ICANN81 | Reunión General Anual – Avances en proyectos relativos a los IDN para el programa para nuevos gTLD: siguiente ronda
Miércoles, 13 de noviembre de 2024 – 15:00 a 16:00 TRT

PITINAN KOOARMORNPATANA: Vamos a comenzar con la sesión. Por favor, inicien la grabación. Hola y bienvenidos a la sesión sobre avances de los proyectos relacionados con los IDN para el programa de nuevos gTLD. Soy Pitinan. Soy la encargada de liderar esta sesión. Tengan en cuenta que esta sesión está siendo grabada y se rige por los estándares de comportamiento esperado y por la política antiacoso de la ICANN.

Durante la sesión solo se leerán en voz alta las preguntas y comentarios que se presentan en la ventana de preguntas y respuestas. La interpretación para esta sesión incluye inglés, francés y español. Si desean tomar la palabra durante la sesión, por favor, levanten la mano en la sala de Zoom. Cuando se les llame por su nombre, los participantes virtuales tendrán permiso para habilitar su micrófono en Zoom. Los participantes presenciales utilizarán un micrófono en mesa para hablar. Por favor, digan su nombre para los registros y el idioma en el que van a hablar si no lo hacen en inglés. Hablen a un ritmo razonable. Habiendo dicho esto le voy a dar la palabra a Sarmad Hussain. Sarmad, le doy la palabra.

SARMAD HUSSAIN: Gracias, Pitinan. Para que tengan una idea de los temas que vamos a cubrir hoy, vamos a comenzar explicando lo siguiente. Para que todos estén en tema vamos a comenzar con algunas recomendaciones que provienen del grupo del PDP para IDN y procedimientos posteriores para que sepan qué es lo que se recomienda desde el punto de vista de políticas para los IDN, específicamente en el primer nivel y también en el segundo nivel.

En función de esto, luego hablaremos acerca de algunos proyectos. Hablaremos acerca del conjunto de reglas para la generación de etiquetas de la zona raíz. Hablaremos acerca del proyecto de revisión de similitud de cadenas de caracteres y finalmente hablaremos acerca de la mejora de la herramienta para las reglas para la generación de etiquetas. Terminaremos hablando acerca de cómo se integra al sistema.

Comenzamos con un poco de información general para todos. Básicamente hay una serie de recomendaciones de políticas sobre las cuales está basado el programa para la nueva ronda de gTLD. Hay una serie de recomendaciones de políticas que forman parte del tema 24, que tiene que ver con similitud entre cadenas de caracteres, y el tema 25, que tiene que ver con los IDN como parte del informe final sobre nuevos gTLD. Informe final sobre los procedimientos posteriores para los nuevos gTLD y su proceso de desarrollo de políticas.

Esto fue cedido por recomendaciones más detalladas a través de la fase uno del informe final del proceso de desarrollo de políticas expeditivo para los IDN que se conoce como IDN EPDP fase 1 y 2. Las recomendaciones de la fase 1, la mayoría fueron ya aceptadas por la

junta en función de esta tabla de clasificación. Hay algunas recomendaciones pendientes que tienen que ver con las tarifas.

Luego el informe final sobre la fase 2 ya fue finalizado por el equipo de IDN EPDP. Incluye más detalles acerca de las registraciones de segundo nivel además de otras recomendaciones relacionadas con las guías de implementación. Esto fue aprobado en la última sesión por el consejo de la GNSO. Todavía está atravesando este proceso de adopción.

Estos son algunos de los documentos de políticas que sientan las bases para la forma en que se implementarán los IDN en la próxima ronda de nuevos gTLD. Para hablar acerca de alguna de las recomendaciones más centrales, por supuesto que hay muchas más recomendaciones y recomendaciones más detalladas en el EPDP IDN fase 1, Pueden leer todos los documentos correspondientes pero comenzamos por ejemplo con la recomendación que está en SubPro 25.2, que habla acerca del cumplimiento con las reglas para la generación de etiquetas en la zona raíz, que son necesarias para la generación de TLD y variantes de etiquetas, incluida la determinación de si una etiqueta está bloqueada o es distribuible.

Después tenemos la recomendación final 1.1 de fase 1 que dice también que el LGR de la zona raíz debe ser la única fuente para calcular las etiquetas de variantes y los valores de disposición para todos los gTLD existentes. Esto significa que las nuevas cadenas de caracteres que se van a solicitar en la próxima ronda y sus variantes, ya sea que sean distribuibles o bloqueadas, van a estar definidas por el LGR de la zona raíz, al igual que si un gTLD existente quisiera pedir

etiquetas en el nivel superior. Esto también estará determinado por el LGR de la zona raíz. También esto determinará si son bloqueadas o distribuibles.

