Thomas Garvis, Ngram Word Prediction Algorithm (Revised)
IT1. Artificial Intelligence; Machine Learning; Natural Language Processing
Key Words: n-gram, natural language processing, computational linguistics

Overview: The goal of this project is to create an algorithm that can shift through numerous ngrams to predict which word will most likely be typed next. The Ngram Word Prediction Algorithm will predict a user's next word in any number of given circumstances. Using almost all the books in the world as a collection of data, this algorithm will be able to predict the most common word that would appear next in real time. The user would put in a phrase one to five words long. This algorithm reads through the collection of words then provides the user with a couple of options the algorithm believes should be the next word. Ngrams make this possible. An ngram is a contiguous sequence of n items. In this case, words and sentences from books are used as ngrams. These ngrams can be used to create a new word prediction and query completer algorithm. Once this algorithm is developed, it can be put into many different applications to solve different problems.

Intellectual Merit: This Small Business Innovation Research Phase I Project creates a new way to word predict and opens the possibility to find better prediction algorithms than the ones we have now. This algorithm would help spark interest in the new ngram technology. Google released a large amount of data, through ngrams and google books, to the public for any use. Since all this data is still new, this is the first autocomplete project utilizing this valuable information to produce results. Some difficulties that could arise would be the speed of the algorithm. When shifting through the copious amounts of data sets, running a precise and speedy program will be ideal. After this algorithm is done, it will be implicated into different application to solve different problems.

Broader Impact: This algorithm could be applied to countless application to perform and solve different problems. An in-time word prediction algorithm would fit into a writing program to help authors and writers with their work. With the proper collection, this algorithm could rival search engine's word predictions. When added to a children learning application, this algorithm could enhance the teaching of children. Kids can type and see possible completion of sentence, thus learning sentence structure. They would be able to substitute words to see how it would affect a sentence. Ngrams are a powerful technology and this project will promote them and encourage other people create more products that use Ngrams. This algorithm is very flexible and would be able to support different collection such as number or characters, not only words.