

From: Jeffrey Smith [<mailto:JSmith@dotshop.com>] **Sent:** Tuesday, March 12, 2013 05:49 **To:** Christine Willett **Subject:** Committe on String Confusion and Section 2.2.1.1.1 issue

Christine,

In following up on our previous conversation, I think we have discovered what may have caused concerned for us all. Please take a look at the below script and see if there simply was some sort of mistake made on the final Applicant Guidebook. Your insight and assessment is greatly appreciated.

On June 4, 2012 the final version of the Applicant Guidebook was release which stands as a contract for anyone wishing to apply for the delegation of new gTLD's in or about 2012/2013. In this guidebook it has several sections that describe String Similarity:

See <http://newgtlds.icann.org/en/applicants/agb>

In Section 1.1.2.10 String Contention is defined as:

String contention refers to the scenario in which there is more than one qualified application for the identical gTLD string or for similar gTLD strings. In this Applicant Guidebook, "similar" means strings so similar that they create a probability of user confusion if more than one of the strings is delegated into the root zone.

In Section 2.2 Initial Evaluation as:

Whether the applied-for gTLD string is so similar to other strings that it would create a probability of user confusion;

In Section 2.2.1.1 String Similarity as:

This review involves a preliminary comparison of each applied-for gTLD string against existing TLDs, Reserved Names (see subsection 2.2.1.2), and other applied-for strings. The objective of this review is to prevent user confusion and loss of confidence in the DNS resulting from delegation of many similar strings.

Note: In this Applicant Guidebook, "similar" means strings so similar that they create a probability of user confusion if more than one of the strings is delegated into the root zone.

The visual similarity check that occurs during Initial Evaluation is intended to augment the objection and dispute resolution process (see **Module 3, Dispute Resolution Procedures**) that addresses all types of similarity.

This similarity review will be conducted by an independent String Similarity Panel.

An issue arises in Section 2.2.1.1.1 Review Performed as:

The String Similarity Panel's task is to identify **visual** string similarities that would create a probability of user confusion...

In the previous version of the Applicant Guidebook it states:

This review involves a preliminary comparison of each applied-for gTLD string against existing TLDs, Reserved Names (see subsection 2.2.1.2), and other applied-for strings. The objective of this review is to prevent user confusion and loss of confidence in the DNS resulting from delegation of many similar strings.

Note: In this Applicant Guidebook, "similar" means strings so similar that they create a probability of user confusion if more than one of the strings is delegated into the root zone.

The visual similarity check that occurs during Initial Evaluation is intended to augment

the objection and dispute resolution process (see Module 3, Dispute Resolution Procedures) that addresses all types of similarity.

This similarity review will be conducted by an independent String Similarity Panel.

<http://archive.icann.org/en/topics/new-gtlds/dag-en.htm>

Changes from the Guidebook to the Final Guidebook are based upon the Analysis of Comments in the Section **String Similarity and String Contention Resolution** whereby it summarizes key points on Page 68:

For the initial evaluation, the proposed position is to keep the similarity assessment restricted to visual similarity only, especially in view of the complexities involved with assessing for example aural similarity, which can be invoked in the subsequent string similarity objection process. Final decisions on similarity will be made by a panel, as string similarity algorithm outcomes are only indicative, not authoritative. Community discussions have made it clear that human assessment is a necessity.

Allowing for agreements between applicants to have confusingly similar strings coexist as TLDs would imply risks for registrants and end users and can only be considered when policy has been developed on provisions and procedures to reduce or eliminate such risks. Similar concerns may relate to linguistic variations of a string from a single applicant, while also noting that there are special provisions in the current approach for variant handling for IDN strings.

Please see <http://archive.icann.org/en/topics/new-gtlds/summary-analysis-agv6-30may11-en.pdf>

It is important to realize that the Analysis of Comments on Page 70 are the decisions, the Comments section beforehand are only for referential purposes. **It appears that there simply is a mistyping in the final Applicant Guidebook as visual similarity was only supposed to be in the initial evaluation – the committee of string confusion should have indeed determined similarity of string based on all criteria, visual, aural and similar meanings.** Further detail is included below on this issue which appears on Page 70 under Analysis of Comments.

The comments regarding the scope of the similarity assessment are well taken. As has been stated in relation to previous public comment periods, the string similarity assessment in the initial evaluation is solely focused on visual similarity. The support from many for that approach is noted, as is the diverging view that aural similarity be considered, an approach that is controversial in principle and very difficult to perform in practice, while such similarity can indeed be invoked in a subsequent string similarity objection process. The proposed position is to keep the established approach unchanged. One comment suggests that, —Aural and meaning similarity should not be considered at all. As reinforced by community discussion, possible examination for these types of similarity was included in the policy recommendations of the GNSO that was approved by the Board. The idea is that user confusion should not be likely to occur – no matter what the cause of that confusion, Therefore, absent other policy advice, the current objection model that includes all types of confusion will remain in place, although the **similarity assessment during initial evaluation will be limited to visual similarity.**

Regarding suggestions that applicants can agree on coexistence for confusingly similar strings, it has repeatedly been clarified in responses to previous public comment periods that a finding of confusing similarity cannot as such be resolved thru mutual agreement by the involved applicants. Such an approach would not make the strings appear less confusingly similar to the internet user, which is the fundamental aspect to consider,

especially given the considerable security risks for registrants and end user that such similarities can entail over the whole lifespan of the involved TLDs. A policy basis for agreement provisions and safeguards to eliminate such risks must be developed before such an approach can be considered. This matter has already been addressed in previous public comment analyses and the proposed position is not to change the current approach in the Applicant Guidebook in this regard.

Regarding the noted high similarity scores provided by the algorithm for strings that arguably can coexist, given that they have coexisted on the second level under .com without causing problems, it must be emphasized that the algorithm score is only one input to be considered by the string similarity panel and not authoritative in any way regarding findings of similarity. Community discussions have made it clear that confusion is a human reaction and that consideration by humans is indispensable for truly assessing similarity, which will thus be the task of a panel. It is the intention to refine the algorithm in view of the panel's findings and thereby improve it for future rounds, but the algorithm outcomes will be considered as solely indicative for now. No change in that approach is foreseen for the first round.

For the claim that —linguistic variations of a string from a single applicant should not be put in contention set, one has to distinguish between a couple of different cases. If the intended meaning is —variant TLD strings declared by an applicant as described in the Guidebook (see separate section), thus occurring within a single IDN gTLD application, they will be handled according to those rules and not be put in a contention set based on the applicants declaration of variants, while still being considered as a basis for assessing similarities with other applied-for strings. If the intended meaning is translations/transliterations/transcriptions, the strings would appear in separate applications and be assessed for visual similarity and may indeed be found to be confusingly similar, for example in the case of an ASCII string and a Cyrillic string. Such strings will not be permitted to coexist as gTLDs in the DNS, regardless of whether they are put forward by the same applicant. Future policy development may potentially change this approach, provided sufficient safeguards can be identified, but for the first New gTLD application round no change in this approach is foreseen.

Thank you

Jeffrey Smith



Jeffrey Smith
President, Commercial Connect, LLC.
1418 South Third Street
Louisville, KY 40208-2117
(502) 235-3475 Cell
<http://www.dotShop.com>