También hubo una recomendación acerca de los TLD de un solo carácter. La recomendación de SubPro sugiere que esto podría estar permitido para los códigos ideográficos con ideogramas siempre y cuando se resuelvan los riesgos de confusión. Esto debe ser congruente con el informe de SSAC. El IDN EPDP fase 1, la recomendación final 3.17, aclara que en el contexto de códigos de escritura con ideogramas se debe cumplir con todos los criterios. Hablamos de código de escritura han y el que se utiliza en chino, japonés y coreano.

Básicamente, lo que dicen es que las recomendaciones sugerían como parte de la implementación la ICANN debe conectarse con la comunidad china, japonesa y coreana, y sus paneles de generación correspondientes para ver básicamente cuál sería el mecanismo para implementar un han de un solo carácter. En función de esto tuvimos dos periodos de comentarios públicos. En uno hubo una manifestación de los paneles de generación de chino, japonés y coreano. Se hizo un seguimiento como parte de esta recomendación en particular con estos paneles de generación. Publicaron una declaración que presentamos a comentario públicos. Desde entonces recibimos comentarios de la comunidad más amplia.

Más recientemente publicamos también el módulo IDN de la guía para el solicitante. Lo presentamos para comentarios públicos y esto también se refiere al han de un solo carácter. Está basado en la

recomendación. Recibimos comentarios públicos adicionales a través del segundo periodo de comentarios públicos que se cerró más recientemente.

Recibimos bastantes aportes sobre este tema. Lo que estamos haciendo ahora es recabar los datos, trabajar con la comunidad china, japonesa y coreana, que también nos hizo aportes sobre este tema, y luego trabajaremos con el subárea del IDN EPDP y luego lo presentaremos para fines de noviembre, principios de diciembre al panel completo para ver cuáles son los siguientes pasos con respecto a este tema.

Además, en SubPro hay una recomendación que indica que los IDN gTLD que se identifican como variantes solo pueden ser solicitados por la misma entidad. Es decir, por el mismo operador de registro. Además, las variantes deben ser administradas por el mismo proveedor de servicios de registro de backend.

Por otra parte, básicamente hay algunas recomendaciones para las registraciones de segundo nivel. La primera es la 25.6 de SubPro. Básicamente se refiere a que si una etiqueta de segundo nivel es registrada en un TLD, su TLD variante, la etiqueta debe ser registrada por el mismo registratario o debe ser retenida para ser registrada por el mismo registratario. La siguiente, por favor.

Esto se extiende un poco más en la recomendación 25.7 de SubPro que dice que no solamente la variante debe ser registrada por el mismo registratario sino que también etiquetas variantes. Las etiquetas variantes también deben ser registradas por el mismo registratario.

Finalmente, la recomendación 25.8 indica que si bien estos nombres de dominio van a ser registrados por el mismo registratario, esto no significa que tengan que actuar de la misma manera, lo cual significa que estos nombres de dominio pueden ser utilizados de diferentes maneras por el registratario y una de las razones es que una de estas variantes podría estar en un idioma y la otra en otro idioma, lo cual podría incluso atravesar regiones geográficas y podría ser utilizado de forma diferente por el mismo registratario porque podría tener diferentes públicos, por ejemplo.

Hace poco tiempo publicamos también algunos módulos preliminares relacionados con temas de IDN. Ya se cerró el periodo de comentarios públicos pero si ustedes quieren ver estas versiones preliminares hay un módulo sobre la revisión de similitud de cadenas de caracteres que ya se publicó. Hay una versión preliminar sobre los nombres de dominio internacionalizados y otra sobre las reglas para la generación de etiquetas de la zona raíz. Estos son tres módulos para la guía para el solicitante. Ya hablamos con el IRT sobre este tema y más recientemente lo publicamos para recibir comentarios públicos.

Obviamente vamos a tomar en cuenta todos los aportes del periodo de comentarios públicos y luego actualizaremos estas versiones preliminares que atravesarán un segundo proceso de comentarios públicos el próximo año. Calculo que en ese momento tendrán otra oportunidad de hacer una revisión de todo este material. Habiendo dicho esto le voy a dar la palabra a Pitinan para hablar acerca de los LGR de la zona raíz.

PITINAN KOOARMORNPATANA: Muchas gracias, Sarmad. Voy a hablar acerca del proyecto de las RZLGR. Estas RZLGR se utilizarán para determinar las variantes de cadenas de caracteres solicitadas. Estamos trabajando con la versión 5, que se publicó en mayo de 2022. Esta versión incluye los 26 códigos de escritura que vemos en la pantalla. Muchos de ustedes en la sala participaron en este proceso. Muchas gracias. Insumió más de 10.000 horas de trabajo de los voluntarios de 43-44 países, más de 270 miembros participaron.

Ahora tenemos 26 códigos de escritura pero un código de escritura puede abarcar múltiples idiomas. Tenemos más de 386 idiomas abarcados en todos estos códigos de escritura. Esperamos que sean la mayoría de los idiomas que se utilizan actualmente. Con respecto a la actualización, Veo que Hadia está pidiendo la palabra. Hadia, si me permite, voy a terminar esta sección y luego le doy la palabra.

