## PROPOSAL FOR SMARTER NON-EXACT MATCHES
An Addendum submitted by: Greg Shatan

The RPM Working Group has been discussing a proposal to add non-exact matches to Sunrise and/or Claims.  So far, this has largely focused on "word-contained" matches, where the trademark string appears anywhere in the matching string.  As noted, these "dumb matches" raise a number of issues.  It may be possible to control for those issues.  This proposal does not deal with word-contained matches, the issues associated with them or with potential remedies for these issues, nor does it express an opinion on the advisability (or not) of adopting word-contained matches.  Instead, this proposal suggests another path.

Members of the Working Group have noted the lack of a formal (or even informal) proposal regarding other types of non-exact matches.  This proposal seeks to fill that gap, by proposing other, more tailored forms of non-exact match strings.  These could be adopted instead of, or in addition to, word-contained matches.  The basis for this proposal is not new.  Rather it is based on the GAC's May 2011 Advice to the ICANN Board.

This proposal recommends the incorporation of the following non-exact match criteria into the TMCH services to be used for Claims and/or Sunrise:

1. **Missing-dot typos**: These variations simulate an Internet user omitting a period in a domain address (e.g., www.domain.com becomes wwwdomain.com). In the first variation, "www" is appended to the beginning of the trademark string. In the second variation, "com" is appended to the end (e.g., [www.domain.com](www.domain.com) becomes www.domaincom.tld.
2. **Fat-finger Typos**: These variations take advantage of "fat-finger" characters (the characters immediately surrounding a character on the QWERTY keyboard). These variations simulate an Internet user accidentally hitting a nearby key when typing a domain name by replacing one character in a trademark string with each possible fat-finger character.
3. **Character Duplication**: For every character in the original string, a character is duplicated (i.e., "domain" becomes "ddomain," "doomain," etc.).
4. **Character Swaps**: For every adjacent pair of characters in the original string, their positions are switched (e.g., "domain" and "odmain").
5. **Character Removal**: One at a time, remove each character from the original string (e.g., "domain" becomes "omain, "dmain," etc.).
6. **Plurals**: An "s" is added to the end of the original string.
7. **Digit Addition**: A "1" is added to the end of the original string.
8. **"Cheap" and "Buy"**: "Cheap" is added to the beginning of each string and to the end of each string, respectively. The same is also done with "buy."
9. **Non-Latin Character Substitutions**: In Latin-character strings, one or more characters is replaced by a non-Latin character that appears similar to the replaced character(s) (e.g., a Latin character is replaced by a similar or identical character in the Bulgarian or Cyrillic alphabet).
10. **Latin Character Substitutions**: A character is replaced by one or more characters that appear similar to the replaced character (e.g., "w" is replaced by "vv", etc.).

11. **Goods and Services and Industry Keywords**: A keyword associated with the goods and services sold by the trademark holder, or with the trademark holder's industry, is added to the end of each original string (e.g., "apple-computer" for the trademark string "apple" registered by Apple, Inc.). Use of the descriptions associated with the Nice classification codes is not recommended for various reasons.[1] Similarly, wholesale adoption of goods and services descriptions in trademark registrations that are not directly derived from the Nice classifications is also not recommended. Rather, the recommendation is that a limited number of keywords be developed for each industry, using information from industry representatives, brand owners, watching and monitoring services, court and UDRP records, etc., that will focus on the type of keywords actually used in cybersquatting. These lists will need to be reviewed in a multistakeholder process.

12. **Commonly Abused Terms**: A commonly abused term, such as CAREERS, JOBS, or HOME is added to the end of each original string . A list of commonly abused terms would need to be developed; as with goods and services and industry keywords, ample resources exist from which these list can be created.

Many of these variations (1-8) can be developed mechanically. Variations 9-11 will require human intervention. One can assume that the resources available to ICANN will be at least as good as those available to cybersquatters who have developed these variations over the years. That is not to diminish the amount of work involved, since ICANN has to be rigorous, transparent and accountable – traits not required of cybersquatters.

If these proposals are well-implemented, problems of false positives should be minimized and even could even be eliminated. An exploration of the details is beyond the scope of this proposal, but can developed when the need arises.

**Rationale:**

Since the dawn of the domain name system, the DNS has been plagued by "cybersquatters"; individuals and entities that register domains related to trademarks in which they have no rights. They may register for the purpose of seeking to sell the domain to a trademark holder with rights in the domain, or they may register for more nefarious purposes, such as fraud, malware distribution, data breach, data theft, identity theft, phishing, spear phishing or the sale of counterfeit goods, among other things.

A system based solely on "exact match" excludes many types of "non-exact-match" domains used for these purposes – both those listed above, others not listed here, and others yet to be dreamed up. Yet these non-exact-match domains are a significant part of the problem.

---

[1] The descriptions associated with the two-digit Nice codes provide very high level industry characteristics that are not always product names that are likely to be included in domain names. For example, Nice Class 1 is described as "Chemicals used in industry, science and photography, as well as in agriculture, horticulture and forestry; unprocessed artificial resins, unprocessed plastics; manures; fire extinguishing compositions; tempering and soldering preparations; chemical substances for preserving foodstuffs; tanning substances; adhesives used in industry." Also the language of these descriptions is often formalistic.

This is not just a brandowners' problem.  It is a consumers' problem, it is a law enforcement problem, it is a financial services and retail industry problem, it is a criminal enterprise problem, it is a money-laundering problem.  It is a trust and security problem.

Since this is an addendum to the original "non-exact-match" proposal, this document does not contain all the information required of a free-standing proposal.  However, these can be developed if and when the need arises.