



UPDATE & CONSULTATION ON
New gTLD Subsequent Procedures
Consensus Building on Recommendations

STRING SIMILARITY (v2)

Justine Chew
20 April 2020

String Similarity in SubPro: Background

- What is the New gTLD Subsequent Procedures (“SubPro”)?
 - ❖ The set of rules and mechanisms applicable to the next round for New gTLDs, i.e. they DO NOT apply to legacy TLDs, ccTLDs, or delegated new gTLDs or those still unresolved from the 2012 application round
 - ❖ “An update” to the 2012 Round rules and mechanisms
- Recap of String Similarity Review in 2012 Round
 - ❖ One of 4 string reviews undertaken on applied-for string within the Initial Evaluation Process
 - ❖ Undertaken by an independent String Similarity Panel:
 - Identify if an applied-for string is “similar” to any existing TLD, any other applied-for strings, reserved names, and in the case of 2-character IDNs, any single character or any 2-character ASCII string
 - Standard used: visual similarity; where “Similar” means string similarities that would create a probability of user confusion if more than one of the strings is delegated into the root zone
 - Informed by Sword Algorithm – “algorithmic score for visual similarity”
 - If found to be similar, applied-for strings are placed in contention sets
 - Ultimately found 4 strings in contention: .hotels v .hoteis, and .unicom v .unicorn

String Similarity in SubPro: Fast Forward to Current PDP



ALAC STATEMENTS touched on:

Singular v. Plural

- Reducing risk of foreseeable consumer confusion due to plurals and singulars of the same word within the same language/script being allowed

No Overlap

- No overlap in rounds for a string that is still being processed from a previous application
- CPE as appropriate method for resolving contentions (if administered properly)
- Concerns and need to draw up rules to disincentive gaming or abuse of private auctions in respect of string similarity/contention sets

Synonyms to verified TLDs

- Treatment for synonym strings associated with a highly regulated sector or are verified TLDs



RELATED SubPro Areas/Topics include:

- String Confusion Objections & challenges
- Dispute resolution review
- SSAC: RFC 5894 IDNA: Background, Explanation & Rationale
- Public Interest Commitments (PICs)



COMPETITION, CONSUMER CHOICE & TRUST (CCT) RECOMMENDATIONS

- Rec #35: Consider adopting new policies to avoid potential for inconsistent results in string confusion objections:
 - 1) Determine through initial string review process, singular and plural versions of same string which should not be delegated
 - 2) Avoid disparities in similar disputes, ensure similar cases of plural versus singular strings are examined by same expert panelist
 - 3) Introduce a post dispute resolution panel review mechanism

Impact of SubPro Recommendations * as at 20 Apr 2020

SubPro PDP WG

Affirmation #1:

- WG affirms Recommendation 2 from the 2007 policy, “Strings must not be confusingly similar to an existing top-level domain or a Reserved Name.”
- Subject to the following recommendation, WG affirms standard used in the String Similarity Review from 2012 to determine **whether an applied-for string is “similar” to any existing TLD, any other applied-for string, reserved names**, and in the case of 2-char IDNs, any single-char or 2-char ASCII string.
 - ❖ Per s. 2.2.1 of the 2012 AGB, “similar” means “strings so similar that they create a probability of user confusion if more than one of the strings is delegated into the RZ.”
- In 2012 round, the **String Similarity Panel was tasked with identifying “visual string similarities that would create a probability of user confusion.”** WG affirms the visual standard for determining similarity and recommends that the Panel additionally consider as part of the standard **whether strings are intended to be used as the singular and plural version of the same word.**

WG’s Rationale

- Believes that existing policy and implementation related to the String Similarity Review remain appropriate, so affirms Recommendation 2 from 2007 and the existing evaluation standard described in the AGB, as amended herein.

