FR

ICANN81 | Réunion générale annuelle– Progrès sur les projets liés aux IDN pour la prochaine série du programme des nouveaux gTLD
Mercredi 13 novembre 2024 – 15h00 à 16h00 TRT

PITINIAN KOOARMORNPATANA : Bonjour et bienvenue à cette session sur le progrès réalisé sur les projets liés aux IDN pour le nouveau programme gTLD. Cette session est enregistrée et est régie par les normes de comportement attendues de l'ICANN ainsi que la politique anti-harcèlement de la communauté.

Lors de cette session, les questions et réponses ne seront lues à voix haute que si elles sont soumises dans la fenêtre questions/réponses. Cette session sera interprétée en français, anglais et espagnol. Si vous souhaitez vous exprimer, veuillez lever la main dans Zoom, et pour les participants à distance, ils pourront prendre la parole quand on appellera leur nom en activant leur micro, les participants en présence prendront un micro physique pour s'exprimer.

Donnez votre nom et la langue dans laquelle vous allez parler si c'est une autre langue que l'anglais. Parlez clairement et à un rythme raisonnable.

Sur ce, je donne la parole à Sarmad Hussain.

SARMAD HUSSAIN: Merci. Diapo suivante. Alors nous désirons vous donner un aperçu de ce que nous allons aborder aujourd'hui.

Nous allons commencer, pour que tout le monde soit sur la même page, par des recommandations des SubPro, donc des procédures pour des séries ultérieures de nouveaux gTLD et l'EPD, processus accéléré d'élaboration de politique sur les IDN, noms de domaine internationalisés, notamment de premier niveau mais aussi de second niveau. Et, par la suite, nous allons partager avec vous les progrès réalisés dans certains de nos projets, nous allons notamment parler des règles de génération d'étiquette pour la zone racine ainsi que le projet d'examen de similarité de chaines et nous allons aussi parler de l'outil d'amélioration des règles de génération d'étiquettes qui va bientôt être intégré dans notre système d'application.

Commençons par quelques informations de contexte. Diapo suivante s'il vous plait.

Il existe différentes recommandations politiques sur base desquelles la mise en œuvre de l'IDN est basée pour la nouvelle série du programme des gTLD. Les recommandations, notamment pour le sujet 24, sur la similarité 25 pour les IDN, et tout ceci se trouve dans le rapport final sur les nouveaux gTLD ainsi que les procédures pour les séries ultérieures de nouveaux gTLD.

Ensuite, nous avons émis des recommandations plus détaillées, grâce à la phase 1 du rapport final sur le processus accéléré du rapport politique sur les IDN et bien sûr les procédures pour les procédures ultérieures des nouveaux gTLD, on fait référence à cela sous le nom de SubPro et EPD IDN. Ensuite nous sommes passés à la phase 2.

Et, sur base de notre fiche de suivi, nous savons qu'il y a encore quelques recommandations qui sont en suspend et liées notamment aux coûts. La

phase 2 du rapport final a été terminée par l'équipe EPDP IDN. Certaines recommandations qui se trouvent dans les lignes directrices ont été modifiées et ceci a été approuvé récemment lors de la dernière réunion du conseil de la GNSO.

Bientôt, nous allons passer au processus d'adoption.

Voilà certains des documents politiques qui représentent la base pour la mise en œuvre des IDN lors de la nouvelle série de gTLD.

Alors, par rapport aux recommandations fondamentales, bien sûr les recommandations sont plus détaillées dans la phase 1 du rapport sur les EPDP, je vous recommande de consulter ces documents ainsi que les documents des SubPro.

Mais commençons par la recommandation 25.2 qui porte notamment sur la conformité aux règles de génération d'étiquette dans la zone racine qui sont nécessaire pour la génération de TLD et d'étiquettes de variantes ainsi que pour déterminer si l'étiquette est bloquée ou peut être allouée.

Ensuite, nous avons la recommandation 1.1 de la phase une du rapport IDN EPDP, les RZLGR seront la source unique permettant de calculer les étiquettes de variante et les valeurs de disposition pour les gTLD existants.

Cela signifie que lors de la prochaine série de gTLD nous aurons des chaines, des variantes et il s'agira de savoir si ces variantes sont bloquées ou peuvent être allouées, cela sera basé sur les règles de génération d'étiquettes, notamment au premier niveau. Et tout ceci sera déterminé grâce aux règles de génération d'étiquettes, les LGR, pour savoir si ces étiquettes sont bloquées ou si elles peuvent être allouées.

Diapo suivante.

Nous avons aussi la recommandation sur les gTLD à caractère unique, recommandation SubPro 25.4 qui suggérait que cela pourrait être permis pour les écritures idéographiques et, du moins, pour autant qu'il n'y a pas de risque de confusion, ce que vérifierait le SSAC, par exemple.

