MeSHgram: An Open Source Tool to Visually Browse Co-occurrence of MeSH Terms in PubMed

Satyajeet Raje, PhD, Ravi Teja Bhupatiraju, MBBS, PhD, Abdelrehman Hosny, MS, Ben Busby, PhD

National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland

Background

MeSHgram is a tool for visual and interactive exploration of co-occurrence of Medical Subject Heading (MeSH) terms in PubMed. It uses the MeSH terms associated with PubMed articles to visualize co-occurrence over time. The potential applications are:

- Visual browsing / querying of PubMed
- Support metadata analysis of literature in PubMed
- Hypotheses generation

Methods

We parsed the NLM PubMed corpus[1] and extracted the ID, year of publication and MeSH terms associated with each document. As of Jan 2017, the corpus had approx. 24.5 million publications from 1809 to 2016[2]. We excluded duplicate items such as revision entries and those with no MeSH terms, resulting in approx. 23 million publications.

We use a simple architecture (Figure 1) for implementation. The extracted data is stored in a **MongoDB** object store. We use a Python web controller (**CherryPy**) and JavaScript (**nvd3, jqcloud**) for the front-end.
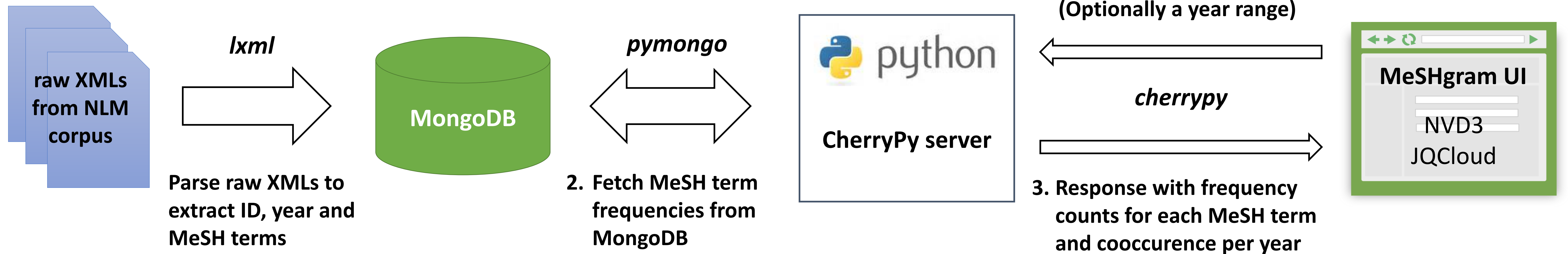
Figure 1. MeSHgram conceptual architecture

