

Instant Search Bar

Jeffrey Slentz

Introduction

The project's goal is to gauge user response to finding similar words in the search results. User's responses could be used to inform a good UI design that made it clear what was happening with their search query.

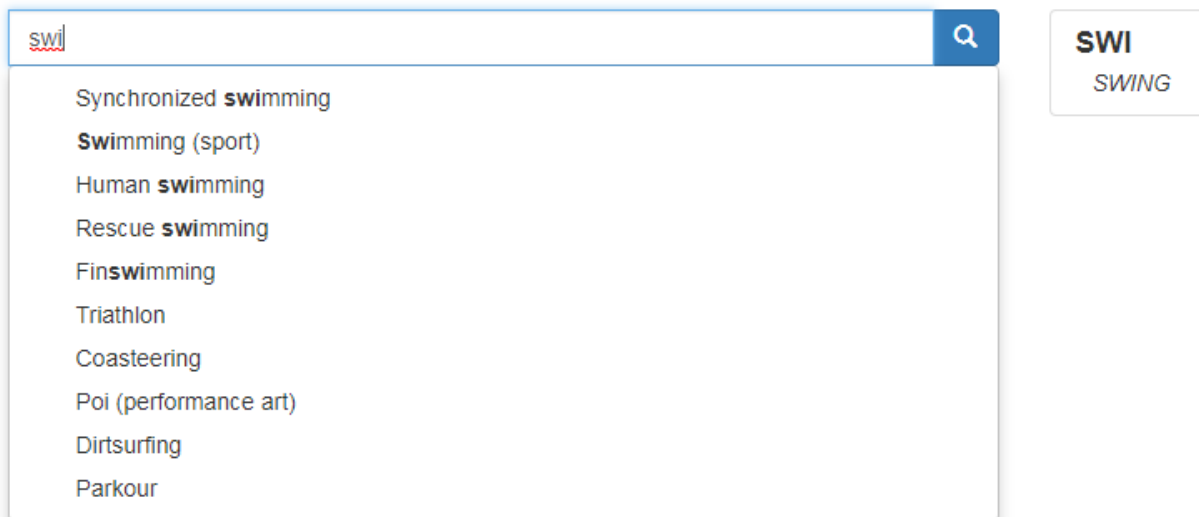
Design

The main functionality

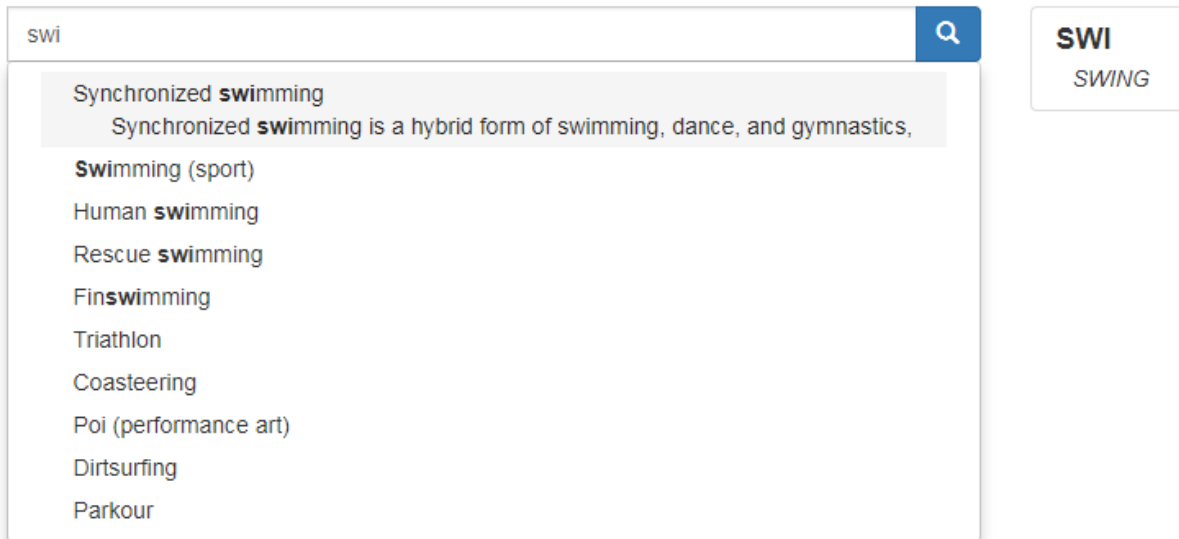
The webpage initially consists of one element: the search bar.

