

Text Mining and STEM Education

A Brief Introduction & Demo

Dr. Shaun Kellogg
2021/08/11 (updated: 2021-08-11)



Overview

1. Why text mining and **what is it** exactly?
2. What are **common techniques** used for mining text?
3. What are some of the **technical and ethical considerations**?
4. How is text mining **used in STEM education** research?

1. Why text mining and what is it?

Why should I care?

There has been **unprecedented increase in text-based data** generated by educational processes and digital learning systems, resulting in...

New **sources** of data

- Discussion Forums
- Online Assignments
- Instant Messaging Tools
- Social Media
- Etc.

Potential **advantages** for analysis

- Massive
- Always On
- Non-reactive
- Social relations
- Etc.

Krumm, A., Means, B., & Bienkowski, M. (2018). [Learning analytics goes to school: A collaborative approach to improving education.](#) Routledge.

What is text mining?

According to my graduate students...

Text mining **is**...

- a process for gaining insight into large amounts of text
- an exploration of text in search of patterns
- like computer-assisted reading
- magic

Text mining **is not**...

- a substitute for traditional qualitative analysis
- as "automated" as one would think
- unlike reading tea leaves
- magic

Fesler, L., Dee, T., Baker, R., & Evans, B. (2019). [Text as data methods for education research](#). Journal of Research on Educational Effectiveness, 12(4), 707-727.

A central question in text mining is:

How do we **quantify** what a document is about?

Silge, J., & Robinson, D. (2017). [Text mining with R: A tidy approach](#). " O'Reilly Media, Inc.".

2. What are common techniques used for quantifying text?

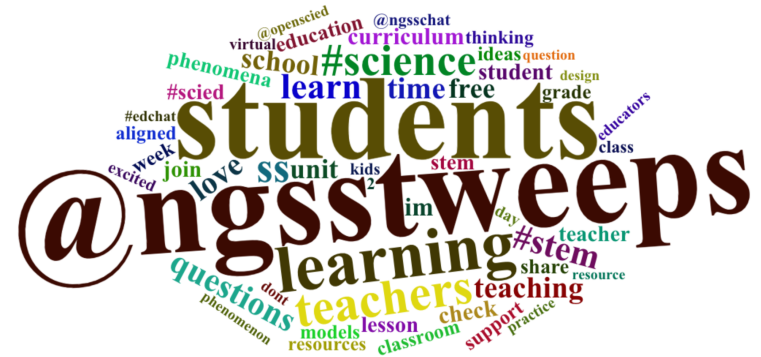
Basic Text Analysis

... it's a mistake to imagine that text mining is now in a sort of crude infancy, whose real possibilities will only be revealed after NLP matures.

Wordcounts are amazing!

- Text Preprocessing
- Word Counts
- Term Frequencies
- TF-IDF

Yep. A Word Cloud.



Underwood, T. (2013). *Wordcounts are amazing*. The Stone and the Shell.

Dictionary-Based Methods

Dictionary-based text analysis uses predefined list of words, or **lexicons**, to assign a particular meaning, value, or category to each word in your data:

- Custom Dictionaries
- Stop Words
- Sentiment Lexicons
- Linguistic Inquiry Word Count (LIWC)

Bing Lexicon Example:

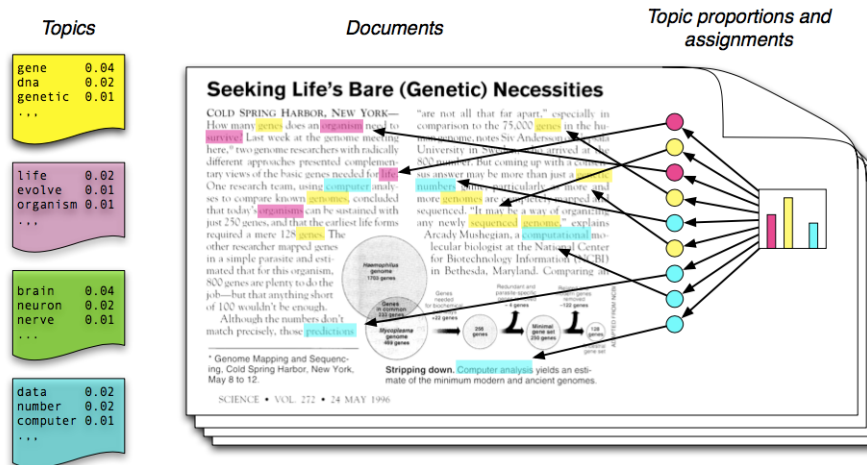
word	sentiment
saintly	positive
excruciating	negative
disdainful	negative
ineffectively	negative
hysteric	negative

Bail, C. (2019). [Dictionary-Based Text Analysis in R](#). Text as Data.

Topic Modeling

With a bit of tongue-in-cheek, Meeks and Weingart describe topic modeling as:

leveraging occult statistical methods like 'dirichlet priors' and 'bayesian models'... to provide seductive but obscure results in the form of easily interpreted (and manipulated) 'topics.'



Latent Dirichlet Allocation

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. Information Systems, 94, 101582.

3. What are some of the technical and ethical considerations of text mining?

Text Mining Considerations

Despite the potential advantages of text-based data captured by educational technologies, **TM poses a number of challenges** for STEM Ed researchers.

Logistical & Technical

- Unstructured
- Inaccessible
- Non-Representative
- Incomplete

Ethical & Legal

- Bias (algorithmic, positivity)
- Sensitive
- Terms of Use
- FERPA

Bail, C. (2019). [Strengths and weaknesses of digital trace data](#). Text as Data.

4. How is text mining used in STEM education research?

Applications in STEM Ed Research

Performance

Feedback

Engagement

Other Uses

TM has been largely used to **evaluate academic performance** in different contexts, especially to essays and online assignments.

- writing style
- use of argumentation
- plagiarism detection
- peer interaction

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). [Text mining in education](#). Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(6), e1332.

Learning Lab Demo