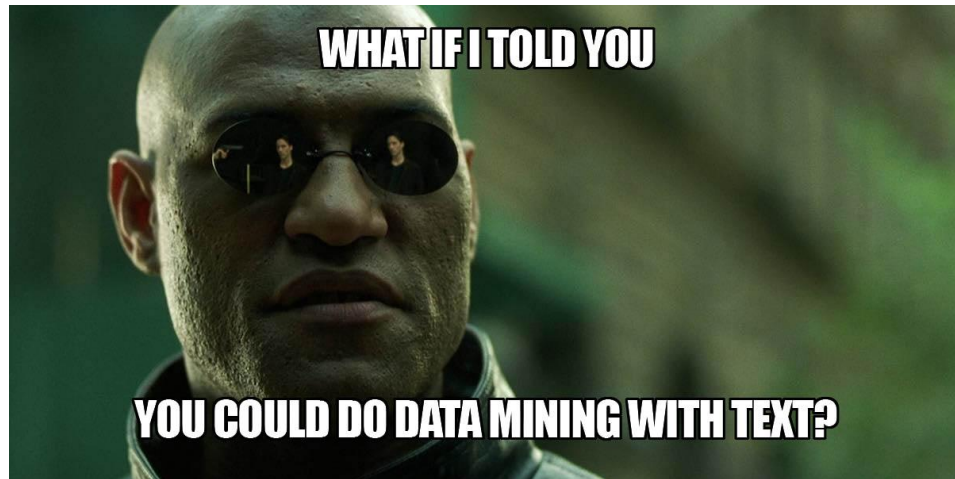


# TEXT MINING WORKSHOP



# Outline

- Overview of Text Analysis
- Examples in scholarship and teaching
- Designing instruction
- Hands-on demo and playtime
- Discussion

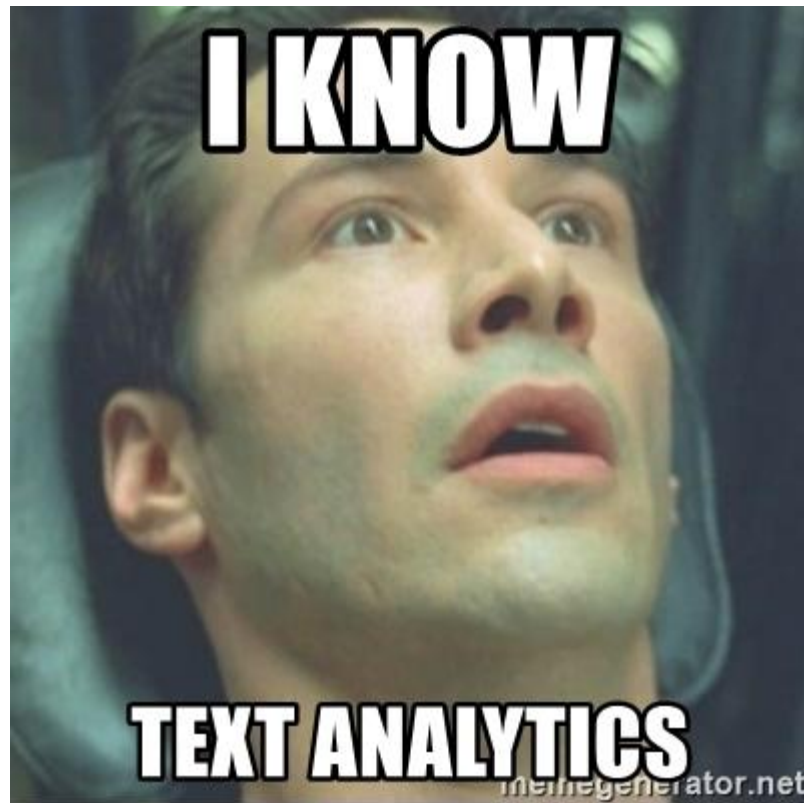


# Overview of Text Analysis

- Analog
- Distant Reading
  - Method for analyzing literature created by literary scholar Franco Moretti
  - Different from close reading, in which you carefully read, analyze, reflect on lit
  - Uses computer analysis to look at LOTS of texts and draw inferences
- Text Retrieval (libraries!)
- Term Frequency
- Topic Modeling
- Metadata analysis