La recommandation 3.17 EPDP IDN pour les écritures idéographiques. Actuellement, la seule écriture qui répond aux critères est le script Han, du chinois, japonais et du coréen. Et cette recommandation suggère que dans le cadre de la mise en œuvre l'ICANN va s'adresser au panel de génération, aux communautés chinoises, japonaises et coréennes, à leur panel de génération d'écriture afin qu'ils se penchent sur les mécanismes qui nous permettraient d'utiliser le Han à caractère unique.

Sur base de cela, nous avons reçu deux commentaires publics. Tout d'abord une déclaration de la part du panel de génération d'écriture chinoise, japonaise et coréenne dans le cadre du suivi de l'application de cette recommandation. Ils ont publié un communiqué, une déclaration que nous avons considéré comme un commentaire public. Et puis nous avons aussi reçu un commentaire de leur communauté.

Et, récemment, nous avons publié le module IDN pour le guide de candidature, l'AGB.

Nous avons reçu d'autres commentaires publics, un deuxième commentaire auquel nous avons pu mettre un point final récemment.

Nous avons aussi reçu beaucoup d'informations à ce sujet. Et, actuellement, nous sommes en train d'organiser les informations des

communautés chinoises et d'autres communautés, de plus grandes communautés, qui nous ont fourni beaucoup d'informations à ce sujet. Nous avons l'intention de soumettre ceci à l'IRT, l'équipe de révision de la mise en œuvre, pour l'équipe de l'EPDP IDN et puis de parler avec eux en décembre et novembre pour les prochaines étapes à suivre.

Il y a aussi la recommandation SubPro 25.5 selon laquelle les gTLD IDN identifié comme étant des variantes de gTLD existants ne peuvent être utilisés que pour la même entité, donc le même opérateur de registre. Outre cela, les variantes devraient être gérées par le même fournisseur de registre back-end.

Diapo suivante.

Aussi, certaines recommandations passent aussi en revue les enregistrements de second niveau, la première recommandation de ce genre est la 25.6 selon laquelle il est dit que si une étiquette de second niveau est un gTLD enregistré et qu'il y a une variante TLD, alors elle devrait être allouée à la même entité, au même titulaire de nom, ou retenue dans l'éventualité qu'elle soit un jour allouée à cette entité.

Diapo suivante.

Ceci est davantage développé dans la recommandation 25.7 selon laquelle les mêmes étiquettes pour les variantes gTLD devraient être utilisées pour le même titulaire de nom au second niveau aussi. C'est-à-dire que ces étiquettes ne devraient être utilisées que pour le même titulaire de nom.

Enfin, la recommandation 25.8 qui dit que même si ces noms de domaine sont enregistrés par un même titulaire de nom, cela ne signifie pas que leur

comportement doit être le même. C'est-à-dire que ces noms de domaine peuvent être utilisés de différentes façons, le titulaire de nom choisira leur utilisation. Une de ces variantes par exemple pourrait dans une langue, une autre dans une autre langue. Cela peut-être même pourrait transcender les frontières et être utilisé de manière différente par le titulaire de nom, car cela dessert des communautés différentes.

Récemment, nous avons également publié des projets de modules liés aux IDN, nous avons mis fin à la période de commentaires publics sur ces projets de modules, mais si vous désirez les consulter, il en existe un concernant l'examen de similarité de chaine qui a été publié, un projet de module sur les noms de domaine internationalisés, et un module sur les règles de génération d'étiquette sur la zone racine. Donc trois modules à l'état de projet dont nous avons déjà parlé avec l'équipe de révision de la mise en œuvre. Et, bien sûr, grâce au processus de commentaires publics, nous avons reçu des commentaires dont nous allons tenir compte pour mettre ces projets à jour puis passer au deuxième processus de commentaires publics plus tard dans l'année.

Et, à ce moment-là, j'imagine que vous aurez encore la possibilité de passer ces documents en revue.

Je passe la parole à Pitinan qui va vous parler des LGR.

PITINIAN KOOARMORNPATANA : Merci, Sarmad. Nous allons faire une mise à jour sur le projet des règles de génération d'étiquette pour la zone racine. Nous allons utiliser cette version pour déterminer les candidatures vers les chaines. La version courante que nous avons est la version 5 qui a été publiée en 2022. Pour

l'instant nous incluons 26 scriptes et beaucoup d'entre vous ont été impliqués, je vous remercie de votre travail.

Et ce travail implique plus de 10 000 heures de travail de bénévolat à travers 43 pays et plus de 200 membres. Nous avons 26 scriptes, mais un script peut couvrir plusieurs langues.

Nous avons des documents qui ont été traduits en 3876 langues. Nous espérons pouvoir couvrir la majorité des langues.

Lorsqu'il s'agit des mises à jour, je vois qu'il y a une main levée de la part de Hadia ? Peut-être que je vais finir ma présentation d'abord.

Récemment nous avons le panel de génération de script Thaana, formé en septembre 2024. Avec ce script Thaana, nous avons du travail qui est en cours et voilà la mise à jour. Il y a 25 consonnes, 11 voyelles, ce n'est pas un très grand répertoire, à savoir ce qui va être inclus ou non dans la discussion. Donc le travail est toujours en cours dans cette version 6.

