

SciTrend: Visualizing scientific trends in publications

Mimi Yang, and Benjamin Wang
Stanford University



MOTIVATION

We have developed a system and process of visualizing scientific publication trends use the PubMed API to access recent publications concerning a seed keyword. A histogram of the number of papers published per year related to the seed keyword demonstrates the growth and maturation of that material or technique in the bio-medical field.

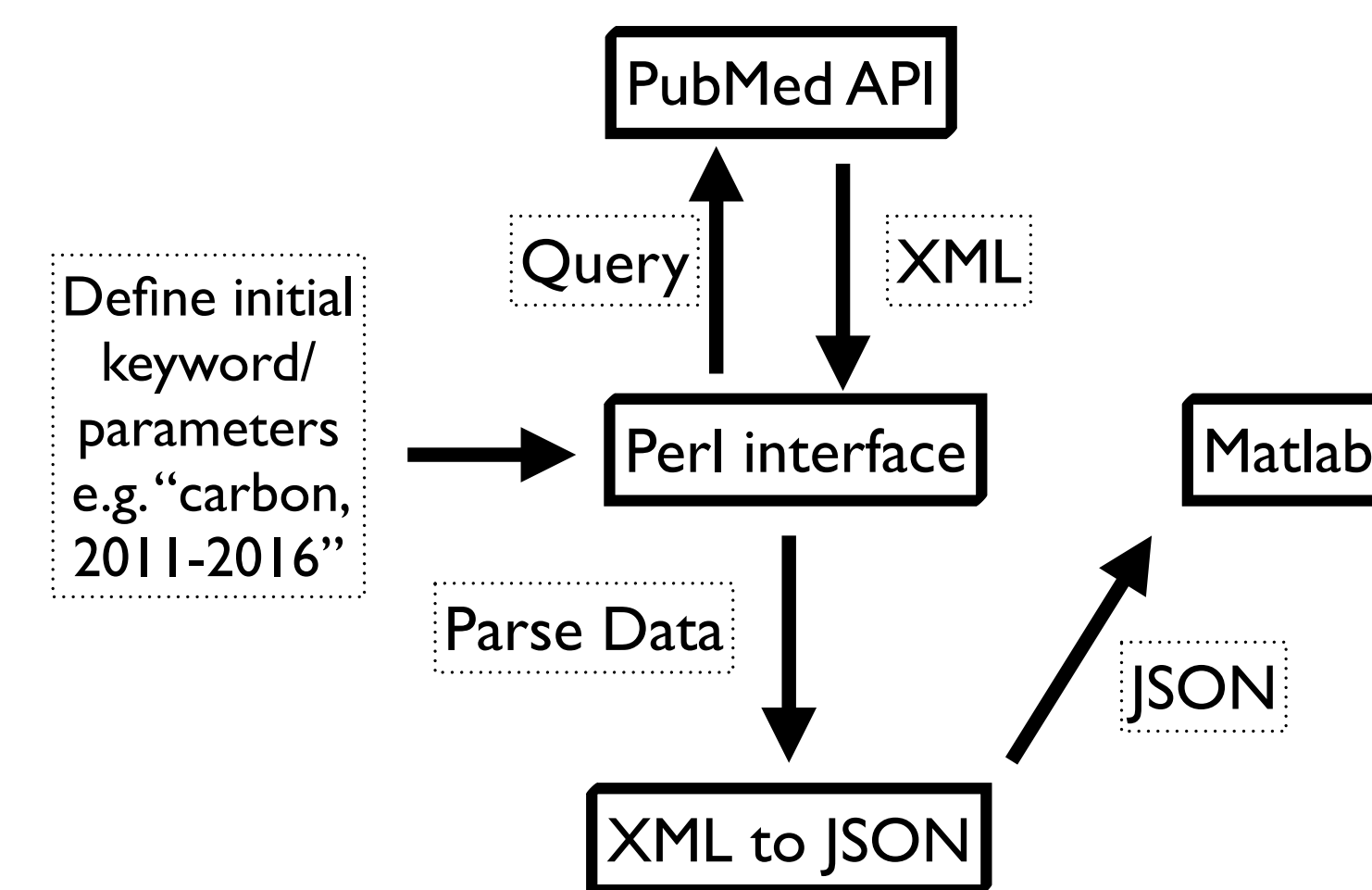
Further understanding of research trends is obtained by parsing the publication titles and enumerating the associated words by time, showing the temporal popularity of terms associated with the seed keyword. Advanced analysis can be conducted by examining relative popularity of common title terms per year using TreeMaps. The outlined method guides a researcher in understanding and discovering the history and development of past research, popular current research directions, and future applications in a chosen research field.