A horizontal search bar with a light gray border. Inside the bar, the text "Search Hobbies" is displayed in a light gray font. To the right of the text is a blue square button with a white magnifying glass icon.

As the user begins typing, the next two elements populate. The dropdown has the results and the suggestions box holds the words being used to score the results. In this case, an incomplete word swi is being autosuggested as swing.

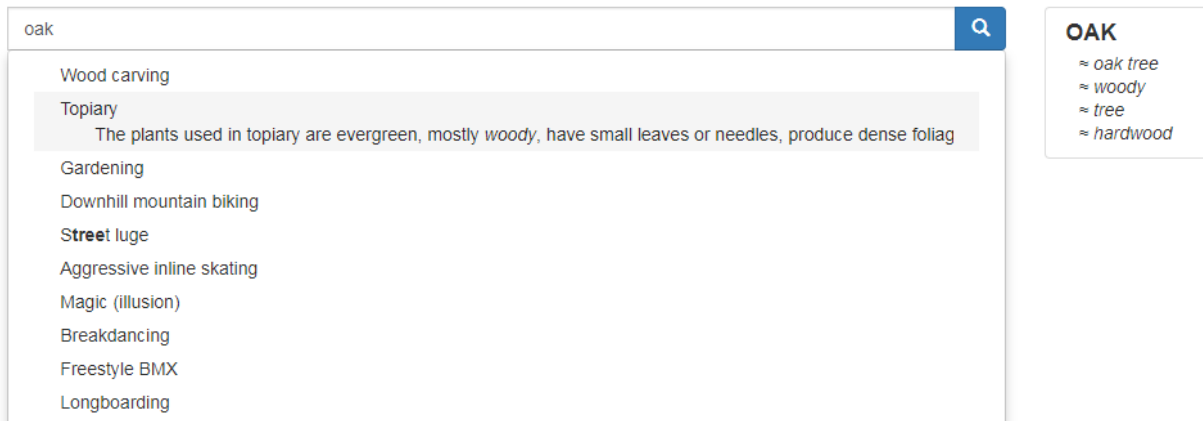
The search bar is now active. The input field contains the text "swi" with a red underline. A blue search button is to its right. Below the input field is a white dropdown menu with a light gray border, containing a list of suggestions: "Synchronized **sw**imming", "**Sw**imming (sport)", "Human **sw**imming", "Rescue **sw**imming", "Fins**sw**imming", "Triathlon", "Coasteering", "Poi (performance art)", "Dirtsurfing", and "Parkour". To the right of the dropdown is a white box with a light gray border containing the text "SWI" in bold and "SWING" below it.

If an element is hovered over, a snippet of the summary is shown.