Quand il s'agit de l'échéancier pour ce script et pour le panel de génération, depuis sa création en septembre, le groupe a travaillé sur la finalisation du premier jet, nous avons déjà du travail qui est fait au deuxième niveau, nous avons déjà des solutions. Nous avons des modifications qui vont être faites. Probablement d'ici janvier, ce premier jet sera publié pour consultation publique. Et donc, à la fin du premier trimestre, cela devrait être finalisé. Ensuite, d'ici le deuxième trimestre intégré dans la zone racine. Donc cela deviendra la base pour la prochaine série de gTLD.

Maintenant, je vais répondre à la question de Hadia. Vous voulez prendre la parole ? Vous voulez bien allumer votre micro, s'il vous plait ? Très bien,

donc Hadia, si vous pouviez inscrire votre question dans le chat, nous y reviendrons plus tard. Nous allons donc continuer avec la prochaine partie de notre présentation. Sarmad ?

SARMAD HUSSAIN: Merci. Nous allons parler de la similarité des chaines et de la révision. Prochaine diapo, s'il vous plait.

En fait, les SubPro ont réaffirmé que les chaines ne doivent pas être similaires ni porter à confusion pour ce qui est des noms de domaine de premier niveau existants et ceux qui sont réservés.

Donc la similarité doit être visuelle. Dans ce sens, le langage explique que la similarité ne doit pas être seulement possible, mais probable.

Il y a eu d'autres détails qui ont été rajoutés à cette déclaration dans le SubPro, nous sommes dans la phase 1 de l'EPDP IDN, il y a eu plusieurs recommandations finales. Aussi, le travail doit être fait manuellement par un panel de révision. Il y avait un outil qui a été utilisé par le passé, il ne sera pas utilisé pour la prochaine série, on sait qu'il y a beaucoup de comparaisons que l'on peut faire avec des similarités de chaines et donc on peut utiliser ces nouvelles directives pour éviter ou inclure ces comparaisons.

Il y aura donc de nouvelles directives à suivre.

Voilà donc comment nous avons expliqué cette recommandation et cette description est dans le document.

Lorsqu'il s'agit de la comparaison de la similarité des chaines, voilà un résumé du travail et des comparaisons qui seront faites. Vous devez porter

votre candidature pour une chaine de TLD et cette candidature va représenter un ensemble, la chaine gTLD principale pour laquelle vous portez votre candidature. Cela inclut une chaine de variantes qui vont être allouables, et cela inclut aussi des chaines de variantes qui vont être bloquées. Leur chaine principale va représenter votre point de base, de départ, et nous allons utiliser cette chaine principale et la règle de génération d'étiquette de la zone racine et cela fonctionnera de cette manière.

Veuillez noter que vous voyez cette colonne pour tout ce qui est des chaines de variante qui sont attribuable qui ne distingue pas entre les chaines de variantes attribuables et les chaines pour lesquels il n'y a pas de candidature.

Ensuite, nous allons comparer cela avec tout un ensemble de chaines afin de voir s'il y a des similarités. Si vous regardez la colonne à gauche, voilà la comparaison avec les chaines existantes. Ensuite il y aura comparaison avec les chaines gTLD pour lesquelles il y a eu des candidatures durant la première série et qui sont toujours en cours, pas encore déléguées, mais déjà candidatées. Ensuite, elles seront comparées avec les ccTLD existants, ensuite avec les ccTLD IDN qui ont été demandés, si la candidature est encore en cours. Elles seront aussi comparée avec les mêmes demandes dans la même série, avec les noms réservés, les noms bloqués, et bien sûr toutes les chaines de deux caractères ASCII.

Et tout cela est conforme avec l'ISO 3166.

Donc toutes ces chaines seront considérées comme des chaines principales, primaires. Et en utilisant la LGR de la zone racine, ces chaines

primaires vont générer des chaines de variantes attribuables et ensuite des chaines de variantes bloquées.

Ensuite, pour chaque chaine pour laquelle il y a une candidature, qui sont attribuables, il y aura des comparaisons avec toute sorte d'autres chaines, les chaines primaires, attribuables et pour les variantes bloquées. Les seules comparaisons non faites seront celles des variantes bloquées, comme vous le voyez sur la colonne de droite.

Alors, ça, ça sort du champ, donc il n'y aura pas de comparaison faite. D'autres comparaisons seront faites par la suite. Vous voyez qu'il y a des zones grises où vous avez une réponse oui. Ces blocs en gris sont les comparaisons entre les blocs de variantes et les chaines attribuables ainsi que les variantes du niveau primaire. Il y a aussi des cas où ces comparaisons ne seront pas faites. Bien sûr, il faut de bonnes raisons pour ce faire et nous suivrons les directives.