Recientemente, el panel de generación para el código de escritura Thaana se creó en septiembre de 2024. Este panel de generación está trabajando todavía y así es como se ve su trabajo. Tienen 26 consonantes, 11 vocales. No es un gran repertorio. Todavía hay que decir qué se va a incluir y qué se excluirá. Este es un trabajo que potencialmente veremos reflejado en la versión seis de las RZLGR.

Bien. Este panel de generación se creó en septiembre. Están trabajando en la versión final de su documento. Ya están trabajando en el segundo nivel, con lo cual algunas soluciones ya están listas. Con esto se van a modificar las LGR para la zona raíz. Esperamos que la versión preliminar del documento esté lista el año próximo, que se publique para comentario público y que entonces ya la podamos

integrar a la versión 6 de las RZLGR. Esta será la base para la nueva ronda de nuevos gTLD. Ahora sí, voy a darle la palabra a Hadia. Hadia, habilite su audio, por favor. Hadia, por favor, habilite su audio en la sala de Zoom. A lo mejor puede escribir su pregunta en el chat, Hadia. Si no, la volveremos a llamar para que tome la palabra más adelante. Ahora le doy la palabra a Sarmad, quien estará a cargo de la próxima parte de la presentación.

SARMAD HUSSAIN: Muchas gracias, Pitinan.

Pedimos disculpas. Hay un inconveniente con el audio.

Hablemos de la revisión de cadenas de caracteres similares. Pasemos a la próxima diapositiva, por favor. Básicamente, el equipo de los procedimientos posteriores confirmó que no debe haber cadenas de caracteres confusamente similares que sean similares a un dominio de alto nivel ya existente o reservado. Hay que tener en cuenta la similitud visual y básicamente lo que dice el texto es que esta similitud visual no tiene que ser solo posible sino también probable para considerar que las cadenas son confusamente similares.

Se agregaron detalles a estas definiciones del equipo de los procedimientos posteriores y la primera etapa del EPDP sobre los IDN. En primer lugar se identifica el alcance de las comparaciones. Luego se indica que esta decisión la tiene que tomar un panel encargado de hacer la revisión de estas cadenas de caracteres y, por supuesto, entonces no se las utilizará en la próxima ronda. También se indica que

si bien ahora hay muchas más comparaciones, el panel de revisión de similitud entre cadenas de caracteres incluirá algunas comparaciones. Estos casos excepcionales tienen que seguir ciertas pautas también y hay que documentarlos. Aquí vemos en otras palabras las recomendaciones que les acabo de mencionar. Aquí vemos el texto extraído de los documentos de políticas.

Con respecto a la comparación de cadenas de caracteres similares, aquí vemos un resumen de todo lo que se va a comparar y con qué se lo va a comparar. Tenemos la cadena de caracteres solicitada a la derecha. Luego tenemos la cadena de caracteres solicitada, que representa un conjunto. Por ejemplo, está la cadena primaria o principal. La cadena de caracteres primaria o principal también representa a todas las cadenas de caracteres variantes y también a todas las cadenas de caracteres variantes bloqueadas que se generan a partir de esa cadenas de caracteres principal o primaria.

Con lo cual, empezamos con la cadena de caracteres principal, la primaria. Utilizamos las LGR de la zona raíz para ver qué cadenas de caracteres variantes se pueden asignar y cuáles están bloqueadas. La columna del medio que dice: “Cadenas de caracteres asignables”, no diferencia entre las cadenas de caracteres variantes solicitadas y no solicitadas sino que todas las cadenas asignables están ahí.

Luego esto se compara con un conjunto completo de cadenas de caracteres para ver dónde están las similitudes. En la primera columna a la izquierda vemos que se las va a comparar con cadenas de caracteres de gTLD existentes y con cadenas de caracteres de gTLD que han sido solicitadas en las rondas previas y que aún se las está

procesando, que aún no han sido delegadas pero que sí fueron solicitadas en rondas previas.

Además, se las va a comparar con los dominios de alto nivel con código de país o ccTLD existentes y con los ccTLD solicitados y que están en proceso. También se las va a comparar con todas las demás cadenas de caracteres solicitadas en la misma ronda y también con los nombres reservados y los nombres bloqueados. También todo dominio de dos caracteres en código ASCII reservado en virtud de la norma ISO 3166.

No solo se van a considerar estas cadenas de caracteres que son las primarias sino que también utilizando las LRG de la zona raíz van a generar cadenas de caracteres variantes asignables y bloqueadas. Por cada cadena que se solicite y que esté en las cadenas variantes asignables se hará esta comparación con todas las otras categorías de cadenas de caracteres.

En lo que respecta a la revisión de la similitud entre las cadenas de caracteres no se van a comparar las variantes bloqueadas de todas estas cadenas de caracteres como vemos en pantalla. Esto está fuera del alcance. Esa comparación está fuera del alcance. Todas las demás comparaciones que les mencioné sí se van a realizar.

