# Instant Search algorithm attempt

1. Grab search string.
2. Split into words on the space character.
3. Discard articles/superfluous words (The, it, a, etc).
4. Get 1-3 common synonyms for each search word.
5. With words and synonyms in hand, search database. Score each database entry by two metrics. User search words are scored more highly than synonyms.
   1. One metric is for if the entry has all the words in it.
   2. The other is for number of times the search words were found.
6. Sort the data by metric 4.b, then 4.a.

The scoring part of the algorithm will likely need tweaking to find the right values.

This approach also doesn’t take order of search words into consideration.

This approach does highly value just having all the search words in the data entry.

Rough Example:

The user’s search string is “Bee sting”.

The string is split up into words “bee” and “sting”.

According to datamuse.com, synonyms for “bee” are “honeybee” and “wasp”. Synonyms for “sting” are “bite” and “twinge”.

A page is found that has multiple occurrences of both “bee” and “sting”. It scores a 2 (out of 2) in matching words and a high score in quantity of occurrences.

Another page is found that has multiple occurrences of “wasp” and “sting”. It scores a 1.5 (out of 2) in matching words and also a high score in quantity.

A page with just the word “bee” would score a 1.