Et, comme vous voyez, dans les cases en blanc, voilà les comparaisons où le panel de révision fera son travail pour déterminer quelles sont les similarités des chaines.

Voilà le champ de travail pour cette révision de similarité des chaines. Une partie de ce travail est manuel. Vous allez voir que cela va créer toute sorte de comparaison, car on ne va pas juste comparer les chaines existantes avec celles qui existent déjà, celles qui vont être proposées. Cela va créer des milliers de comparaisons.

Avant d'avancer et de vous montrer comment nous allons faire le travail, savoir comment tout ce travail va être géré, je vais m'arrêter et demander si vous avez des questions sur tout ce champ de comparaison.

S'il n'y a pas de question, continuons.

En fait, nous avons élaboré un ensemble de directives pour cette révision de similarité des chaines, nous avons mis en œuvre un processus de consultation publique, nous avons reçu et intégré des commentaires dans nos directives, nous avons rapporté cela à la communauté et nous allons utiliser ces grandes lignes pour le développement de notre projet.

Mais, bien sûr, toutes ces chaines de variantes font partie, maintenant, du processus. Le nombre de comparaisons a augmenté énormément.

Pour appuyer le panel de révision de similarité des chaines, afin qu'il puisse faire son travail de comparaison, nous avons suggéré un processus en deux étapes. La première étape permettra de faire une vérification précoce, par exemple il pourrait y avoir une étiquette en arabe qui ne va pas porter confusion avec d'autres scriptes, par exemple en chinois. Dans ce cas-là, il n'y a pas de révision manuelle, mais il pourrait y avoir deux révisions de scripts latins qui pourraient poser confusion, que ce soit du latin ou du cyrillique, qui pourraient être comparés ou liés. Dans ce cas, il nous faut faire une révision avec notre panel d'experts pour savoir s'il y a une similarité.

Donc durant notre première étape, nous faisons un screening et nous rassemblons les données sur certaines similarités de chaines. Et nous rentrons un peu plus dans les détails.

Nous avons aussi élaboré un outil qui va nous permettre d'utiliser un mécanisme qui va examiner cette règle de génération d'étiquette sous le format RFC. C'est le format 7/9/4/0. Nous allons utiliser le même sorte de mécanisme pour étudier les similarités des données. Donc cet outil et ces données vont pouvoir être calculés et nous allons pouvoir déterminer s'il y a des étiquettes qui portent à confusion. Donc ce screening sera fait en amont.

Et, encore une fois, la dernière révision sera faite de manière manuelle et nous pourrons ainsi gérer tout cela.

Très bien. Nous avons deux intervenants ici avec nous, Mickaël Bauland, de NIM, qui va nous parler du travail qui est fait pour ce qui est de la collecte des données. Nous avons aussi Julian Bernard qui va nous parler un peu plus de cet outil que nous allons utiliser pour étudier et réviser la similarité des chaines.

Donc je vais passer la parole à Mickaël qui va nous parler de la collecte des données.

MICKAËL BAULAND: Merci beaucoup. Comme Pitinan l'a dit déjà, les règles de génération d'étiquette pour la zone racine couvrent 26 scriptes et les répertoires de ces scriptes doivent être analysés. Cela couvre un bon nombre de langages et, de toute évidence, il n'y a pas une seule personne capable de connaitre les détails de tous ces scriptes.

C'est la raison pour laquelle on a choisi de déléguer cette tâche à des experts pour chacun de ces scriptes. Et chacun de ces experts a reçu un mandat, en quelques sortes, pour analyser leur répertoire qui est composé

de 4 éléments, de 4 volets. Tout d'abord, tous les caractères du répertoire doivent être comparés aux caractères ASCII, les majuscules et les minuscules, de A à Z. Ensuite il y a les comparaisons de variantes dans un même scripte ou une même écriture, il faut donc comparer tous les éléments d'un même répertoire entre eux.

Il y a aussi des familles de scriptes liés, il y a aussi des comparaisons de variantes communes à différents scriptes qui doivent être effectuées. Et puis, enfin, des comparaisons plus diverses, plus variées, qui ne s'appliquent pas à tous les scriptes, mais qui, par exemple, exigent que les majuscules dans certaines écritures doivent être comparées par rapport aux minuscules ; il en va de même pour les caractères composés ou conjoints. Pour d'autres scripts, on doit aussi se pencher sur certaines formes simples, par exemple des cercles ou des lignes qu'on peut retrouver dans n'importe quel scripte. Et puis il y a la version soulignée de ces différents caractères, parce que très souvent, les noms de domaine sous forme d'URL sont affichés avec un élément souligné.

Diapo suivante.

Pour cette catégorisation, nous avons créé 5 catégories. Et donc, bien sûr, il faut vérifier à quelle catégorie appartient chaque script. Première catégorie, elle concerne les cas identiques ou presque identiques et qui contient déjà toute une série de variantes définies. Dans la suite de ma présentation, je passerai chaque catégorie plus en détail et je donnerai des exemples.