Aquí tenemos algunos recuadros en color gris con la palabra sí. Esas son comparaciones entre variantes bloqueadas versus variantes asignables y otras. Aquí, el panel de revisión de similitud entre cadenas de caracteres podrá tener cierto poder de decisión para ver si hace o

no estas comparaciones siempre y cuando tengan un motivo válido y sigan las pautas correspondientes.

Todo lo que vemos en blanco en los cuadros blancos que dicen sí, esas son comparaciones que el panel de revisión sí debe hacer. Este es el alcance de la revisión de la similitud entre las cadenas de caracteres. Es un proceso en parte manual. Ustedes verán que esto va a generar muchísimas comparaciones porque no estamos comparando cadenas solicitadas versus existentes sino también sus etiquetas variantes versus las etiquetas variantes de las ya existentes. Esto va a generar miles de comparaciones. Antes de avanzar y ver cómo vamos a abordar todo esto de manera tal que se lo pueda gestionar, que sea manejable, quiero detenerme y ver si hay preguntas acerca del alcance de las comparaciones.

Bueno, si no hay preguntas voy a continuar. Desarrollamos un conjunto inicial de pautas para la revisión de la similitud entre cadenas de caracteres. Recibimos comentarios públicos. Los tuvimos en cuenta. Actualizamos el documento. Lo volvimos a publicar para comentario público por parte de la comunidad y vamos a seguir repitiendo este ciclo a medida que avancemos en este proyecto. Todas estas cadenas variantes son parte del proceso ahora. Eso significa que aumentó la cantidad de comparaciones de manera notable.

Para ayudar al panel de revisión que se encarga de revisar esta similitud sugerimos un proceso de dos pasos. En primer lugar van a poder hacer una selección previa. Por ejemplo, pueden tener una etiqueta en árabe, en código de escritura árabe, que quizá no se confunda con otra cadena o etiqueta en código de escritura han, del

idioma chino. Quizá no haga falta una revisión manual. Quizá sí haya dos etiquetas en código de escritura latino que sí están relacionadas y que pueden ser confusas. Ahí sí eso lo va a revisar un panel de expertos en similitudes de cadenas de caracteres.

En el primer paso del proceso hacemos esa selección previa. Recabamos datos que nos dan los expertos en similitud entre cadenas de caracteres y desarrollamos una herramienta que va a utilizar el formato de las LGR para la zona raíz. 7.9.4.0. Vamos a utilizar ese mecanismo y lo vamos a ampliar para incluir datos en materia de similitud.

Con esos datos vamos a poder calcular las posibilidades de que haya cadenas que generen confusión, que se puedan confundir. Esta selección previa, estos resultados se le darán al panel encargado de hacer la revisión de la similitud. La última etapa va a ser manual estará a cargo del panel revisor.

Bien. Tenemos dos oradores aquí: Michael Bauland, que hablará acerca del trabajo con respecto a la recopilación de datos por parte de los expertos, y también Julien Bernard, que hablará acerca de la herramienta para revisar la similitud entre cadenas de caracteres. Le doy la palabra a Michael para que hable sobre este tema.

MICHAEL BAULAND: Muchas gracias, Sarmad. Tal como se dijo, y tal como dijo Pitinan, la versión LGR de la zona raíz versión 5 cubre 26 códigos de escritura. Todos los repertorios de todos estos códigos de escritura deben ser

analizados. Esto nos habla de los diferentes idiomas, puntos de código y obviamente no hay una sola persona que sea capaz de conocer todos los detalles de todos los códigos de escritura y por eso se tomó la decisión de delegar la tarea de cada código de escritura en especialistas en sus respectivos códigos de escritura.

A cada especialista se le asignó un trabajo para que analice sus repertorios que básicamente están formados por cuatro elementos. Primero todos los caracteres del repertorio deben compararse con caracteres ASCII, tanto en minúscula como en mayúscula, de la A a la Z. Luego, por supuesto, tenemos las comparaciones dentro del código de escritura. Es decir, se comparan todos los elementos del repertorio entre sí. Algunos de los códigos de escritura se desarrollaron dentro de la misma familia con algunos códigos de escritura relacionados. Por eso hay que hacer otro análisis allí.

Finalmente, hay algunas comparaciones varias que nos aplican a todos los códigos de escritura si no algunos elementos en mayúscula que en el código de escritura deben compararse con los que están en minúscula, al menos en el caso de aquellos que tienen letras en mayúsculas y en minúsculas. Lo mismo con características compuestos y todos deben analizar formas simples como círculos, líneas, curvas, que podrían darse en cualquier código de escritura. Finalmente, también es necesario ver las versiones subrayadas de los diferentes elementos porque a veces en el URL los nombres de dominio aparecen con un guion bajo o un subrayado y aquí tenemos entonces algunos elementos.