PREVIOUS WORK



Tools such as Citeology[1] and CiteRivers[3] organize the papers temporally. This enables the inclusion of individual points for each paper along the time axis, allowing for further analysis of specific papers. Laying out the publications along the x-axis permits addition of a relational notation along the y-axis.

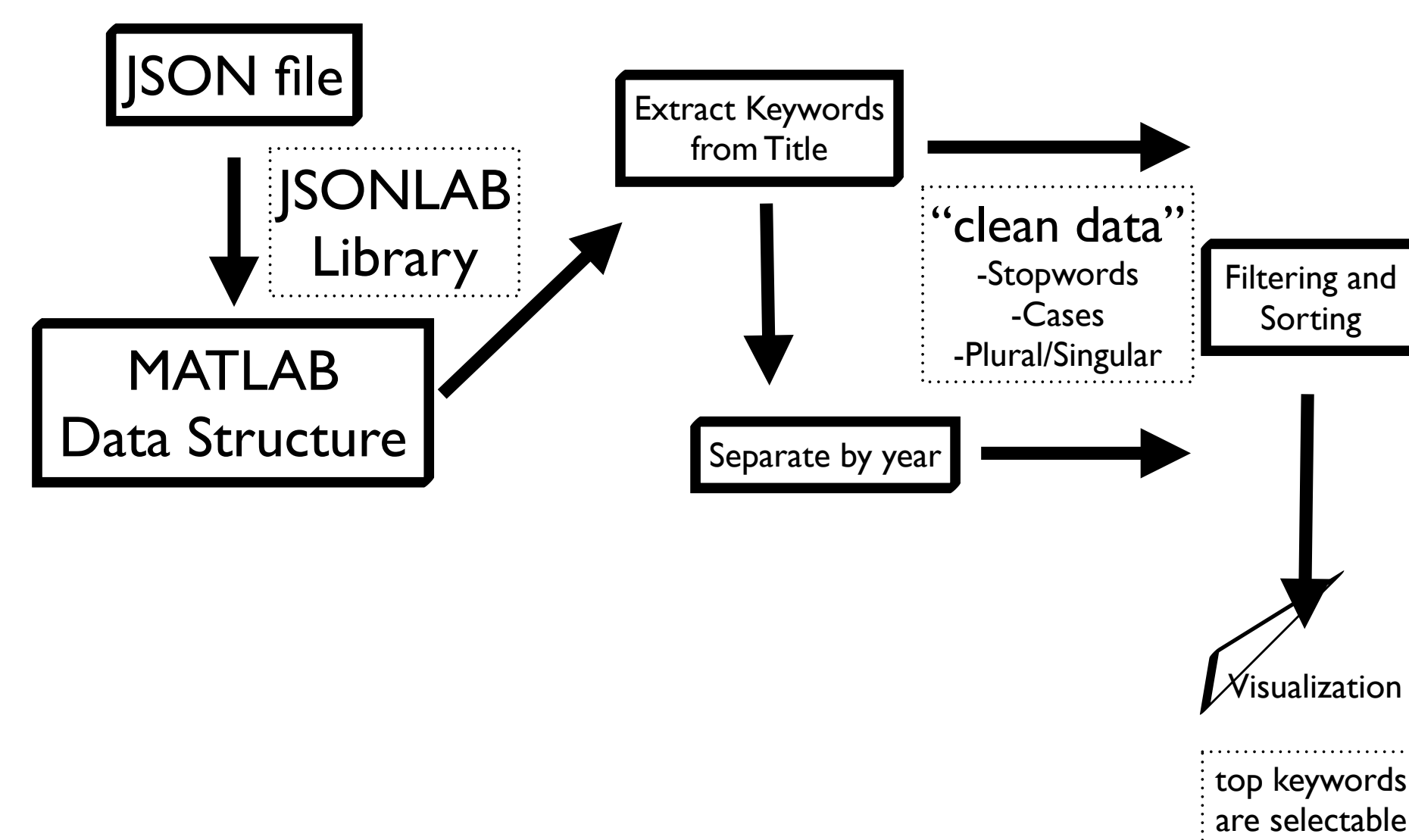
METHODS - EXTRACTION



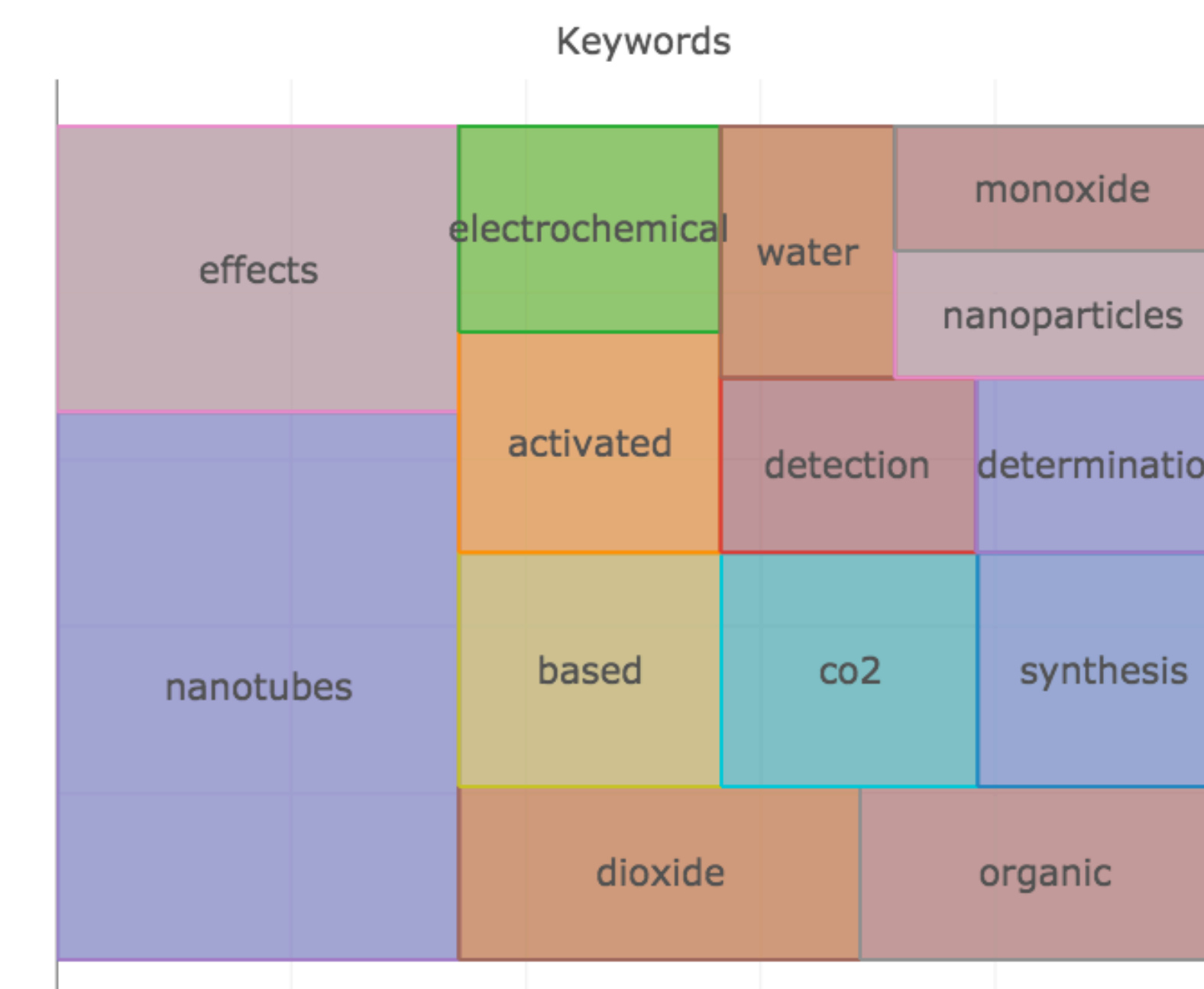
The National Center for Biotechnology Information (NCBI) provides access to a plethora of biomedical and genomic information.

We refine the list of publications included in the visualization by querying according to publication keywords (known as other text [OT] in PubMed). As the authors are responsible for selecting keywords that they feel are representative of the core concepts and innovations of their paper, we argue using keywords as a refinement metric produces a list of relevant and important work.