Nous avons ensuite la catégorie 2 avec les caractères qui sont très fortement similaires ou qui sont presque des homographes. Ensuite, catégorie 3, similaire, catégorie 4 : peu similaires. Catégorie 5 : distincts.

S'agissant de la première catégorie, par défaut, toutes les variantes existantes qui sont déjà incluses dans les règles de génération d'étiquettes pour la zone racine sont dans cette catégorie, quelle que soit leur apparence visuelle. Puis, pour la suite de la catégorisation, on ne va se pencher que sur la similarité visuelle et aucun autre type de similarité, par exemple la signification.

Au sein de cette catégorie, on a le cas de caractères identiques. La plupart de ces caractères identiques, à mon avis, devraient déjà avoir été définis. Mais peut-être qu'il y en a qui manquent.

Étant donné l'ampleur de cette révision, il est probable que d'autres caractères soient introduits, par exemple sous la forme de majuscule, si on se penche sur la lettre E et sur la lettre grecque Epsilon, quand elles sont sous la forme majuscule – contrairement quand elles sont sous minuscules où on parvient à les distinguer – elles sont identiques, comme vous le voyez à l'écran dans cet exemple. Aussi, il est clair que ces caractères appartiennent à la catégorie 1.

Nous avons aussi le cas de caractères presque identiques, qu'il est facile de confondre, et ce même s'ils se trouvent l'un à côté de l'autre, par exemple dans le cas de caractères soulignés. Ici à l'écran on a ici un E et un E en dessous duquel il y a un point. Et malheureusement ici on n'a pas montré le soulignage dans la suite de l'exemple. Mais cela devrait être visible. Et puis aussi, si on utilise une police différente, peut-être qu'on ne va pas voir

ce point en dessous, et notamment si le caractère est souligné, on ne verra pas ce point sous le E. Donc cette paire de caractères devrait également se trouver dans la catégorie 1.

Ensuite, on a la catégorie 2 avec les caractères qu'il est fort probable de confondre. Cela signifie que ces caractères sont confondus par une grande majorité d'utilisateurs attentifs lorsqu'ils ne regardent pas ces lettres ou caractères placés à côté ; par exemple un E avec un accent grave versus un E avec un accent aigu, donc la barre qui va du bas vers le haut. Quand on les met les uns à côté des autres, on constate la différence, mais si on a une étiquette précise en tête, par exemple point coupé et qu'on a point coupè et point coupé, il est fort possible qu'une grande partie des utilisateurs confonde ces deux versions et ne reconnaisse pas que l'accent part dans une direction différente. Effectivement, en français, ce serait pertinent.

Nous avons aussi d'autres exemples, d'autres scriptes que vous voyez. Il s'agit du deuxième et du troisième exemple à l'écran qui ont également été classés dans la catégorie 2 par les experts.

Diapo suivante.

Nous en arrivons à la catégorie 3 qui représente les cas de caractères similaires et qui peuvent prêter à confusion par une majorité simple d'utilisateurs attentifs. Ici, vous avez aussi des exemples classés comme appartenant à cette catégorie par les experts respectifs de chaque script représenté à l'écran. Vous en avez trois. Il y a aussi d'autres cas de similarité de catégorie trois et donc c'est, par exemple, les solutions de repli, disons les remplacements, notamment pour le cas de l'alphabet latin, car souvent si vous avez un caractère d'un scripte diacritique, lorsqu'on écrit un nom

de domaine, très souvent en fait on va l'écrire sous une forme non diacritique pour utiliser plutôt les caractères ASCII, de façon à ce que l'algorithme du code ne les convertisse pas en version XN tiret/tiret.

Diapo suivante.

Catégorie 4. Ce sont des cas confondus par une minorité d'utilisateurs attentifs ou qui ne seraient peut-être même pas confondus. Les experts ne recherchent pas spécifiquement ce genre de cas, car cela concerne probablement une grande quantité de caractères. Mais, très souvent, lorsqu'on s'est penché sur une paire de caractères et qu'il a été décidé que cette paire n'appartenait pas à la catégorie 1, 2 ou 3, mais qu'on décide tout de même de les inscrire, alors on les place dans cette catégorie 4. C'est-à-dire qu'ils sont peu similaires et cela signifie que lorsque d'autres personnes par exemple dans le cadre du processus de consultation publique se penchent sur cette paire, ils peuvent être rassurés, cette paire a été examinée et la décision de ne pas l'inclure dans les catégories 1, 2 ou 3 était bien une décision consciente. Et vous avez trois exemples ici à l'écran fournis par différents experts.

Et, enfin, la grande majorité des comparaisons appartiennent à la catégorie 5, caractères distincts. Il s'agit là de paires d'éléments pris un peu au hasard et, bien sûr, tous ces caractères sont bien distincts, comme vous le voyez ici à l'écran. Il n'est pas nécessaire d'examiner ces paires ni de les inscrire dans notre base de données, parce que cela n'importe pas pour notre examen de similarité des chaines.

