

# Internationalized Domain Names Expedited Policy Development Process

Continued Discussion of String Similarity Small Group Outputs



IDN-EPDP Team Meeting #49 | 26 August 2022

# Agenda

---

1. Roll Call & SOI Updates (2 min)
2. Welcome & Chair Updates (5 min)
3. Continued Discussion of String Similarity Small Group Outputs (80 min)
4. AOB (3 mins)

# Small Group Recommendation: Hybrid Model

**Summary:** *The small group recommends the hybrid model, which is a mixed-level approach between level 2 and level 3*

**Goal:** *Mitigate any possibility of confusing similarity between one IDN TLD and another IDN TLD or any of its valid variant(s)*

*In practice, the string similarity review must be modified to compare:*

- **An applied-for primary IDN gTLD and all of its allocatable variant label(s)**

*Against:*

- **Existing TLDs and all of their allocatable and blocked variant labels;**
- **Strings requested as IDN ccTLDs and all of their allocatable and blocked variant labels;**
- **Other applied-for gTLDs in the same round and all of their allocatable and blocked variant labels;**
- **Reserved Names; and**
- **Any other two-character ASCII strings and all of their allocatable and blocked variant labels (if the applied-for primary IDN gTLD is a two-character string)**

*In addition, compare:*

- **All of the blocked variant label(s) of an applied-for primary IDN gTLD**

*Against:*

- **Existing TLDs and all of their allocatable variant labels**

**Note:** *Blocked variants of one IDN TLD should NOT be compared against blocked variants of another IDN TLD*

# Relevant Background for Consideration of Hybrid Model

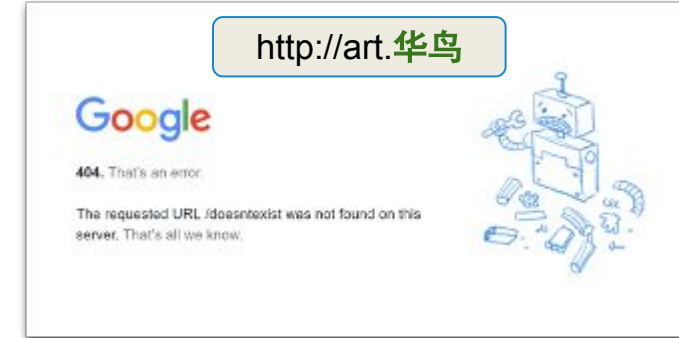
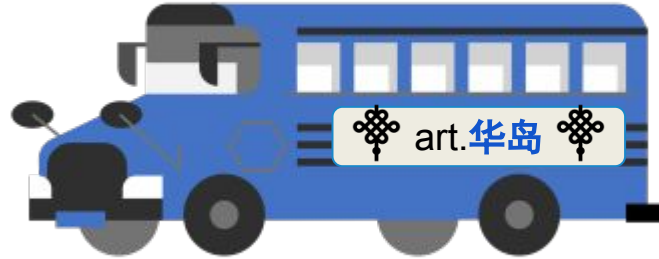
- ❑ **RFC 5891**: Any domain name registry, including that of the root zone, should develop and apply additional **restrictions** as needed **to reduce confusion and other problems** (part of IDNA2008 standard)
- ❑ **RFC 6921**: Zones higher in the DNS tree tend to have **more restrictive rules**...the context is that the root zone serves the entire Internet population
- ❑ **SAC089**: **Confusability cannot be considered in isolation from other issues related to security**. Phishing and other social engineering attacks based on domain name confusion are a security problem for end users
- ❑ **Staff Paper**: Variant implementation must be done in a way that **operation and maintenance of the DNS not be adversely impacted by the introduction of variants**; it should **avoid including variant TLDs in a manner that would create user vulnerabilities or a probability of confusion**

# Discussion Recap

- ❑ EPDP Team had **general support** for comparing the **primary** labels and **requested allocatable** variant labels
- ❑ Some EPDP Team members expressed **reservations** about including **non-requested allocatable** variant labels and **blocked** variant labels in the String Similarity Review
- ❑ String Similarity Small Team recommends the hybrid model to meet the **singular goal of risk mitigation of failure modes**, which are 1) **denial of service**, and 2) **misconnection**
- ❑ String Similarity Small Team believes that the string similarity review provides an opportunity to mitigate such risks as much as possible
- ❑ String Similarity Small Team **did not consider implementation complexity** of the hybrid model and deferred this to EPDP Team to deliberate

# Denial of Service: Example & Illustration

A user attempts to visit `http://example.X`, reading it as being the same as the `http://example.Y` that, for example, he or she saw in an advertisement. After typing the address (`http://example.X`), the connection does not work as `http://example.X` is not registered.



Denial of service will likely cause user confusion and frustration but not harm

# Misconnection: Example & Illustration


A user attempts to visit <http://example.X>, reading it as being the same as the <http://example.Y> that, for example, he or she saw in an advertisement. After clicking on <http://example.Y>, the user arrives at a site controlled by a registrant different to <http://example.X>.