METHODS - PROCESSING

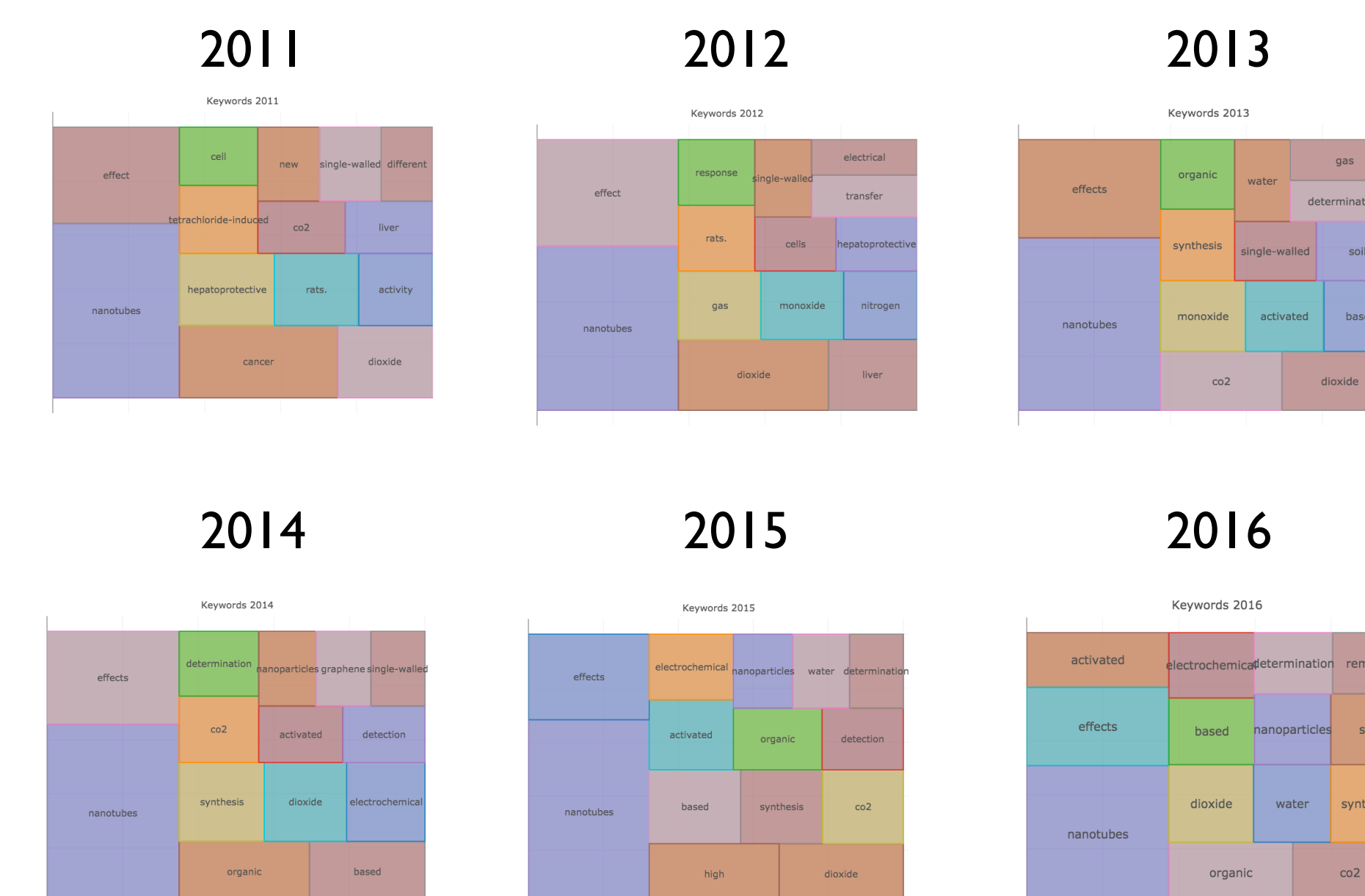


KEYWORD TREEMAP



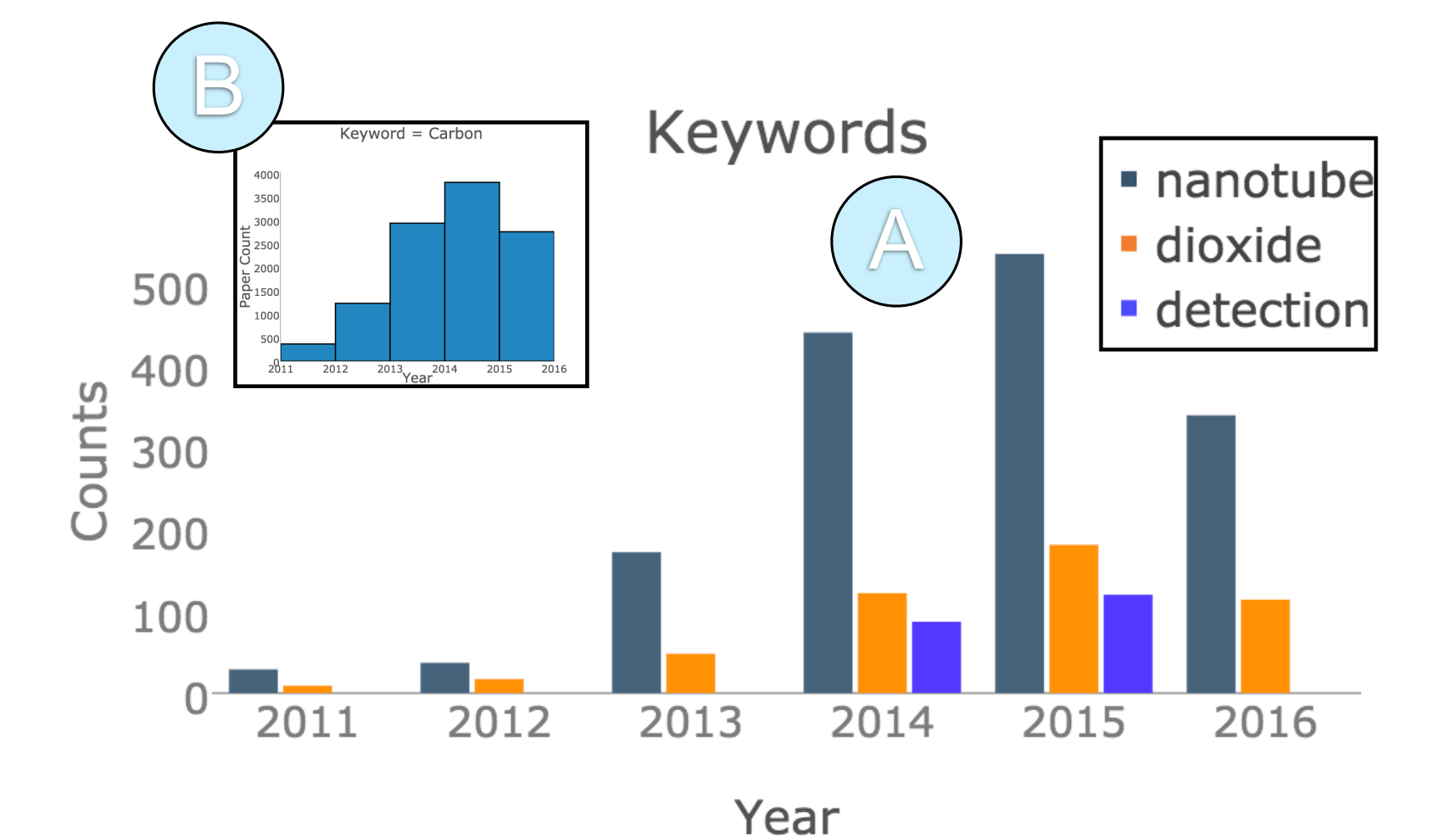
Each of our TreeMaps had a total of 15 bins. It is easily concluded that the top two sub-keywords for the entire time period are "nanotubes" and "effects".

YEARLY KEYWORD TREEMAP



Comparing TreeMaps by year, which gives the user a sense of how various trends change over time. At this level the keywords are not weighted and are simply shown as a number count. It would be interesting to weight the keywords based on search popularity, journal impact factor, or another metric that would allow for user control of weights.

KEYWORD HISTOGRAM



We have a histogram interface that has a main histogram(A) and sub-histogram(B). Comparing individual keywords between TreeMaps is rather difficult due to difficulties comparing area between maps. We have implemented a histogram/bar graph representation that shows various keywords as a function of year.

CONCLUSIONS

The designed method and visualizations guide a researcher in understanding and discovering the history and development of past research, popular current research directions, and future applications in the various research fields.

REFERENCES

- [1] J. Matejka, T. Grossman, and G. Fitzmaurice. "Citeology: visualizing paper genealogy". Proceedings of ACM CHI '12 Extended Abstracts, pages 181-190. ACM, May 2012.
- [2] J. Stasko, J. Choo, Y. Han, M. Hu, H. Pileggi, R. Sadanaand and C. D. Stolper, "CiteVis: Exploring conference paper citation data visually", Posters of IEEE InfoVis, 2013.
- [3] F. Heimerl, Q. Han, S. Koch and T. Ertl, "CiteRivers: Visual analytics of citation patterns", IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 01, 2016.