Exploratory Corpus Analysis

School of Information Studies
Syracuse University

Corpus Construction

Collect text data from the Web

- Web crawling with Python

Clean text data

- Regular expression with Python

School of Information Studies
Syracuse University

Corpus Analysis

You can examine the frequency of each word
- Most/least common words/phrases

You can also build a lexicon for words of particular interest, for example:
- Sentiment
- Family relationship
- …

Trend analysis
- Word/phrase frequency change over time

Comparative analysis
- Compare word frequencies in different subsets

School of Information Studies
Syracuse University