Lexical similarity

- greedy string tiling
- longest common subseq.
- Jaccard coefficient
- word containment
- idf-weighted intersections
- cosine similarity
- Word [1–4]-grams

Syntactic similarity

- Syntactic trees (PTK)
- Chunks to POS to lemmas

Semantic similarity

- Latent semantic analysis
- Qatar Living data
- GloVe and Composes
- Common crawl