Idea

- Problem: You can post anything on the internet
- Verify content against an established information source
- Solve the problem of Internet reliability
Implementation

- Reddit API
- Freebase API
- Natural Language Toolkit (NLTK)
- Stanford CoreNLP Parser
Assumptions

● Sentences have a core topic
● Our corpus is reliable
● Similarity of a query & reliable document of the same topic = validity
Results

The Good
● Comparison of two sentences with a shared topic

The Bad
● The score isn’t normalized/objective
● Finding the topic is hard
● Negative clauses are “ignored/unpenalized”
Improvements

- corpus size
- HITS/PageRank for authority
- NLP usage
- topic generation
Future Work

- Create an autonomous reddit bot
- Expand corpus beyond Freebase
- More in-depth NLP
Questions or Suggestions?

- Implementation
- Algorithm
- Freebase/Reddit API
- App Engine
- NLTK