En cuanto a la categorización, creamos cinco categorías. Es necesario verificar a cuál de estas categorías corresponde cada código. La categoría 1 son los homógrafos idénticos o muy parecidos y las variantes ya definidas. En las próximas diapositivas voy a entrar en mayor detalle sobre cada una de estas categorías y les voy a dar algunos ejemplos. Luego tenemos la categoría 2, que son los casos en los que hay mucha confusión porque son muy parecidos o son casi homógrafos. La categoría 3 son similares. La categoría 4 no son tan similares y la categoría 5 son bastante distintos o no similares.

En el caso de la categoría 1, de manera predeterminada todas las variantes existentes que ya forman parte del LGR de la zona raíz corresponden a esta categoría. Esto es independientemente del aspecto visual que tengan. Estas son todas las variantes. En cuanto a las otras categorías ahí sí observamos la similitud visual y no ningún otro tipo de similitud como el significado o ese tipo de cosas.

Luego, dentro de esta categoría, tenemos los casos de caracteres idénticos. La mayoría ya debería haber sido definido como variante en los distintos paneles de generación pero podría faltar alguno. Debido al alcance extendido de la revisión probablemente se introduzcan otros casos como por ejemplo debido a alguna forma en mayúscula. Si vemos la E latina y la épsilon griega, quizá haya una cierta similitud pero no son muy confundibles en minúscula. Ahora, si vemos las mayúsculas, que son 0045 y 0395, son idénticas y por lo tanto corresponden definitivamente a la categoría 1.

Otro caso sería la situación de algunos que son casi idénticos que son fácilmente confundibles y se los ve uno al lado del otro. Un ejemplo

sería el subrayado. Si tenemos la E y la E con un punto debajo, lamentablemente en esta diapositiva falta el subrayado. En la segunda parte las dos E sin punto. Con punto debajo deberían estar subrayadas. Esto dependiendo del tipo de letra a veces quizá no sea posible ver la diferencia entre la E y la E con el punto debajo porque el punto quedaría tapado por el subrayado o por el guion bajo. Este par entonces debería corresponder a la categoría 1.

Ahora vemos la categoría 2, que son aquellos que son muy confundibles. Esto significa que tienen que ser confusos para una gran mayoría de los usuarios cuando no están mirando las etiquetas una al lado de la otra. Un ejemplo sería la é con acento de arriba hacia abajo, de izquierda a derecha, y el otro al revés. Cuando los vemos uno al lado del otro vemos que son diferentes pero si estamos pensando en una sola etiqueta, por ejemplo, como [.coupé], ahí vemos que el otro es diferente pero la gran mayoría de los usuarios se confundiría y no reconocería que el acento va en el sentido contrario. Quizá esto sea más común para el francés. Sí. Hay otros ejemplos de otros códigos de escritura que son los que figuran aquí en el segundo y tercer ejemplo. También fueron incluidos en la categoría 2 por los especialistas respectivos.

Categoría 3. Aquí tenemos los casos similares que los confunde una mayoría pero no la gran mayoría de los usuarios. Aquí tenemos también algunos ejemplos que se clasificaron como casos de categoría 3. Esta es la clasificación que hicieron los especialistas de los paneles de generación. Aquí tenemos tres ejemplos. Luego tenemos otro caso

de similitud de categoría 3 que es el remplazo de letras que se usa principalmente en latín.

Tenemos un carácter con un acento diacrítico. Si lo tenemos en un nombre de dominio podríamos tender a escribirlo sin el acento diacrítico y, por lo tanto, podríamos usar solamente caracteres ASCII y no sería necesario utilizar el Punycode. Todos estos casos también deberían estar dentro de la categoría 3.

Luego pasamos a la categoría 4, que son casos que podrían resultar confusos para la minoría de los usuarios que prestan atención o quizá ni siquiera serían confusos en absoluto. Los especialistas aquí no buscan activamente estos casos porque habría que incluir una gran cantidad de caracteres pero cuando la decisión es que no son categoría 1, 2 o 3, se lo pone como categoría 4 que tiene una similitud lejana. La razón es que cuando otras personas miran los datos, por ejemplo, en comentarios públicos, ven que no se olvidaron de esta comparación sino que la analizaron y la decisión consciente fue no incluirla en la categoría 1 a 3.

Algunos ejemplos. Tenemos aquí algunos ejemplos más. Tenemos tres renglones con ejemplos de diferentes especialistas en los diferentes códigos de escritura. Finalmente llegamos a la gran mayoría de las comparaciones que corresponden a la categoría 5, que es la categoría de diferentes. Podemos tomar cualquier par de elementos aleatoriamente y casi todos estarán bien diferenciados, como vemos aquí, en este ejemplo. Definitivamente no es necesario volver a mirarlos. No es necesario tampoco escribirlos dado que esto no es importante en materia de similitud de cadena de caracteres.