For At-Large Consensus Building

Impact

- Expands scope of String Similarity Review to reduce risk of foreseeable consumer confusion due to plurals and singulars of the same word within the same language/script being allowed
- Addresses CCT Rec #35 **(1) Determine through initial string review process, singular and plural versions of same string which should not be delegated**

Singular v. Plural

Additional intervention

- **Need understanding of impact to IDN variants?**
- **Any concerns? What else needs to be done?**

* From SubPro PDP WG, not limited to recommendations, but also affirmations and implementation guidance

Impact of SubPro Recommendations as at 20 Apr 2020

SubPro PDP WG

Recommendation #2:

- WG recommends updating the standards of both (a) confusing similarity to an existing top-level domain or a Reserved Name, and (b) similarity for purposes of determining string contention, to address singular and plural versions of the same word, noting that this was an area where there was insufficient clarity in the 2012 round.
- Specifically, WG recommends **prohibiting plurals and singulars of the same word within the same language/script in order to reduce the risk of consumer confusion.**
 - For eg, the TLDs .EXAMPLE and .EXAMPLES may not both be delegated because they are considered confusingly similar.
- This **expands the scope of the String Similarity Review to encompass singulars/plurals of TLDs on a per-language/script basis.**

WG's Rationale

- Neither GNSO policy from 2007 nor the 2012 Applicant Guidebook defined a specific rule regarding singulars and plurals of the same string, and in the 2012 application evaluation process, the String Similarity Evaluation Panel did not find singular and plural versions of strings to be visually confusingly similar. The GAC, the ALAC, The ICANN Board, and the Final Issue Report on New gTLD Subsequent Procedures have raised that existing guidance does not address the issue of singulars and plurals of the same word and that additional guidelines may be needed.
- WG's recommendation to prohibit singulars and plurals of the same word within the same language/script and to expand the scope of the String Similarity Review to include singulars/plurals provides a clear, consistent standard for subsequent procedures that will provide greater predictability for applicants

For At-Large Consensus Building

Impact

- Prohibition on plurals and singulars of the same word within the same language/script to reduce risk of consumer confusion

Singular v. Plural

Impact of SubPro Recommendations as at 20 Apr 2020

SubPro PDP WG

Recommendation #2 (cont'd):

- This expands the scope of the String Similarity Review to encompass singulars/plurals of TLDs on a per-language/script basis.
 - ❑ An application for a single/plural variation of an existing TLD or Reserved Name will not be permitted if the intended use of the applied-for string is the single/plural version of the existing TLD or Reserved Name.
 - For eg, if there is an existing TLD .SPRINGS that is used in connection with elastic objects and a new application for .SPRING that is also intended to be used in connection with elastic objects, .SPRING will not be permitted.
 - ❑ Applications will not automatically be placed in the same contention set because they appear visually to be a single and plural of one another but have different intended uses.
 - For eg, .SPRING and .SPRINGS could both be allowed if one refers to the “season” and the other refers to elastic objects, because they are not singular and plural versions of the same word.

- However, if both are intended to be used in connection with the elastic objects, then they will be placed into the same contention set.
 - Similarly, if an existing TLD .SPRING is used in connection with the season and a new application for .SPRINGS is intended to be used in connection with elastic objects, the new application will not be automatically disqualified.
- WG recommends using a dictionary to determine the singular and plural version of the string for the specific language.

WG's Rationale

- The recommendation that singular/plural versions of the same string should be considered confusingly similar only applies when both strings are intended to be used in connection with the same meaning of the word.
- In the case where two applications are submitted during the same application window for strings that create the probability of a user assuming that they are single and plural versions of the same word, but the applicants intend to use the strings in connection with two different meanings, both strings may be permitted to proceed.