Alors où on en est dans le cadre de ce projet de collecte de données ? Les experts sont encore en train de travailler sur les données pour 8 scripts. Par

exemple pour le Han, le chinois, dont le répertoire est immense et requiert donc plus de temps pour se pencher sur tous le cas. Ensuite, pour 16 autres scripts, les experts ont déjà soumis des données et, à l'heure actuelle, ces données sont examinées dans le cadre d'un processus de révision en interne et, pour deux autres scriptes, les données sont prêtes, l'examen a été effectué.

Alors qu'allons-nous faire avec ces données ? Je pense que Julien sera mieux à même que moi de vous expliquer tout ceci, aussi, je vais lui donner la parole pour la partie prochaine de cette présentation.

JULIEN BERNARD: Merci. Je suis consultant IT et je vais vous parler un peu de cet outil de similarité de chaines. Cet outil prend en compte la liste des TLD, gTLD, ccTLD existants, les noms réservés et les noms bloqués, comme Sarmad l'a mentionné. Il y a également la liste des chaines qui font l'objet d'une candidature ainsi que la liste des données collectées pour la similarité de chaines dont vous a parlé Mickaël.

Cet outil va prendre donc ces données sur la similarité collectées et va les appliquer aux chaines, cela va se pencher sur les noms existants, réservés ou bloqués pour savoir s'ils sont similaires entre eux et aussi entre les différentes étiquettes de variantes. Et puis nous allons avoir un rapport sur l'ensemble conflictuel potentiel qui sera soumis au panel de révision.

Sarmad, je vous rends la parole.

SARMAD HUSSAIN: Il y a un détail à rajouter à cette liste, lorsqu'il s'agit des 5 niveaux présentés par Mickaël, nous allons utiliser les trois premiers niveaux pour faire une liste plus courte des données pour ce qui est des ensembles conflictuels

potentiels, donc pour finaliser le panel. Donc nous n'allons pas considérer les niveaux 4 et 5, nous allons nous aligner avec les recommandations qui proposent d'étudier les confusions visuelles et non les confusions visuelles possibles.

Prochaine diapositive, s'il vous plait.

Comme vous le voyez ici, le travail de la collecte de données et du développement de cet outil est en cours pour ce dernier trimestre. Nous aurons collecté les données au début de l'année prochaine et bien sûr nous partageons les informations sur les données et sur l'outil et nous ferons cela aussi à l'ICANN 82. Cet outil devrait être terminé, les données devraient être intégrées et tout cela devra être testé très bientôt. Nous avons pour cible le mois de juin 2025.

Voilà la dernière diapo pour cette section de la présentation. Y a-t-il des questions sur la similarité des chaines ?

NABIL BENAMAR: Merci de votre présentation. J'ai noté que vous avez basé votre étude sur ce qui est des standards visuels des similarités. Je me posais la question, y a-t-il des standards, des normes scientifiques pour qu'il n'y ait pas d'interprétation faite par des parties intéressées sur le sujet de cette similarité et sur ce qui est de cette confusion ?

Ce n'est pas une science exacte, cela dépend de l'expérience de l'utilisateur. Chaque utilisateur peut mettre une étiquette sur toutes les catégories, surtout sur la catégorie 1, 2 ou 3. Quand vous passez à la catégorie 4, vous basez votre définition sur les utilisateurs attentifs.

Comment garantir qu'un utilisateur va toujours être attentif ? Cela ne va pas être une garantie à chaque fois, pour chaque expérience.

Vous avez des défis et j'aimerais savoir ce que vous en pensez par rapport au travail déjà fait dans différents scripts.

SARMAD HUSSAIN: Merci, Nabil. C'est une thématique que nous avons couverte dans ce document qui a été publié, quand nous l'avons fait. Nous avons utilisé des révisions similaires, nous avons répondu, débattu sur ces questions quand nous avons établi ces grandes lignes.

Par exemple nous avons parlé avec des personnes familières avec ce script ou moins familières. Nous avons aussi discuté avec d'autres personnes qui sont très attentives à ce qui est écrit, et pas forcément des personnes qui ne font pas très attention aux détails, qui n'observent pas les choses en détail. Et lorsque nous avons proposé nos grandes lignes, nos directives, ce que nous avons publié auparavant, nous avons bien sûr étudié les similarités basées sur ce que nous avaient dit les lecteurs dans ces langues.

Donc nous basons nos recherches aussi sur les personnes qui lisent vraiment attentivement et qui font très attention dans la lecture des noms de domaine. Donc si nous pouvons dire cela, c'est parce que les recommandations de politique émanant du PDP sont très claires. En ce qui concerne les similarités, il ne faut pas juste étudier les possibilités de similarité, mais les probabilités de similarité.

