegadas en la zona raíz. Entonces si es única, si no se le usa, la aceptamos, si no es única la rechazamos y se considera otra variante de esa etiqueta ya delegada, en ese caso, básicamente la rechazamos.

La segunda parte es muy clara, al mismo tiempo, como paso opcional el solicitante también puede proponer variantes. ¿Cómo se determina? Se puede hacer a través de la propia investigación para saber cuáles son las variantes en las etiquetas solicitadas, al crear esas variantes primero hay que asegurar que es válida... Y luego asegurarse de que es asignable, que no esté bloqueada porque si no también será rechaza.

Si es asignable se convierte en una variante asignable, es decir, básicamente se le puede delegar en la zona raíz. Hay algo importante, hay que tener cuidado qué se elige como etiqueta original y como etiqueta de variante porque el mapeo entre variantes no es simétrico.

Hay distintos tipos de mapeo entre la etiqueta original y la variante para cada conjunto de miembros, entonces a veces hay que tener cuidado al elegir la etiqueta original que debe permitir la etiqueta de variante porque, por ejemplo, a veces se elige una variante que al formarse en la zona raíz con la etiqueta original no constituye una variante asignable.

Entonces no es exactamente simétrico, hay que tener cuidado qué elegir, que ambas puedan ser asignables. Hay que consultar la definición en las LGR de los tipos de mapeo para utilizar esos conjuntos de variantes. Siguiente diapositiva, por favor.

Básicamente en esta diapositiva tenemos cuatro códigos de escrituras, latino, griego, cirílico y armenio que ahora se han completado, creo que el código de escritura latino ya está para comentario público, en cuanto al griego estamos a punto de finalizarlos; si es que ya no se terminó, y para el cirílico finalizó hace un tiempo.

Una breve explicación al respecto. Las variantes que están dentro del sistema están plenamente enumeradas en cada uno de los cuatros LGR, entonces básicamente las variantes están relacionadas o son interdependientes de estos cuatro códigos de escrituras y vamos a publicar estos LGR que tienen conjuntos que son coherentes y pueden, por ejemplo, encontrar un LGR o algo contenido en un LGR que tiene correlación en otro LGR.

Nosotros hicimos una excepción para lo que se denominan las formas genéricas, que están solamente enumeradas en el LGR latino como algunos derivan, el griego deriva en otros códigos de escrituras y se pueden encontrar, pero no hemos relacionado ninguna de estas formas genéricas, así que decidimos, por lo tanto, como mecanismo de integración enumerarlos en el LGR latino cuando esté integrado.

Estos no sabemos si van a estar integrados en el repertorio, pero algunas de estas variantes van a ser nuevas y bueno, vamos a presentarles luego más información. Como dije anteriormente, también determinar el nivel de colisión de etiquetas requiere el uso de un archivo que esté unido o unificado.

También hay una situación de derivación para dos LGR para evitar ciertas incompatibilidades, como no sabemos cuáles van a ser las situaciones que se plantean de las variantes dentro de los códigos de escrituras, la situación que se da es bastante común. Por ejemplo, el LGR latino puso ciertas variantes, pero en el cirílico no se habían propuesto. Entonces básicamente se agregan estas variantes al código de escritura y al LGR cirílico, esto quiere decir que estos LGR definidos o derivados no deberían requerir ninguna propuesta actualizada.

Pero también vemos que si se requieren mapeos van a ser agregados como parte de esta integración a este LGR original. Básicamente es negociar con los miembros del GP para ver si se integran o no, y el conjunto totalmente integrado se va a presentar para revisión pública como RZLGR 5.

En cuanto al contenido del RZLGR 4 vemos que hay 18 códigos de escritura existentes, hay dos nuevos códigos o alfabetos de códigos de escrituras, tenemos el coreano y el japonés, y también vamos a agregar dos códigos de escrituras alfabéticas que serían el latino y el griego, dos que fueron previamente derivados, dos códigos de escrituras alfabéticas, el cirílico y el armenio.

También tenemos uno que está en progreso que es el de Myanmar y esperamos ver esto publicado para comentario público, es decir, el de Myanmar, en breve. Seguramente estos sean mensajes para poder brindar luego.

Y el LGR 5 anticipa que va a contener unos 25 códigos de escrituras y dos códigos de escrituras futuros, que son el thaana y el tibetano, así que, serían 27 o quizás 28 porque creo que falta alguno, alguno quedó perdido por ahí.

PITINAN KOOARMORNPATANA: Michel, también tenemos el katakana y el hiragana.

MICHEL SUIGNARD: Sí, sí, me faltaba algún par y me preguntaba cuáles eran. Esto simplemente es una diapositiva para referencia, con esto concluye mi

presentación, creo. Así que, bueno, termino, creo que ahora sí podemos dar un espacio para las preguntas, en caso de que haya preguntas.

PITINAN KOOARMORNPATANA: Muchas gracias, Michel. Vamos ahora sí a pasar a las preguntas, creo que ya todas o muchas han sido respondidas.

SARMAD HUSSAIN: Voy a leer las preguntas, algunas ya fueron respondidas por Michel en cierto punto, pero quiero darle a Michel la oportunidad de ver si quiere agregar algo más. La primera pregunta viene de Donna Austin, que dice lo siguiente: “Si esta es la conclusión, es decir, que no se permiten variantes en el LGR japonés, ¿ha sido logrado en algún otro código de escritura? Es decir, esta conclusión de que no hay o no se permiten variantes, ¿se ha logrado también en otro código de escritura?”