## Uses

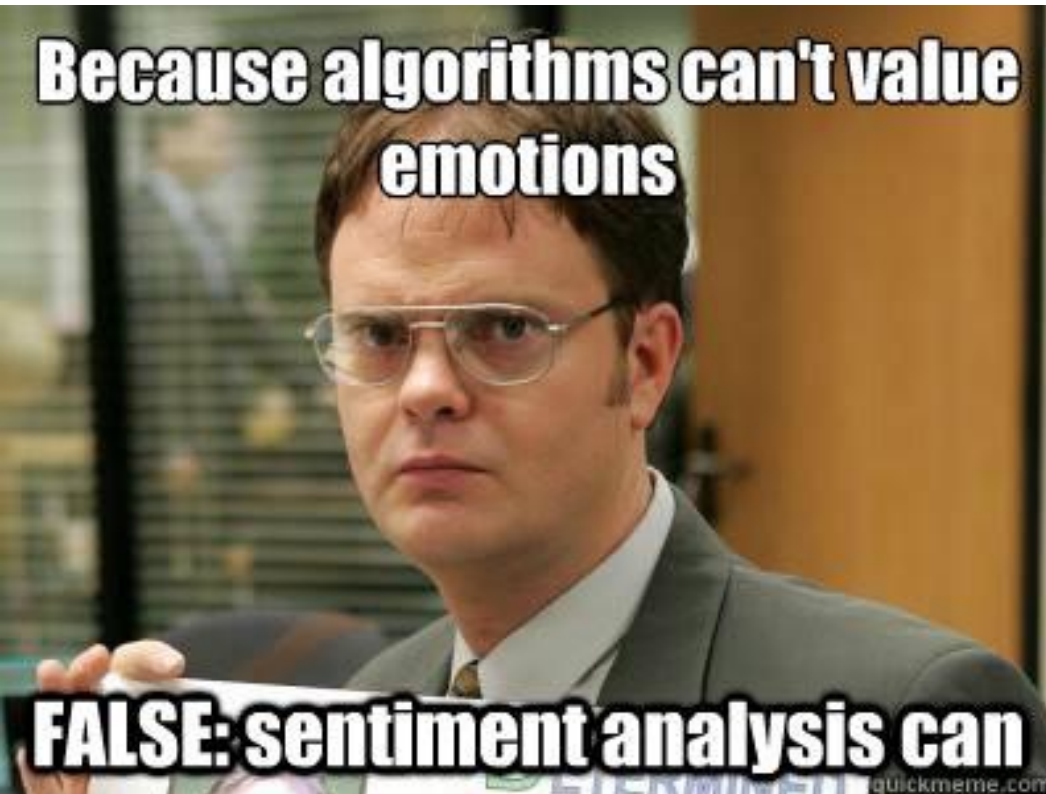
- Analyze changes in vocabulary over time
- Prove authorship
- Look for gender differences in vocabulary, sentence structure, novel structure
- Find discriminatory and biased language
- Compare different editions or translations of a work
- Evaluate your own work!
- Get ideas



# Designing Instruction for Accessibility

- Provide handouts prior to class
- Make sure handouts and files are accessible
- Caption/transcript multimedia
- Provide instructions for keyboard or speech use only
- Describe visual cues in detail
- **Provide students with info about browser extensions to aid in color perception**
- Allow students to complete assignments outside of class