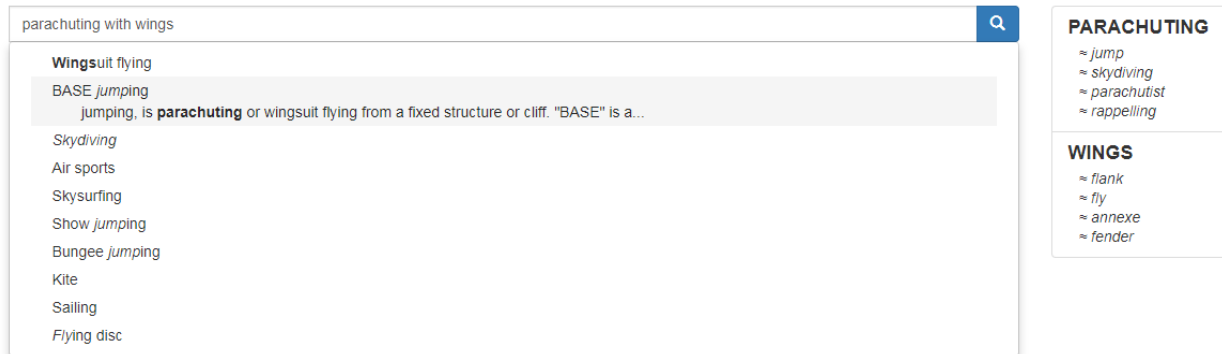
When a word is completed, the suggestions change from auto-complete to similar words. Sometimes, an entry only contains similar words. These are italicized instead of bolded.

(Btw, all those weird search results near the bottom? They have 'street' in them, which contains *tree*. This is not useful behavior for this search term, but may be useful for compound words like firetruck)



The bar can handle multiple words and ignores less useful words. Here, 'with' is ignored. The highlighted word in the summary seems arbitrary, but it's actually the word with the highest "score" value.

A direct match, 'parachuting', is worth more than the similar word 'jumping'.



How are similar and suggested words generated?

The similar and suggested words come directly from an API hosted by Datamuse.

Suggested: <https://api.datamuse.com/sug?s=word>

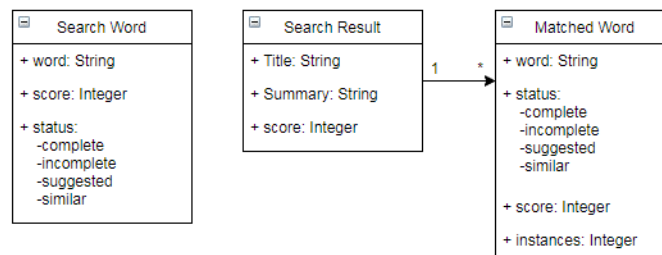
Similar: <https://api.datamuse.com//words?ml=word>

The algorithm

1. Grab the user's query.
2. Split it into words.
 - a. Discard less useful words according to a list from <http://www.stopwords.org>
3. Categorize the words.
 - a. Every word but the last is trusted as a "complete" word. (Unsafe, to be sure.)
 - b. The last is either "incomplete" or "complete", as determined by the results from <https://api.datamuse.com/sug?s=word>
4. Generate related words.
 - a. "similar" words are generated from sending "complete" words to <https://api.datamuse.com//words?ml=word>
 - b. A "suggested" word is generated from sending "incomplete" words to <https://api.datamuse.com/sug?s=word>
5. Assign scores to the words, to simulate relevancy.
 - a. "complete" words are worth the most.
 - b. "similar" words are worth almost as much as complete words.
 - c. "incomplete" words are not worth very much.
 - d. "suggested" words are worth even less.
6. Each word found in a data entry adds that word's score to the entry's cumulative score according to a few rules:
 - a. The first word found is worth more than the next words.
 - b. If the first character matches, the entry receives a lot more points. (To reduce the 'tree is in street' problem.)
 - c. Title matches are worth more than summary matches.
7. Sort the data entries by their score and display them to the user with matching words highlighted.

Data structures

Three main objects are used to hold information.