# Misconnection: Potential Consequences

---

 Misconnection may be **more problematic than denial of service** and cause more harm to the user beyond confusion and frustration

 Arriving at the wrong site, even legitimate, can **result in credential compromise and accidental exposure of information**

 If confusing similarity is maliciously leveraged, it can be a **DNS abuse** vector. When confusion is at the top-level, the possibility of DNS abuse is **much greater** than that at second-level



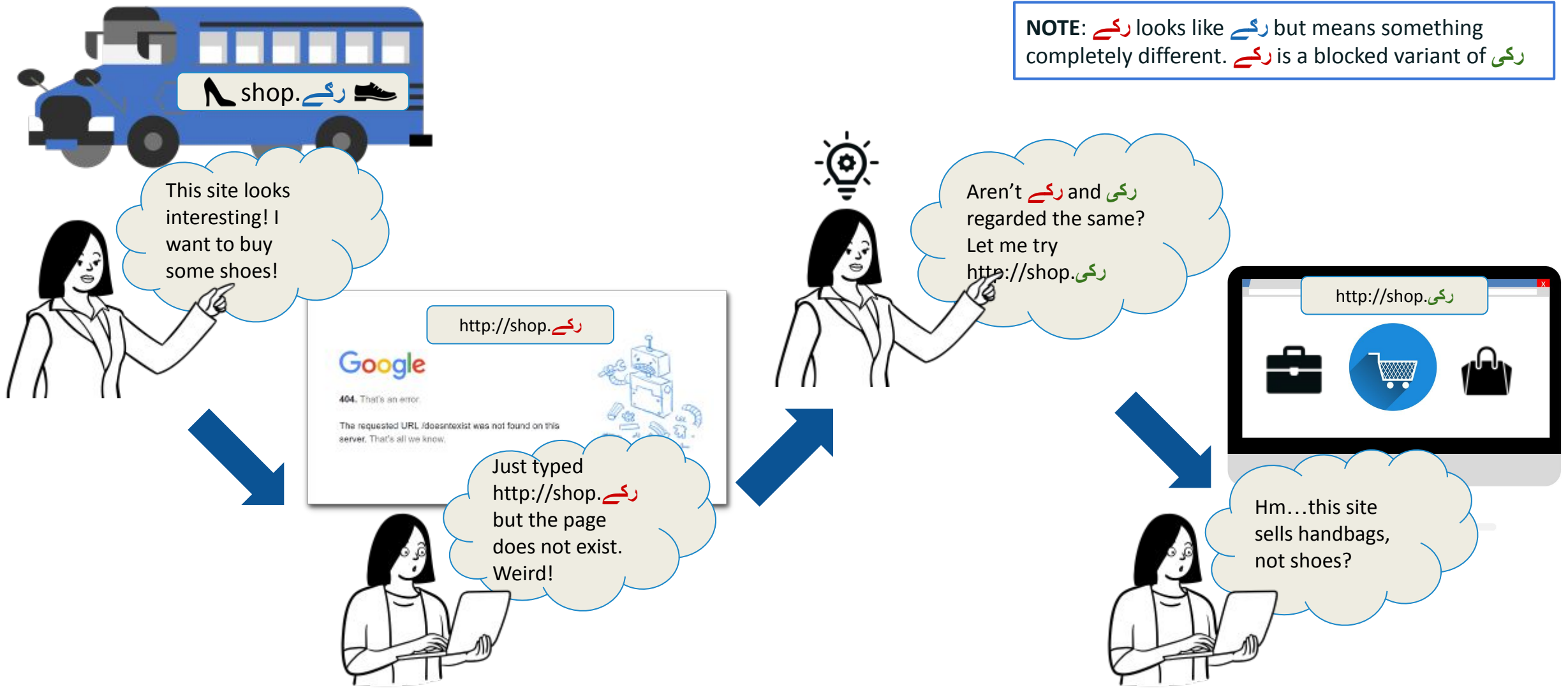
# Why Should Blocked Variants Be Part of String Similarity Review?

A label used in everyday life can be considered a blocked variant label by RZ-LGR calculation. End users can perceive and intend to access a blocked variant label domain name without knowing that it does not exist in the root zone

Variants are blocked by RZ-LGR for various reasons, but they still have a potential to cause confusion or even harmful consequences for end users. Reasons for blocking variants include, but not limited to:

- Apply more restricted rule to the root zone
- Reduce mutation issues
- Avoid overproduction issue
- Reduce redundant, infrequently used, or less desired characters
- Provide more flexibility for gTLD applicant
- Consider usability, correctness, manageability, and other factors

# Misconnection Involving Blocked Variants: Example and Illustration



# Next Steps

---

EPDP Team to determine if the hybrid model is the appropriate path forward, taking into account factors such as:

- ❑ Risk analysis – likelihood and impact of failure modes, especially the misconnection risk
- ❑ Operational impact of the hybrid model
- ❑ Cost and benefit of the hybrid model

# 2012 New gTLD Program Implementation Review

- InterConnect Communications and the University College London acted as the String Similarity evaluation panel
- The panel identified 234 contention sets, composed of 754 applications
  - 230 exact match contention sets (including ASCII strings and IDN strings)
  - 2 non-exact match ASCII contentions sets (.HOTELS/.HOTEIS; .UNICOM/.UNICORN)
  - 2 non-exact match IDN contention sets, which have variant relationship (.盛贸饭店/.盛貿飯店; .点看/.點看)
- **Evaluation results were published later than originally scheduled** (forecasted in Nov 2012, but not published until 26 Feb 2013)
- Delay due to volume of unique strings – 1,380 unique applied-for strings resulting in over 1 million combinations requiring review
- Results were released two weeks before deadline to file String Confusion Objection, leaving limited time to prepare an objection
- Community indicated dissatisfaction with the result in regard to singular and plural versions of strings, which were not found to be confusingly similar by the panel; several applicants filed String Confusion Objection to challenge the review results