Figure 2. Screenshot of MeSHgram UI illustrating different features. Showing co-occurrence of "Rubela" and "Rubela Vaccine".
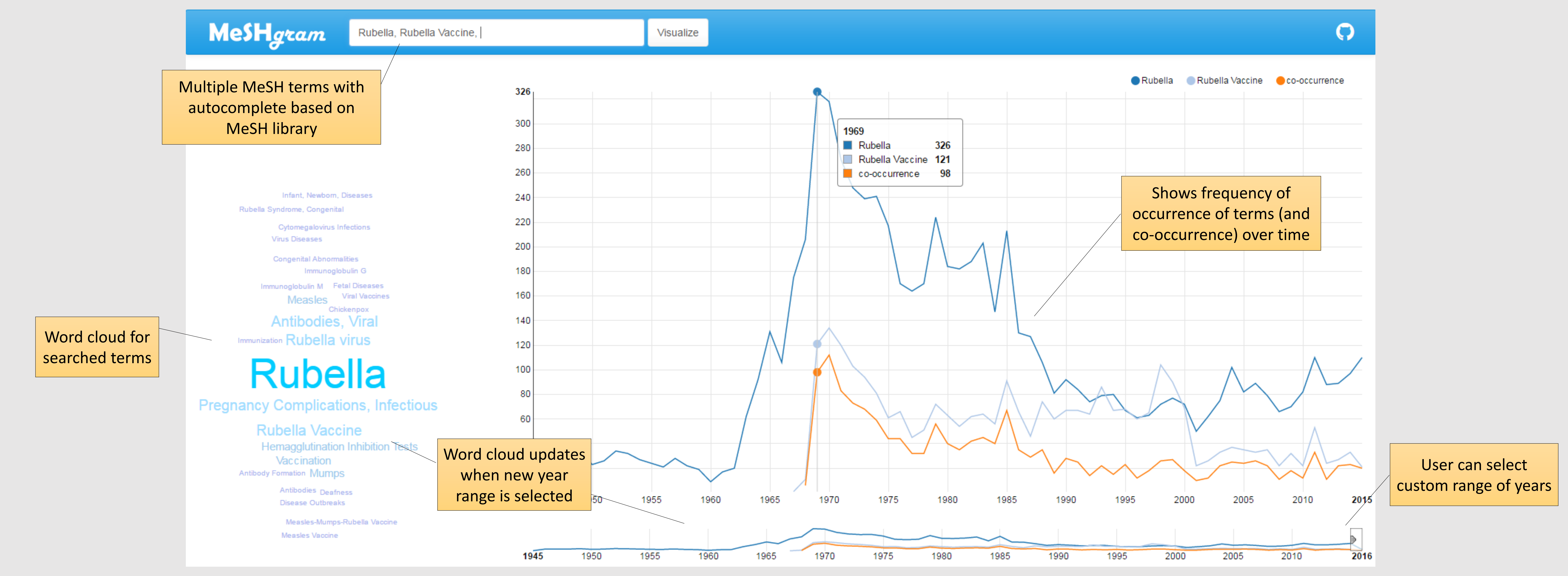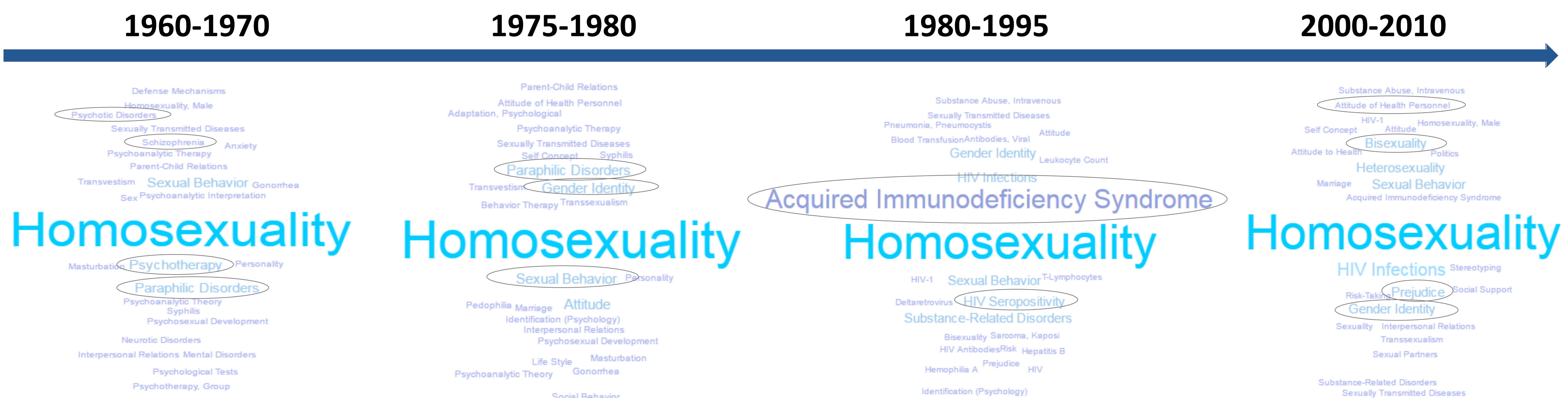

Figure 3. Example retrospective study. Evolution of Homosexuality in medical literature from a MeSH perspective.

References

1. https://www.nlm.nih.gov/databases/download/pubmed_medline.html
2. https://www.nlm.nih.gov/bsd/licensee/2016_stats/baseline_med_filecount.html

Looking for a new home!
Code is available at https://github.com/NCBI-Hackathons/Visualizing_MeSH_Term_Interaction_Over_Time

NIH U.S. National Library of Medicine