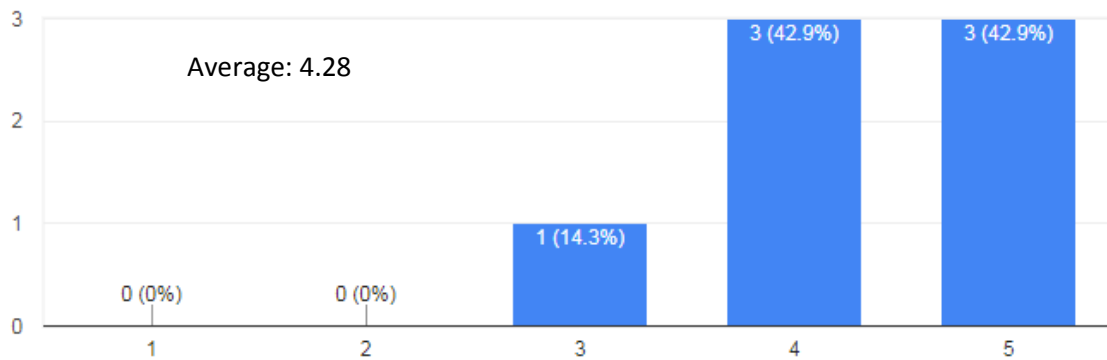
User testing

The user test consisted of a google form with about 13 questions, 5 of them optional and open ended. The questions are intended to determine if the similar words feature is useful or not.

Question 1

On average, were your search results relevant?

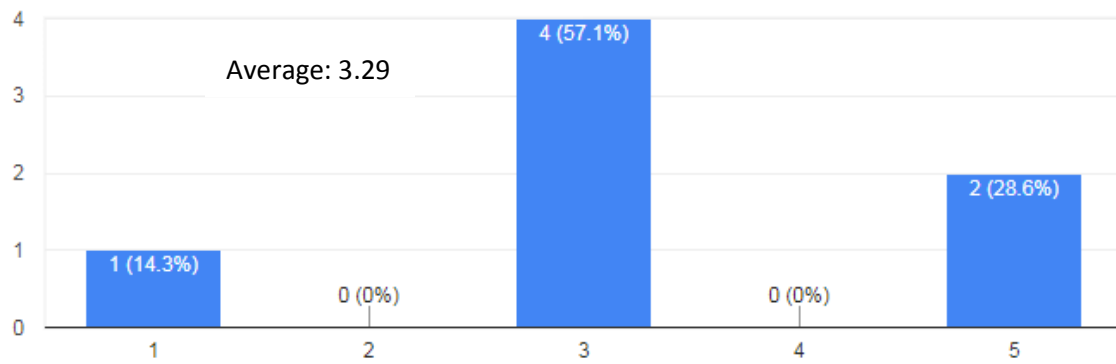
7 responses



Question 2

Was the suggestions box on the right helpful?

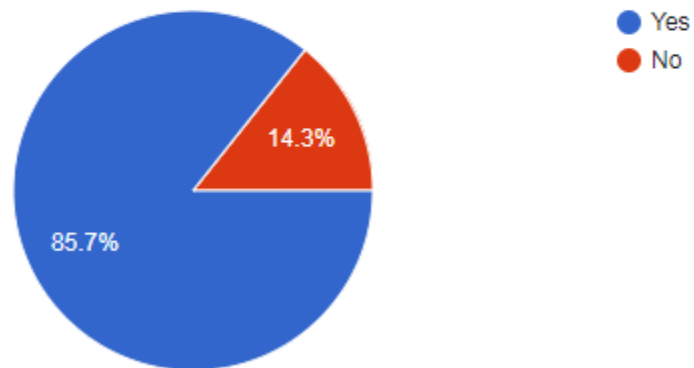
7 responses



Question 3

Did you find any italicized words?