¿Cuál es la situación actual en este proyecto de recopilación de datos? Para ocho códigos de escritura los especialistas están trabajando en los datos iniciales. Esto incluye principalmente códigos de escritura como han, que tienen un gran repertorio y necesitan más tiempo por lo tanto. Luego tenemos 16 códigos de escritura en cuyo caso los especialistas ya presentaron los datos y ahora se está pasando por un proceso de revisión interna. Es decir, la revisión está en curso y hay dos códigos de escritura que ya están listos. La revisión ya finalizó y los datos están listos. Julien va a explicar un poco más qué hacemos con los datos así que le voy a dar la palabra a Julien para que explique la siguiente parte de la presentación. Gracias.

JULIEN BERNARD: Muchas gracias, Michael. Soy Julien Bernard. Soy consultor de tecnologías de la información. Rápidamente voy a hablar acerca de la herramienta para revisar la similitud entre cadenas de caracteres. La herramienta va a utilizar la lista de TLD, gTLD, ccTLD, nombres reservados y bloqueados, tal como la describió Sarmad en la tabla que les mostró en su presentación.

Vamos a ver las cadenas de caracteres solicitadas y también los datos recolectados en este momento por los expertos en similitud entre cadenas de caracteres. La herramienta va a tomar todos estos datos y va a calcular si algunas de las cadenas de caracteres solicitadas, los nombres reservados o los nombres bloqueados son iguales o similares a otros. Incluyendo sus etiquetas variantes. Iba a generar un informe en el cual estarán los posibles conjuntos de cadenas controvertidas y eso se le entregará al panel. Toma la palabra Sarmad.

SARMAD HUSSAIN: Muchas gracias. Creo que un detalle que cabe mencionar aquí es el siguiente. Con respecto a los cinco niveles que nos presentó Michael, nosotros vamos a utilizar los primeros tres niveles, las primeras tres categorías para preseleccionar los datos de nuestros posibles conjuntos de cadenas de caracteres controvertidas. Las categorías 4 y 5 no serán tenidas en cuenta. Así vamos a estar en consonancia con la recomendación que solicita que identifiquemos con funciones no solo posibles sino probables desde el punto de vista visual.

Como pueden ver, el trabajo de recolección de datos y el trabajo para generar esta herramienta es un trabajo que está en curso. Empezamos el trimestre pasado. Seguimos trabajando este trimestre y esperamos tener este trabajo listo para principios del año próximo. Obviamente les vamos a ir informando de las novedades acerca del desarrollo de la herramienta y acerca de la recopilación de datos. Esperamos tener la herramienta finalizada y los ratos recopilados para junio de 2025. Esa es la fecha a la cual apuntamos. Bien. Esa es la última diapositiva de esta parte de la presentación. Ahora me voy a detener para ver si alguien tiene alguna pregunta sobre la similitud entre cadenas de caracteres.

NABIL BENAMAR: Muchas gracias, Sarmad y equipo, por esta excelente presentación. Nosotros estamos basando nuestro estudio en una definición de similitud desde el punto de vista visual. Yo me preguntaba si quizá haya un estándar o un factor científico de manera tal de que no haya

interpretaciones diferentes por parte de las personas interesadas en este tema acerca de esta confusión o similitud. Esto no es una ciencia exacta. Depende de la experiencia del usuario. Cada usuario puede incluir una etiqueta en alguna de las categorías que se mencionaron, sobre todo en las primeras tres categorías.

Por ejemplo, si vamos a la categoría 4, ahí la definición se basa en los posibles usuarios que prestan atención. ¿Cómo podemos garantizar que un usuario siempre preste atención? Eso no puede ser algo que garanticemos en cada experiencia. Me gustaría saber su punto de vista al respecto teniendo en cuenta el trabajo ya realizado en cada uno de los códigos de escritura. Gracias.

SARMAD HUSSAIN: Muchas gracias, Nabil. Bien. Creo que esto queda incluido en las pautas preliminares que publicamos con respecto a la revisión de la similitud entre cadenas de caracteres y justamente, como parte de ese proceso, en esas pautas antes de que se las comparta con la comunidad hablamos de lo siguiente. A ver qué pasa con gente que conoce el código de escritura, que está familiarizada con él, y la gente que no. ¿Qué pasa en esos casos? También está lo que usted planteó. Hay que ver si la gente está prestando atención o no a lo que está escrito. Está leyendo atentamente o alguien que no, que lee a grandes rasgos y no presta atención a los detalles.

Nosotros lo que proponemos en las pautas es estudiar las similitudes que se basan en lo siguiente. En lo que puedan hacer los lectores de ese código de escritura. También en lo que pueden descubrir las

personas que quizá no leen con tanto nivel de detalle pero que le están prestando la suficiente atención a un nombre de dominio cuando lo están leyendo.

