Possible Class Project Topics

- Focused crawlers
- Content mining
- Topical crawlers
- Social Network Analysis
  - On the Web – finding small communities (crime networks)
  - On the e-mail databases
- Co-citation measurement for network analysis
- Bibliographic coupling measurement for network analysis
- Text and Web page pre-processing methods
  - Stemming
  - Stopword removal
- Hit account accuracy
- Inverted index
  - Index construction, index compression
- Web spamming
  - Content spamming, link spamming, combating spamming
- Collaborative filtering
- E-mail spamming