Impact of SubPro Recommendations as at 20 Apr 2020

SubPro PDP WG

Recommendation #2 (cont'd):

WG's Rationale (cont'd)

- In such cases there needs to be a means for the registries to commit to the use stated in the application and a method for enforcing adherence to this commitment. **The WG believes that a mandatory PIC will serve this need.**
- WG notes that Recommendation 35 from the Competition, Consumer Trust, and Consumer Choice Review Team states: "The Subsequent Procedures PDP should consider adopting new policies to avoid the potential for inconsistent results in string confusion objections, in particular:
 - ❑ 1) Determining through the initial string similarity review process that singular and plural versions of the same gTLD string should not be delegated
 - ❑ 2) Avoiding disparities in similar disputes by ensuring that all similar cases of plural versus singular strings are examined by the same expert panelist

For At-Large Consensus Building

Impact

Meets the earlier positions taken by ALAC:

Singular v. Plural

- Prohibition on plurals and singulars of the same word within the same language/script to reduce risk of consumer confusion
- Expanding scope of String Similarity Review to encompass singulars/plurals of TLDs on a per-language basis – applications for a plural version of a singular word in the same language should go into contention set because they are confusingly similar – application for single/plural variation of an existing TLD would not be permitted -- but applications should not be automatically disqualified because of a single letter difference with an existing TLD (eg .NEW and .NEWS)
- To address CCT Rec #35 (Ed. Parts 1 and 2)
- Using a dictionary to determine singular and plural versions of the string for the specific language

Additional intervention

No Overlap

- To confirm disallowing application for a string that is still being processed from a previous application opportunity, to avoid creating unintended contention set. Consequentially, need a way to terminate any application that has little chance of succeeding and which are not withdrawn in subsequent procedures.
- "Use of mandatory PICs" should appear in the recommendation
- **Any other concerns? What else needs to be done?**

Impact of SubPro Recommendations as at 20 Apr 2020

SubPro PDP WG

Recommendation #3

- Eliminate the use of the SWORD tool in subsequent procedures.

WG's Rationale

- WG agreed that there was insufficient correlation between the results of the SWORD Tool and the outcomes of the String Similarity Review, indicating that that tool, as implemented, was not a helpful resource for evaluators and especially for applicants, where the SWORD results could be counter productive.
- Given the limited utility of SWORD Tool to provide consistent and predictable results, the Working Group believes that it should not be used in subsequent procedures.
- WG leaves open the possibility that in the implementation phase, an alternate tool may be leveraged to address the issues experienced in the 2012 round.

For At-Large Consensus Building

Impact

- We did not object to proposal for elimination of SWORD tool given anecdotal evidence / comments of its unhelpfulness.
- We did suggest that a suitable alternative be identified; while no one objected to suggestions, there weren't any immediate ideas on viable alternatives, so punt to IRT.

Additional intervention

- **Any concerns? What else needs to be done?**

Impact of SubPro Recommendations as at 20 Apr 2020

SubPro PDP WG

For At-Large Consensus Building

Recommendation #4

- The deadline for filing a String Confusion Objection must be no less than thirty (30) days after the release of the String Similarity Evaluation results. This recommendation is consistent with PIRR recommendation 2.3.a, “Review the relative timing of the String Similarity evaluation and the Objections process.”

WG’s Rationale

- WG notes that the delay of releasing String Similarity results during the 2012 round caused those wishing to file a String Confusion objection to only have two weeks to file the String Confusion Objection, which many viewed as too short. Therefore, the Working Group recommends that there be at least thirty (30) days between the publication of the String Similarity Evaluation results and the deadline for filing a String Confusion Objection.

Impact

- To help ensure that String Confusion Objection period runs for 30 days.

Additional intervention

- None.

New Issues on String Similarity as at 20 April 2020

SubPro PDP WG

Unanswered still:

Cross-language complications

- Where an applicant may suggest a particular language of a label when applying for a TLD and operating that TLD, but the user might not relate to the label in the same language.
 - How should it be handled if there are two strings which belong to two different languages from the applicant point of view, but they represent singular/plural form of the same word in a particular language?
 - What should be the primary consideration in developing rules -- the intent of the applicant or possible confusion by the user?
- Maybe the only way to address potential concerns about end user confusion in the application process is to look at the intent of the applicant, because the TLD has not yet been launched. But the user may still ultimately be confused by the end result if the sole focus is on the intent of the applicant.