Uno de los motivos por los cuales hicimos este planteo fue que la recomendación en materia de políticas con respecto a los PDP es muy clara e indica que no hay que tener en cuenta solo la posibilidad de una similitud sino también la probabilidad de una similitud. Estos criterios ya fueron plasmados, redactados. Son un tanto prudentes, conservadores. Obviamente, si se hace una lectura que no va en profundidad, obviamente habrá cierto nivel de posible confusión. Queríamos mantener esto a un nivel en el cual la gente preste atención y aun así se confunda. Ahí está la probabilidad de confusión.

NABIL BENAMAR: Gracias. Esto quizá esté más presente en algunos idiomas o códigos de escritura específicos. Hay categorías de códigos de escritura que estamos estudiando. Por ejemplo, el código de escritura árabe que usted conoce tiene los puntos. Nosotros utilizamos puntos. Estos puntos se pueden confundir con un símbolo más pequeño que se utiliza en algunos idiomas que utilizan el código de escritura árabe también. Esto puede ser aun más difícil para este código de escritura en particular. Este ejemplo no se aplica a los códigos de escritura latinos, por ejemplo.

SARMAD HUSSAIN: Muchas gracias, Nabil. Por eso estamos trabajando con los expertos locales que utilizan los códigos de escritura, para que nos den sus opiniones, sus puntos de vista. Soy consciente del tiempo disponible en la sesión. Quiero pasar al próximo tema. Luego de esta próxima parte de la presentación seguiremos respondiendo preguntas. Pitinan, le voy a dar la palabra. Luego seguiremos respondiendo preguntas.

PITINAN KOOARMORNPATANA: Michael quiere decir algo.

MICHAEL BAULAND: Quiero responder a lo que ha dicho Nabil. Algo muy importante a tener en cuenta es que todos los datos y el proceso que estamos desarrollando es para ayudar al equipo revisor de la similitud entre las cadenas de caracteres. Estas decisiones no son decisiones exactas sino que el equipo que revisará la similitud entre cadenas de caracteres será el que tome la decisión. Estamos tratar de facilitarles el trabajo. Simplemente eso.

PITINAN KOOARMORNPATANA: Muchas gracias. El próximo tema es la herramienta para las reglas para la generación de etiquetas y su mejora correspondiente. A esta herramienta la llamamos herramienta LGR. Esta herramienta tiene tres modos. Básico, en el cual verificamos si la etiqueta que nos interesa es válida según las LGR o las RZLGR. Este es el modelo básico. Luego tiene otra funcionalidad mediante la cual podemos revisar las tablas de IDN. La pueden utilizar los operadores de registros y la

pueden comparar con las RZLGR o sus LGR de referencia. Luego está el modo avanzado. Los usuarios desarrollan las LGR. Pueden generar nuevas LGR. Hacer comparaciones, modificaciones, etc.

Este es el estado actual de la herramienta. El enlace es lgrtool.com. Pueden acceder a esta herramienta si tienen una cuenta en el sitio web de la ICANN también. Esta herramienta utiliza código abierto y todo el código abierto está disponible también a través del enlace que ven en pantalla. Hay un manual disponible.

De cara a la próxima ronda, queremos agregarle una funcionalidad a esta herramienta, que puedan ver en primer lugar la etiqueta incluida en código ASCII y en formato IDN. Tenemos que ver si es una etiqueta A o si está en código ASCII. Se la compara con el requisito de protocolo. También hay que ver si cumple con las normas IDN-A 2008 y también si la cadena de caracteres es válida según las LGR de la zona raíz.

Como se dijo, quizá tengamos un requisito diferente con respecto a la longitud de las cadenas de caracteres en código de escritura han. Lo estamos verificando. También vamos a verificar si las etiquetas primarias colisionan con TLD existentes o sus conjuntos de variantes. Vamos a verificar los nombres reservados y sus variantes también. Esta herramienta tiene una funcionalidad específica separada también porque en algunos casos las cadenas de caracteres podrán avanzar.

Por último, vemos qué pasa con los nombres bloqueados y sus variantes. Esa es la funcionalidad para las etiquetas primarias. También vamos a tener una función para verificar las etiquetas variantes. Hará las mismas verificaciones que en el caso de las

etiquetas primarias y se determinará si es una etiqueta asignable. Van a tener que indicar la etiqueta primaria con la cual quieren hacer la comparación. Tenemos planificado mejorar esta herramienta antes de la próxima ronda y estará disponible para que la pueda utilizar la comunidad. Con eso termino esta parte de la presentación.

SARMAD HUSSAIN: Con respecto a los datos sobre similitud entre cadenas de caracteres, una vez que tengamos los datos y hagamos la revisión interna para ver que esté todo completo, que no falte ningún detalle, queremos publicar estos datos para que los pueda revisar la comunidad. Ustedes van a ver esos datos y con todo gusto recibiremos sus comentarios o aportes con respecto al análisis que hayan realizado los expertos en códigos de escritura para ver si están de acuerdo o no con la categorización que están realizando los expertos en códigos de escritura y también si están de acuerdo con el nivel de similitud. Todo esto lo publicaremos a la brevedad y esperamos recibir sus comentarios sobre el proceso para poder finalizar esta etapa.

