

Bookworm: a grammar for text analysis and visualization

Benjamin Schmidt

November 12, 2014

1. Acknowledgements
 - (a) Acknowledgements
 - (b) Description

**platform for the analysis of large textual collections
general users,**
2. Reading Digital Libraries at scale.
 - (a)
 - i. A. At scale means metadata.

before
 - (b) Texts
 - i. Flagship Browser

This is not an advance in visualization
 - ii. Our other browsers
 - iii. Newspapers: Investigating sources
 - A. Presidents example
 - iv. Partner browsers
 - A. Yale University Libraries
 - B. Medical Heritage Library
 - C. US State Department
 - D. Hathi
3. Core Philosophy: texts and metadata.
 - (a) What should the infrastructure for text analysis look like?
 - i. The current state:

- ii. What it should be:
 - A. Librarians as the collaborators
 - 4. A generative grammar for texts
 - (a) Core functions of the grammar.
 - i. 1. Creating a corpus
 - ii. 2. Setting a working definition of texts and tokens
 - A. First, Texts.
 - B. Gender and language
 - C. Different kinds of multivariate approaches
 - useful even without words:**
 - iii. 3. Performing an operation that compares counts in two corpora against each other.
 - A. Text Length
 - B. Log-Likelihood
 - iv. Returning and comparing results
 - 5. Extensibility through a grammar
 - (a) Topic Models
 - i. TEI tags as metadata
 - ii. Geocoding
 - 6. A platform for existing research
 - 7. Conclusion
- infrastructure**