Donc ces critères sont déjà en place et il nous faut être plus ou moins conservateurs dans ce sens. Donc, quand on parle de ne pas être très attentifs, de lire vaguement ces scripts, oui, il pourrait y avoir des

possibilités de confusion, mais nous voulons travailler à un niveau où les personnes font très attention et sont très attentives à l'écriture.

NABIL BENAMAR: Cela pourrait être présent dans d'autres langages ou scripts spécifiques parce qu'il y a beaucoup de différences entre les différents scripts que nous étudions maintenant. Si vous examinez le script arabe, sachez que nous avons les points ou les trémas, donc ces petits points qui peuvent être confus. Cela pourrait porter à confusion avec d'autres petits symboles qui sont utilisés dans d'autres langues au sein même de notre communauté de langue arabe.

Donc pour le latin, nous n'avons pas ces exemples, très peu de lettres utilisent ces petits points.

SARMAD HUSSAIN: Bien sûr, c'est pour cela que nous avons impliqué des experts locaux qui utilisent ces alphabets, et cela nous a permis de recevoir des informations de leur part et nous sommes conscients de ces problèmes.

Nous allons passer au prochain sujet et nous reviendrons à ces questions après. Nous pouvons passer à la présentation de Pitinan.

PITINIAN KOOARMORNPATANA : Je pense que Mickaël a un petit point à rajouter.

MICKAËL BAULAND: Un petit point rapide par rapport à ce que Nabil a dit. Un point important dont il faut se rappeler, c'est que pour tout ce qui est du processus de collecte de données et ce que nous avons développé, ça a été fait pour aider, justement, cette équipe de révision de similarité des chaines, pour qu'elle puisse faire son travail. Ces décisions ne sont pas toujours parfaites, mais notre équipe de révision va prendre les décisions finales. Nous avons

dû faire cela pour faciliter leur travail. Nous n'avons pas encore de décision finale.

PITINIAN KOOARMORNPATANA : Alors, pour le prochain sujet, nous allons parler du FGR, de l'outil d'amélioration pour les règles de génération d'étiquettes. Voilà le contexte.

Pour l'instant cet outil est disponible en ligne. Il y a trois modes fournis. Nous avons le mode de base qui nous aide à valider les étiquettes qui nous intéressent et il y a la référence LGR ou un script LGR dans la zone racine. C'est un outil pour faire la révision de la table des IDN. On peut utiliser cet outil pour faire les comparaisons.

Ensuite, nous avons le mode avancé qui permet à l'expert, à l'utilisateur de faire des comparaisons, des modifications, etc., et faire des tests. Cet outil est disponible sur ce lien que vous avez à l'écran, LGRTOOL.ICANN.ORG. Si vous avez un compte ICANN, vous avez un accès à l'outil.

Pour la communauté technique, il est aussi disponible, c'est open source, les manuels sont aussi disponibles si vous utilisez ce lien.

Pour la prochaine série, nous voulons améliorer cet outil pour inclure ces différentes fonctionnalités ; tout d'abord vous allez pouvoir vérifier une étiquette primaire qui inclut ASCII et IN, elles vont être rassemblées sous un protocole d'exigences, donc est-ce que la chaine valide dans la LGR de la zone racine.

Alors, comme je l'ai dit tout à l'heure, nous allons peut-être avoir des exigences un peu différentes sur la longueur des scripts, nous allons vérifier cela aussi. Et nous allons aussi vérifier que cette étiquette primaire est en collision avec un TLD existant ou avec un ensemble de variantes. Nous

allons faire une révision sur tout ce qui est noms réservés et aussi les variantes des noms réservés. Nous allons étudier tous les critères avant de faire avancer cette chaine. Nous allons aussi vérifier s'il y a collision avec les noms bloqués.

Ensuite, nous aurons aussi des fonctions pour vérifier les étiquettes de variantes, des vérifications comme sur la liste précédente, mais aussi vérifier si la chaine d'une étiquette de variante attribuable de l'étiquette primaire, si cette chaine est attribuable.

Voilà, cet outil sera utile et disponible pour la communauté. C'est tout pour moi. Je vous rends la parole, Sarmad.

SARMAD HUSSAIN: Je voudrais rajouter quelque chose sur la donnée des similarités de chaine. Une fois que nos aurons reçu les données finales et que nous aurons fini notre révision, nous allons faire une révision en interne et nous allons publier ce document afin que la communauté puisse y avoir accès. Donc vous allez pouvoir étudier ces données et nous attendons vos rétroactions sur cette analyse faite par des experts en alphabet.

Et donc, que vous soyez d'accord ou non avec les catégorisations utilisées par ces experts de langage, vous allez pouvoir nous donner votre avis sur le niveau de similarité.

Nous espérons pouvoir proposer cela bientôt et nous espérons que vous allez vous intégrer à ce processus afin que nous puissions finaliser le travail.

Donc ces données sur la similarité sont parfois logiques parce qu'il s'agit du niveau des séquences des lettres ou des caractères, mais parfois ce n'est pas le cas, cela prend une analyse supplémentaire.