Estos datos sobre similitud también están a nivel de caracteres o secuencias de caracteres. A veces una etiqueta quizá no queda incluida a nivel de secuencia de caracteres. Esto lo tendrá en cuenta el panel que revisará la similitud entre cadenas de caracteres. Tenemos una herramienta de selección previa que tiene un alcance limitado y que genera casos potenciales para que el panel revisor los tenga en cuenta y los analice en mayor profundidad. La decisión final se hará de manera manual y la realizará el panel revisor. Tenemos tiempo para una o dos preguntas. Anil, tiene la palabra.

ANIL KUMAR JAIN: Muchas gracias, Sarmad. Tengo una pregunta para Pitinan. La demostración que usted hizo es maravillosa y será de mucha utilidad para la próxima ronda de los nuevos gTLD. Me pregunto si esta herramienta está disponible para ccTLD y gTLD o solo para los gTLD. Por otra parte, si un registro no puede discernir la prueba de las LGR, ¿hay algún tipo de ayuda que vaya a ofrecer la organización de la ICANN para ese operador de registro, para que pueda corroborar su LGR? Gracias.

SARMAD HUSSAIN: Gracias, Anil. En la política IDN ccPDP4, esa política está en curso. Una vez que esa política pase a la etapa de implementación seguramente podremos diseñar esa herramienta si es de utilidad para la comunidad de los ccTLD. Actualmente esta herramienta se focaliza principalmente en la próxima ronda de nuevos gTLD.

PITINAN KOOARMORNPATANA: Quiero agregar algo. Se van a incluir en el caso de los TLD existentes, tanto los gTLD como los ccTLD.

SARMAD HUSSAIN: Anil, usted hizo una segunda pregunta.

ANIL KUMAR JAIN: Si un operador de registro no puede darse cuenta de cómo verificar las LGR antes de solicitar un nuevo gTLD, ¿hay algún plan de ayudarlo de alguna manera por parte de la ICANN ante esa dificultad? Gracias.

SARMAD HUSSAIN: Creo que sí, que habrá una ayuda disponible en la próxima ronda de nuevos gTLD a través de la ICANN. Esta ayuda estará disponible en un canal en el cual también se pueden hacer preguntas acerca de la herramienta. La respuesta es sí. Esto lo va a realizar el equipo a cargo del programa de la próxima ronda de nuevos gTLD.

SARMAD HUSSAIN: ¿Alguien más tiene alguna pregunta o algún comentario?

ASHLEY ROBERTS: Hola. Soy Ashley Roberts. Estaba pensando en esto desde el punto de vista de los solicitantes de gTLD. Obviamente este es un tema complejo. Estaba pensando en cómo haría esto un solicitante para determinar si tiene problemas de similitud de códigos de escritura. Venimos hablando de la similitud de cadenas de caracteres que ustedes vienen desarrollando. ¿Cómo sería el resultado de la herramienta que ustedes están pensando en desarrollar?

SARMAD HUSSAIN: ¿Podemos volver a la diapositiva que mostró Julien? Julien, ¿quiere responder usted la pregunta? Entiendo que la pregunta es cuál sería el resultado de esa herramienta.

JULIEN BERNARD: Este sería un conjunto de solicitudes controvertidas. Se solicita un TLD y hay un TLD similar ya existente. Esto podría generar un cierto grado de confusión. Esto es lo que debe analizar el panel de revisión.

ASHLEY ROBERTS: ¿Esta herramienta sería para el panel de revisión o también estaría disponible para los solicitantes?

SARMAD HUSSAIN: Ahora estamos desarrollando la herramienta para el panel de revisión pero creo que si la comunidad considera que podría ser útil, podríamos ver si también se puede poner a disposición del público en general. El foco actual, por supuesto, consiste en ayudar a hacer el preanálisis para el panel de revisión.

ASHLEY ROBERTS: Debido a la complejidad de este tema, cualquier orientación adicional que ustedes puedan proporcionar sería fantástico. Se lo agradeceríamos porque hemos visto ya esta información, el documento de guía y a mí personalmente me resulta difícil entenderlo. Es un documento difícil. Me preocupa un poco cómo los solicitantes, especialmente gente nueva en el espacio, podrá entender esto y atravesar este proceso de forma eficiente.

SARMAD HUSSAIN: Sin duda podemos llevar su comentario y ver cómo podemos poner estas herramientas a disposición de todos. Ya nos pasamos del tiempo. Vamos a tener que dar por cerrada la sesión. Gracias a todos por participar. Si tienen alguna otra pregunta, vamos a estar aquí para responderlas después de la sesión y si no, envíen sus consultas al programa de idn@icann.org y con todo gusto vamos a responder las preguntas por mail también. Vamos a estar disponibles por aquí en caso de que tengan alguna pregunta adicional. Podemos ahora cerrar la sesión.

[FIN DE LA TRANSCRIPCIÓN]