

## NSF Grants

The abstracts for grants awarded by the NSF are available on the NSF web site. This can be scraped with standard tools like the mechanize and BeautifulSoup libraries for Python. Other collections of grants, pre-prints, and the like are also available from various sources. The idea of collecting this information is to do a text analysis. The project that motivates this was done on clustering of words by year, trying to see whether you could tell what was “hot” in a given year from the types of language used in abstracts. There is of course a huge literature on text analysis, and for graduate students especially analyzing grants and papers for interesting information can never be an untimely topic. This project forces learning enough of a non-statistical programming language to do the scraping and data collection and organization, and then allows a very creative range of possibilities in questions to ask and statistical techniques employed to answer them.