MICHEL SUIGNARD: A ver, debo decir que muchos de los LGR son variantes, pero la mayoría están bloqueados. Las que son asignables son una situación totalmente diferente, pero desde el punto de vista del DNS no es tan simple.

Tratamos dentro del panel de integración ver si se identificaba algún cronograma o algún plan escalable, pero en el caso del chino, por ejemplo, creo que hay un máximo de cinco que pueden existir y si uno, por ejemplo, solicita una etiqueta solamente van a quedar después cuatro variantes asignables posibles.

Creo que en el caso del código latino tenemos ciertas mitigaciones para poder limitar esto, en la mayoría de los casos se puede lograr algo. En el caso del código de escritura japonés esto presentó un gran desafío y tuvimos que trabajar con el panel de generación para poder lograr una solución, aunque seguimos buscando otra solución, trabajamos en la búsqueda de una solución que sea buena para todos porque en el peor de los casos puede haber una explosión masiva generalizada de una etiqueta que básicamente se crea. Ese sería el peor de los casos.

Y también podría ser un tema con las herramientas que se utilizan, la integración es tan complicada en muchas oportunidades que podrían no dar lugar a una solución, pero la situación es esta y por eso llegamos a la conclusión de que la solución sería, dentro de los GP, bloquear estas variantes, pero sí, existen.

Es una situación común que se da, en muchos casos se bloquean. En otros códigos de escrituras también se da esa situación, no es exclusivo del código de escritura japonés.

PITINAN KOOARMORNPATANA: Gracias, Michel. ¿Alguna pregunta adicional?

SARMAD HUSSAIN: Sí, hay otra pregunta de Robert Nkambwe, la pregunta dice lo siguiente: “¿Es este trabajo un subconjunto de UA y viceversa? Entonces, ¿es este un trabajo de un conjunto de UA y viceversa?” Veo que hay algunas respuestas en el chat, no sé si alguien quiere explayarse.

PITINAN KOOARMORNPATANA: Muchas gracias por la pregunta. Bueno, sí, en realidad está relacionado, en el UA este es el mecanismo para hacer que los caracteres que no son ASCII o los nombres de dominios internacionalizados se utilicen a nivel local, técnicamente hablando, pero ahora, ¿cómo hacer que el sistema o la aplicación o el sitio web acepte esto en internet, acepte este IDN? Bueno, aquí hablamos de UA o Aceptación Universal, pero, en este caso, no solo hablamos de un IDN, sino que también hablamos de la interconexión.

En cuanto a la Aceptación Universal es necesario garantizar que los IDN se acepten dentro de una aplicación en internet, pero también esto se aplica a los gTLD, particularmente a lo que son los nuevos gTLD cortos o extensos, si ustedes ven, por ejemplo, .NUEVOALGO ese sería el uso. Ahora tenemos, por ejemplo, nombres más extensos, como .RUN, así que, la Aceptación Universal abraza lo que serían los nombres de dominios internacionalizados con IDN y los que no, están un tanto interrelacionados.

Esto va de la mano con lo que sería la Aceptación Universal y los nombres de dominios internacionalizados.

SARMAD HUSSAIN:

Tenemos otro comentario, es un comentario que probablemente lo pueda abordar Michel y también lo hace Robert, dice lo siguiente: “¿Se espera que haya un testeo riguroso?” Eso no se vio todavía en curso, así que, no sé si Michel... Estamos quedándonos sin tiempo, pero no sé si Michel puede compartir con nosotros algunas de las pruebas que se están haciendo a nivel de LGR.

MICHEL SUIGNARD:

Sí, nosotros le pedimos a cada GP que brinde un conjunto de evaluaciones o de pruebas de etiquetas y queremos que se haga este ejercicio para los LGR que se proponen y también verificamos que los TLD existentes en la zona raíz o estén en el LGR, es decir, nosotros validamos las etiquetas existentes en la zona raíz y también lo que hacemos es verificar los LGR. Básicamente llevamos adelante este esfuerzo para cada código de escritura y, por supuesto, pueden encontrar estas pruebas online.

Los hacemos en el uso común y si encontramos desviaciones volvemos al GP y analizamos esa desviación, a veces esto se da porque el código de escritura tiene una versión o... Hay diferentes razones, incluso históricas, que pueden ocasionar esto. Y la respuesta es que, sí, hacemos ese tipo de testeos, tenemos este tipo de mecanismos y la idea es que se efectúen antes de la integración.

Básicamente para nosotros es algo mecánico, es decir, tiene que ver con mantener en forma el LGR, antes lo hacían los GP y esto es una especie de interacción que existe entre los paneles de generación y el panel de integración.

PITINAN KOOARMORNPATANA: Muchas gracias, Michel, gracias por abordar este comentario. Ya nos hemos excedido del tiempo asignado para esta sesión, así que voy a darla por finalizada. Gracias a todos por participar, ha sido una sesión sumamente interactiva, los invito a que lean los procedimientos para

para la Zona Raíz

ES

comentario público y que, quizás, puedan también brindar ciertos aportes.

Así que, con esto, voy a dar por finalizada la sesión, que tengan un muy buen resto del día y esperamos verlos la próxima semana. Vamos a dar por finalizada la llamada y vamos a detener la grabación.

[FIN DE LA TRANSCRIPCIÓN]