7 responses

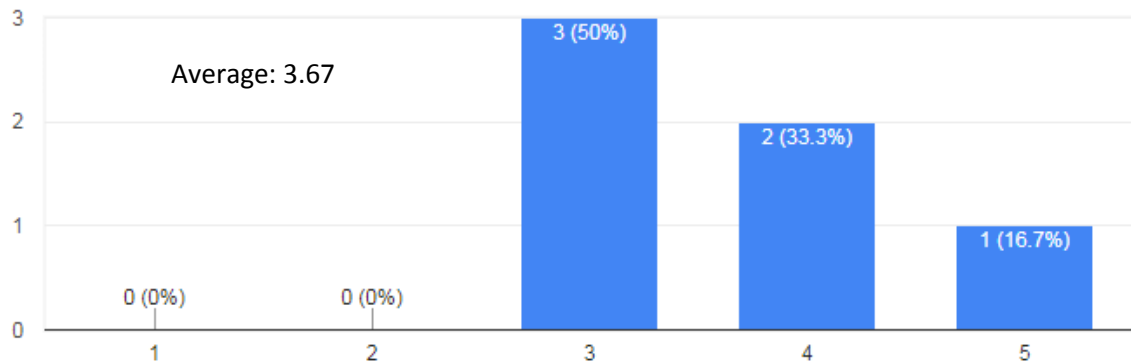


Question 4

This section was only answered by people who answered yes to the previous question.

On average, were the italicized words meaningfully related to your original query?

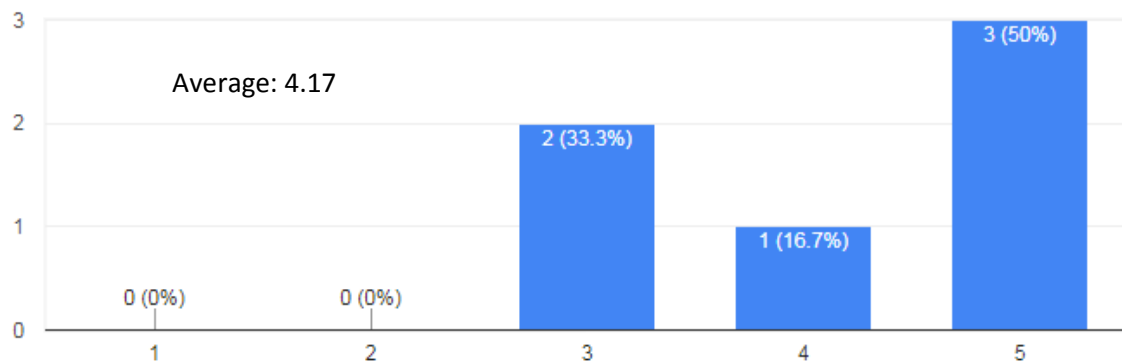
6 responses



Question 5

How likely is it that a user would recognize the difference between bold and italicized words?

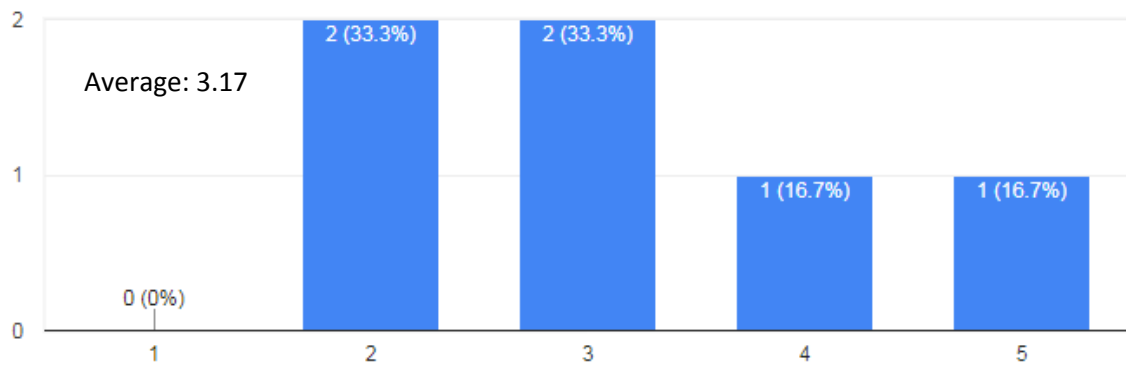
6 responses



Question 6

Is the distinction between similar and exact matching words helpful?