Singular v Plural forms in different languages

- “It may not be possible for rules regarding string similarity to be as simple or straightforward as the above referenced preliminary recommendations state. For example, singular and plural noun forms are represented differently by different languages.” – SSAC
- Would suggestion to use a dictionary to determine singular/plural versions of a word to achieve primary goal of developing policy on this topic is to prevent clear cases where the applied-for TLD is a singular or plural of an existing TLD. Leave edge cases to be handled through additional contract language.

Inflectional morphology - different forms of inflection beyond pluralization

- For example, in addition to inflection associated with singular/plural forms of a word, nouns in some languages inflect for gender

New Issues on String Similarity as at 20 April 2020

SubPro PDP WG

Sufficiently answered?

Inflectional morphology - different forms of inflection beyond pluralization

- For eg, in addition to inflection associated with singular/plural forms of a word, nouns in some languages inflect for gender
- Applies to verbs also - verb conjugation being a form of inflection; as eg, “decide” and “decides” are different forms of the verb inflected for agreement with singular and plural subject.
 - Does it make sense that the “s” would differentiate between two forms of a noun and not two forms of a verb for the purposes of defining confusing similarity?
 - If a grammatical category like singular or plural is confusingly similar, why not also consider other grammatical categories confusingly similar like masculine and feminine or different tenses?
 - Is there a way to make the framework for determining confusing similarity manageable so that it is predictable to the applicant?
- WG received feedback from ICANN org that from a linguistics perspective, inflection on a per-language basis is fairly well understood and bounded. Inflections are given in many dictionaries, which makes it possible to apply rules about inflection consistently

Semantics

- “Beyond visual similarity, trying to determine confusability based on the meaning of words is fundamentally misguided, as domain names are not semantically words in any language.” – SSAC
- WG considered an alternate point of view that the SSAC’s statement may be true from a purely technical perspective, but many of the gTLDs now delegated have semantic intent.

IDN ccTLD

- WG conducted a comparison between the gTLD String Similarity Review and the review for string similarity that takes place as part of the IDN ccTLD Fast Track Process to determine if any additional harmonization between the two processes may be appropriate.
- WG noted that both reviews focus on a similar standard for visual similarity. In addition, both processes compare requested/applied-for TLDs against existing TLDs, reserved names, and other applied-for strings (ccTLDs or gTLDs).
- There is within the ccTLD process, the possibility for a second review of the DNS Stability Panel’s initial review.
- An external and independent Extended Process Similarity Review Panel (“EPSRP”) too conducts a second review.
- Then, there will be new challenge mechanisms in SubPro.

New Issues on String Similarity as at 20 April 2020

SubPro PDP WG

No consensus, no conclusions

Synonyms

- WG considered whether synonyms should be included in the String Similarity Review for those strings associated with highly-regulated sectors and those representing verified TLDs. The example of .DOCTOR and .PHYSICIAN was raised in discussion. Public comments expressed diverging perspectives on this issue.
- The Working Group further considered whether exact translations of these strings should be included in the String Similarity Review, but did not conclude the discussion with any recommendations.

Synonyms to
verified TLDs

Homonyms

- WG considered a proposal put forward in public comment that homonyms should be included in the String Similarity review. From one perspective, homonyms may cause user confusion, for example in the 2012 round an application for .thai phonetically clashed with existing . ๓๐๕ (Thai IDN ccTLD)
- Some WG members felt that there is the possibility of end-user confusion if two TLD strings are spelled differently but pronounced the same.
- Other WG members did not feel that there was a clear problem to address through policy with respect to homonyms. It was raised that even if the WG agreed that there was a well-defined problem that needed to be solved, it might not be possible to develop clear rules on homonyms that could be fairly enforced.
- Some WG members raised that even within a language, there may be different pronunciations of a word. Across languages, it is even more difficult to determine whether words are pronounced the same.
- The WG did not conclude the discussion with any recommendations