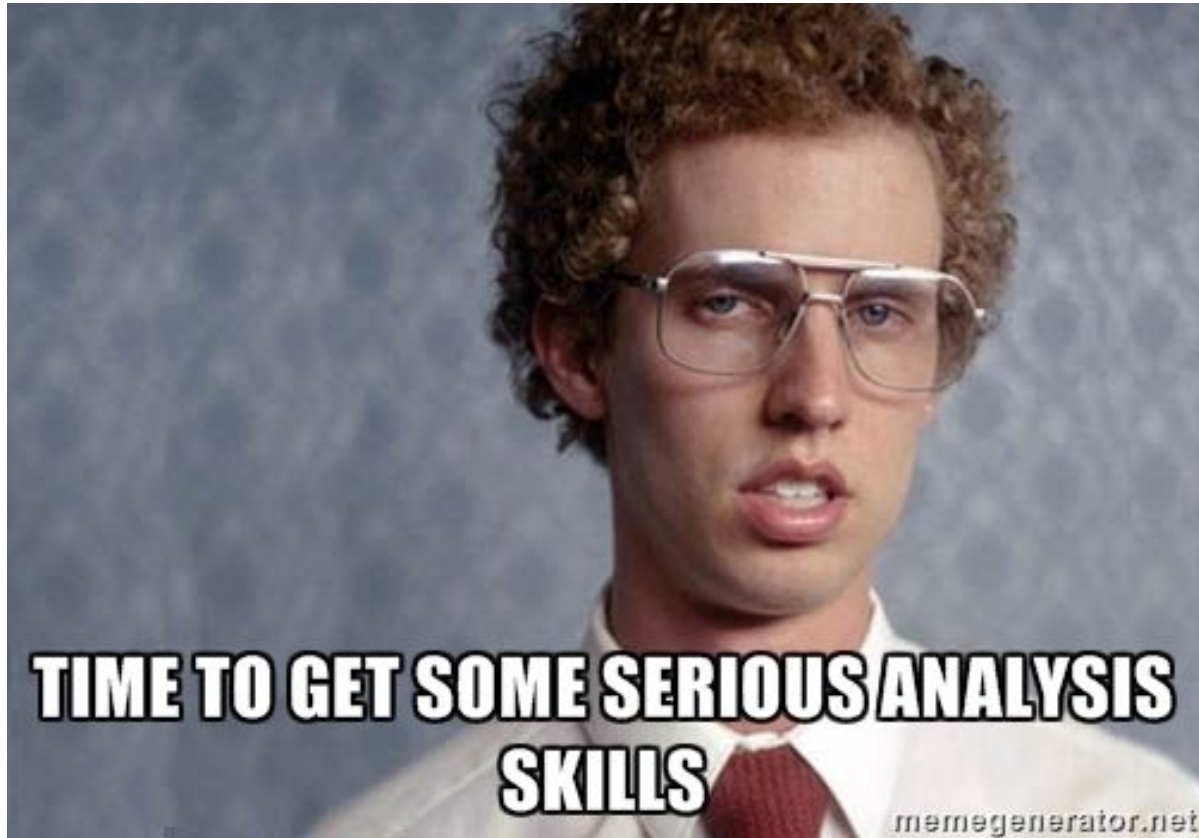
## General Text Mining Instruction Planning

- Finding text
- Digital Editions and Annotation
- Don't reinvent the wheel
- Sneak in standards
- Use fun data
- Clean...?

**QUESTIONS!?**

# Voyant

- Open source
- Web-based version (which we will use today) and also a downloadable VoyantServer
- Four ways to upload corpus
  - open an existing corpus (click on the "Open" button under the text box)
  - type or paste text into the main text box (this creates a corpus with one document)
  - type or paste one or more URLs into the main text box (one URL per line)
  - click the "Upload" button under the text box to use files from your computer
- File formats include plain text, PDF, XML, MS Word, RTF, etc.
- Works with almost any language but does not have language-specific functionality (needed for things like semantic analysis)



[researchguides.loyno.edu/text\\_workshop](https://researchguides.loyno.edu/text_workshop)



## Discussion



**MADE A TEXT DEPENDENT  
ANALYSIS**



**GOT A FOUR OUT OF  
FOUR**