Et donc le panel de révision va examiner cela et cet outil que nous avons utilisé pour faire cet échantillonnage ou ce screening en amont, a donc un champ très limité et il va pouvoir générer des cas potentiels pour des similarités de chaines qui vont être étudiées par le panel.

Donc la décision finale sera faite à l'aide d'un processus manuel.

Si vous avez d'autres questions, n'hésitez pas.

ANIL JAIN: J'ai une question pour Pitinan. Le test de la LGR que vous avez fait est fantastique, je pense qu'on a couvert tous les aspects du problème et je pense que ce sera un outil très utile pour la prochaine série. Est-ce que cet outil est disponible pour les ccTLD et les gTLD ou est-ce que c'est seulement pour les gTLD ? Première question.

Question numéro 2 : si un opérateur de registre ne peut pas gérer ce testing LGR, est-ce qu'il y a des plans de la part de l'ICANN pour appuyer ou aider les opérateurs de registre afin qu'ils puissent vérifier cette règle de génération d'étiquettes ?

SARMAD HUSSAIN: Le ccPDP4 pour les IDN, ces politiques sont en cours en ce moment, une fois que cette politique sera mise en œuvre, nous pourrons étudier la question de savoir si nous allons mettre en œuvre cet outil pour qu'il soit utile à la communauté, à savoir si nous allons aussi le mettre à la disposition de la communauté ccTLD.

Pour l'instant nous sommes focalisés sur la prochaine série de gTLD. Voilà pour la première question.

PITINIAN KOOARMORNPATANA : Il y a les TLD existants qui incluront aussi les gTLD et les ccTLD.

SARMAD HUSSAIN: Il y avait une deuxième question ?

ANIL JAIN: Avec ces directives, si un directeur ou un opérateur de registre ne peut pas comprendre comment vérifier ces LGR, est-ce qu'il y a un plan de la part de l'ICANN pour apporter du soutien pour ces difficultés ?

SARMAD HUSSAIN: Et bien je pense que oui, un soutien sera octroyé dans le cadre de la prochaine série des gTLD et je pense que les canaux de soutien seront rendus accessibles. Donc ma réponse à votre question est oui, et tout ceci sera géré dans le cadre du nouveau programme.

ANIL JAIN: Merci, Sarmad et merci Petinan.

SARMAD HUSSAIN: D'autres questions ou commentaires ?

ASHLEY ROBERTS: Bonjour, je vous parle de la part de la perspective d'un nouveau candidat pour un gTLD, évidemment c'est très complexe, je me demande comment un candidat peut s'en sortir, tenter de déterminer s'il y a un problème de similarité de chaines ? L'outil que vous avez utilisé et sur lequel vous travaillez actuellement, à quoi ressemble-t-il ? Quels seront les résultats présentés ?

SARMAD HUSSAIN: Vous pourriez revenir sur la diapo que Julien présentait ? Julien, vous voulez répondre ? À quoi ressembleraient les résultats que nous donnerait cet outil d'évaluation de similarité des chaines ?

JULIEN BERNARD: Alors, on appellera cela l'ensemble conflictuel et donc vous allez présenter une candidature pour un TLD et, bien sûr, il y aura peut-être des TLD existants, des noms bloqués ou réservés qui pourraient peut-être prêter à

confusion. Et c'est ce qu'on s'attend à obtenir comme résultats de la part du panel de révision.

ASHLEY ROBERTS: Merci. Donc cet outil sera utilisé seulement par le panel de révision ou aussi par les candidats ?

JULIEN BERNARD: Pour le moment, cet outil est développé pour qu'il soit utilisé par le panel de révision, mais si la communauté estime qu'il serait utile pour les candidats, nous ferons en sorte de le rentre utile pour la communauté. Bien sûr l'idée est que le panel puisse faire cette présélection de TLD.

ASHLEY ROBERTS: Merci. Mais bon, voilà, c'est une question très complexe, donc il y aurait il peut-être des lignes directrices supplémentaires, des conseils, des outils que vous pourriez nous donner ? Je me suis déjà penchée sur la question, j'ai consulté les lignes directrices et, personnellement, je l'ai trouvé assez difficile à digérer. Aussi, voilà, cela m'inquiète un peu. Les candidats, les personnes nouvelles dans ce domaine, je m'inquiète de comment ils vont s'en sortir d'une façon efficace.

SARMAD HUSSAIN: Nous prenons note de votre commentaire et nous allons voir ce que nous pourrons faire lorsque cet outil sera disponible.

Nous avons dépassé le temps qui nous été alloué, nous allons clôturer cette session. Si vous avez d'autres questions, n'hésitez pas, je pourrais y répondre après la session, sinon n'hésitez pas à envoyer vos questions à ICANN.ORG ou au programme dont je suis responsable, nous y répondrons. Sinon, je reste à votre disposition si vous avez des questions suivies et nous pouvons clôturer cette session, merci.

[FIN DE LA TRANSCRIPTION]