6 responses



Question 7

Is there a better way to visually mark the difference? (Optional)

3 responses

Maybe have "Do you mean:" and then the italicized words.

Different colours

I'm not sure how you would make it look that clean but more intuitive.

Open Ended Questions

How would you improve the search bar?

3 responses

maybe clean up suggested words

Probably raise the threshold for relevance - some of the lower-down search results weren't very relevant.

Mainly aesthetic and clarity would be improved. Something saying that these are synonyms or like that.

Did you get any astoundingly wrong search results? What were they?

4 responses

Yes. Fisting turned into Fantasy sport.

it knew weed=pot and suggested pottery

I searched "reading" and lower down I got "candle making" after typing in "read" and pausing to look at the results - the article for candle making had the word "readily" in it.

No

Did you get any surprisingly good search results? What were they?

3 responses

above, no bad ones

I searched climbing and a lot of good relevant results came up for that.

Frolf - Frisbee Golf was not expecting that.

Weird or interesting results?

3 responses

suggested words knows lingo and cussing

Searched origami (one of DJ's favorite hobbies) and the italicized options weren't very close! Also after searching climbing I tried searching "canyoneering" and got first, "mountaineering" and "mountain biking" (in italics) and then after that "canyoning" (the most relevant one, and not italicized).

There are street luges that look super dangerous.

Add any additional feedback you'd like to give here:

2 responses

impressive search bar

The suggestions box on the right (climbing ≈ ascending, mounting, scandent, jumping) didn't seem like it was providing useful additional information. But I'd be curious to talk about the analysis that went on behind the scenes. Overall good job!! The design is really clean and simple and the tool works quickly and smoothly.

Positives

Search results on exact matches are at the top and usually very relevant, so it does exact matching very well.

The design is pretty clean.

Because of the similar words, any search result is likely to return *something*.

After a minute or two, it seems like most users were able to intuit what the search bar was doing.

The instant results are pretty quick unless the search query becomes too long. This could be fixed with caching the results of previous words.

Negatives

General Design Problems

Without good search results, user tests would likely not be very useful. So, in the desire to build a good search algorithm, the project quickly became mired in the technical approach to displaying well sorted entries by both the actual search query and similar search queries. Search is complex and it took a lot of time to reach a design that was useful at all.

The choice of data to search introduced another difficulty. The original intention of the similar queries approach was to be used on a set of frequently asked questions or help topics. This might allow

many opportunities for a search query to be close but not quite match a topic. However, with the list of hobbies as the data set, opportunities for similar queries to be useful was limited, as each hobby's title was usually one word only. This is all to say that finding a hobby by name is more intuitive than finding it by related/similar words.

Showing the user what was happening with their search query was time-consuming to implement and never quite reached a crystal clear level of unobtrusive yet useful information.

The Datamuse API's similar words functionality is cool, but sometimes it returns strange, less than useful results. It's probably more useful if the data entries have keywords that contain similar words already. Then those can be a part of the search results. Auto-generating the keywords will be difficult, however.

Specific Functionality Issues

Because the algorithm will only search for similar words once it has deemed a word "complete" the results will tend to jump around.

There are some bugs where the summary doesn't get sliced correctly.

The API requests are synchronous, so the UI will have increased response times if either the number of requests becomes large (5+ "complete" words) or the Datamuse API is responding slowly.

Conclusion

The search works for the most part and I'm pretty happy with it. It shows you what it's searching for and highlights what it found. The user responses seem positive overall. The most interesting result were the answers to Question #6: "Is the distinction between similar and exact matching words helpful?" It only scored a 3.17 out of 5 on average, which implies that either there's a better way to highlight the difference between similar and exact words, or that this highlighting isn't actually necessary. Overall, it was a fun and challenging project and while the original goals weren't met, I'm proud to have built a functional, if quirky, design.