Daniel O’Brien

Multi-tiered text analysis

Text analysis is an increasingly researched field that has just a few barricades holding back its progress. One of these barricades is the intensive computational costs used by its processes. Enterprise-level systems are required to perform any sort of meaningful text-based analysis in a timely fashion. This means that useful text-analysis projects get pushed aside because they lack cost efficiency. The goal for this project is to explore potential efficiency gains that can be found by implementing a multi-tiered check-based analysis system.

Searching for efficiency gains in computer programs is not a matter of saving a few seconds for convenience sake. Companies have realized the economic value of data, and are scrambling to acquire, store, and analyze it to increase profits. Data is being stored and shared at a rate never seen before. In Data Science, more data always leads to a more accurate answer. As such, Companies who can afford it are purchasing incredibly large data sets and handing them off to data scientists to analyze. These Data Scientists are then left with the problem of efficiency. With such large data sources, inefficient functions can render a program useless.

Increasing efficiency of computer programs has always been a fundamental part of any Computer Science discipline. This focus on maximizing efficiency has driven the field to where it is today.