# Word Trees for Visualizing PubMed Search Results

**Edwin Sperr**, MLIS, AHIP (esperr@uga.edu)
Augusta University/University of Georgia Medical Partnership
Athens, GA, USA

AUGUSTA UNIVERSITY
MEDICAL PARTNERSHIP
UNIVERSITY OF GEORGIA

## Rationale

Starting the search process for a complicated question can be daunting, particularly if a searcher is new to a topic and not yet familiar with the language used to describe it. Searches in databases such as PubMed often yield result sets of several hundred citations or more – the challenge is less a matter of finding something than it is making a search specific enough. Therefore, it would be helpful if a user could begin by finding some way to summarize the results of an initial search and see the context into which individual search terms fit.

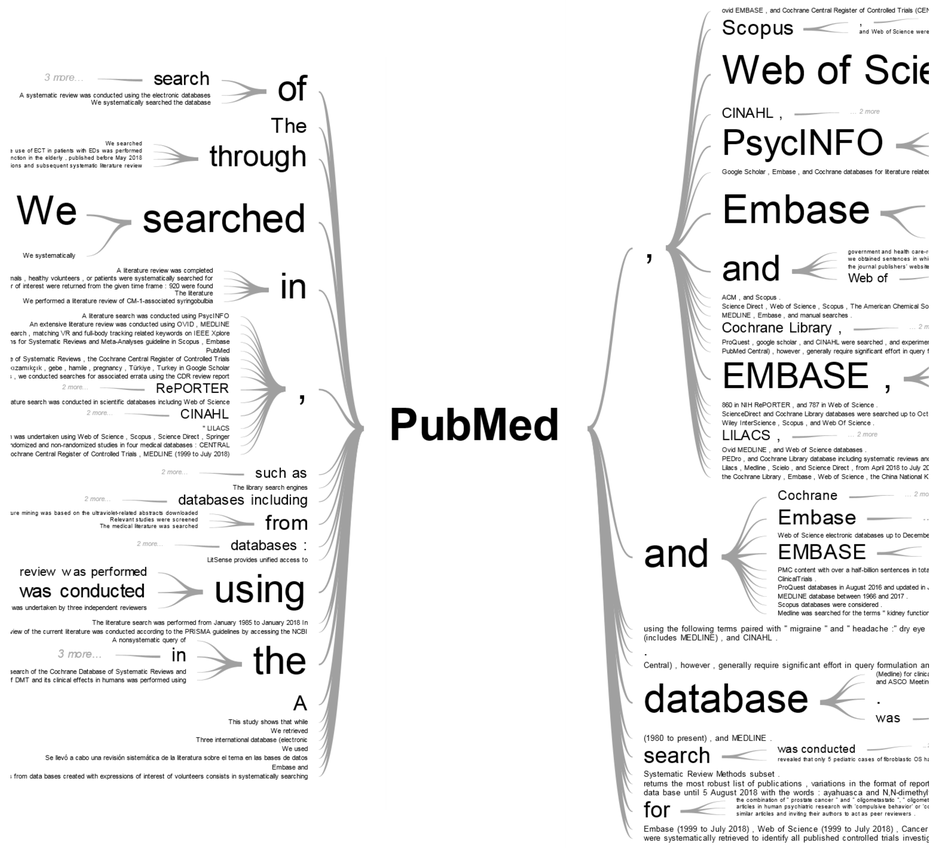
One approach is to use a visualization technique known as Word Trees. Word Trees were developed by Martin Wattenberg and Fernanda Viégas during their work on IBM's Many Eyes platform[1]. As with the more commonly seen Word Clouds, Word Trees show the frequency of the terms they represent by representing them with different sizes. Crucially, they also (like the Keyword-In-Context technique) show them in the context of the words that adjoin them. This method has been previously described in the biomedical context[2] and there are extant web interfaces for viewing individual documents as Trees[3], but there is not currently a straightforward way for an end-user to use Word Trees to visualize a PubMed search.

## Methods

The web application described here (accessible at https://esperr.github.io/pub-trees/) takes a user's search and interactively runs it against PubMed via NCBI's E-utilities API[4]. Using that same interface, it then retrieves the titles and abstracts from the first fifty, hundred, or one hundred fifty results and aggregates them all together. Once a user designates a single word to serve as their Word Tree "Root", a JavaScript regular expression isolates those sentences which contain that word. Finally, those sentences are graphed using the version of Word Trees that is hosted on Google Charts[5].

The resulting diagram is displayed on the page in an interactive fashion – selecting a term will redraw the Tree with that term as Root. Users are also free to manually select a different term around which to redraw the Tree.

Try it yourself!
https://esperr.github.io/pub-trees/

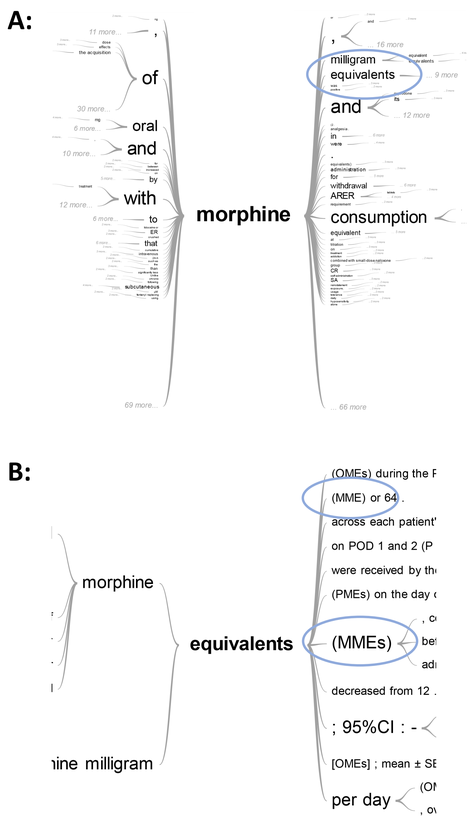



## Results

Simple searches often show results that one might predict – "breast" is commonly followed by "cancer", and those two words are most often succeeded by "cells". However, even simple searches can yield insights if one takes advantage of the interactive nature of the Google Word Tree interface. A search for "morphine" shows "equivalent" as a common succeeding word. Selecting that quickly leads one to the term of art, "MMEs". **(*Fig 1*)**

This example demonstrates how a searcher using this tool might uncover useful terms for search strategy refinement even before they begin examining their first set of results.

**Figure 1: Using Word Trees to explore related terms**




## References

1. Wattenberg M, Viégas FB. The Word Tree, an interactive visual concordance. IEEE Trans. Vis. Comput. Graph. 2008;14:1221–8. DOI: http://dx.doi.org/10.1109/TVCG.2008.172.

2. Harris DR, Kavuluru R, Jaromczyk JW, Johnson TR. Rapid and Reusable Text Visualization and Exploration Development with DELVE. AMIA Summits Transl. Sci. Proc. 2017;2017:139–48.

3. Davies J. Word Tree [Internet]. [cited 2019 Apr 10]. <https://www.jasondavies.com/wordtree/>.

4. What is E-utilities? [Internet]. [cited 2019 Apr 22]. <https://dataguide.nlm.nih.gov/eutilities/what_is_eutilities.html>.

5. Word Trees | Charts | Google Developers [Internet]. [cited 2019 Apr 22]. <https://developers.google.com/chart/interactive/docs/gallery/wordtree>.

Medical Library Association
2019 Annual Meeting
Chicago, IL
**Poster #172**