Paleontology Topic Trends

May 01, 2018
CS4624: Multimedia, Hypertext, and Information Access
Instructor: Edward A. Fox
Virginia Tech, Blacksburg, VA 24061
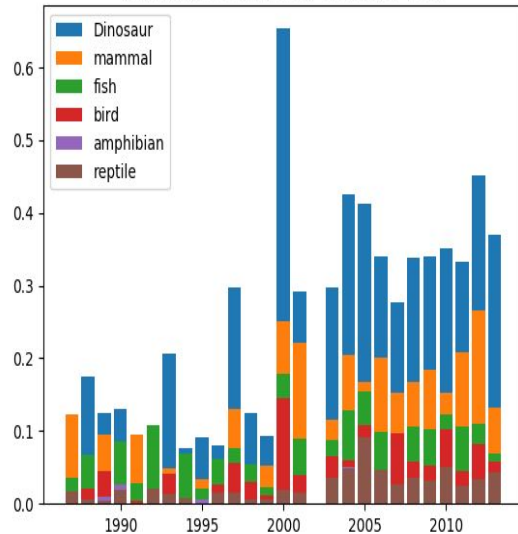James Wilson, Joseph Martin, Eric Weiler, Rudy Cruz

Outline

1. What We're Doing
2. Stacked Bar Graphs
3. Our Experience
4. Future Work
5. Acknowledgements

What We're Doing

- Task: Determine trends in paleontology

- Resources: PDFs containing abstracts for new paleontology publications presented at international paleontology conference (1987-2017)

- Work: Text mining the abstracts and building word counts to perform analysis on.
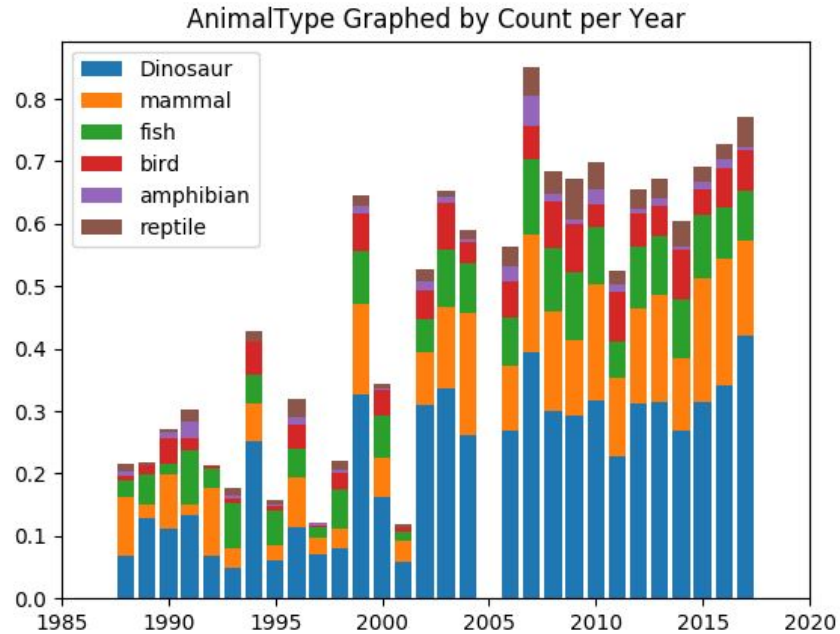
Stacked Bar Graphs

Stacked Bar Graph Cont.

AnimalType Graphed by Count per Year
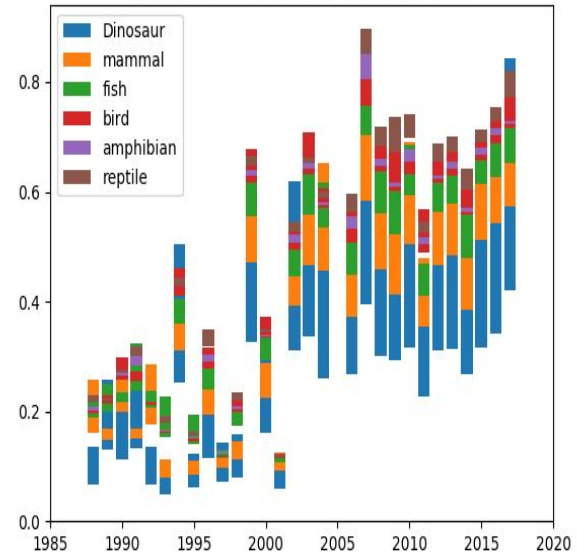
Our Experience

Learned to be client centric

Learned over 5 new Python frameworks

Fostered a better understanding of automated textual analysis

Future Work

- Extract, analyze, and present data relating to the authors of the publications
    - How authors relate to each other
    - How authors relate to a particular subject
    - Show demographics of the authors
    - Can be represented as a graph data structure

- Integrate the code into a web application
    - Users submit their abstracts to the app
    - Users can do it all themselves instead of submitting to devs and waiting for a result
    - Users can customize report parameters and stop words

Acknowledgements

Client: Our client is Professor Sterling Nesbitt, a professor in the geosciences college, with a focus in Paleobiology. Our project is focused around abstracts created by the Society of Vertebrate Paleontology, and the information within them.

References

- http://vertpaleo.org/Annual-Meeting/Abstracts.aspx
- http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf
- http://www.simonlindgren.com/stuff/2017/4/14/topic-modelling-with-mallet
- https://en.wikipedia.org/wiki/Stemming

Questions?