For my project, I'd like to use SVM classification/prediction techniques in the field of text mining, specifically in the context of summarizing research papers. Given a research paper with an omitted conclusion, I would like to be able to automatically generate a conclusion consisting of several key sentences from the paper.

In order to be able to do this, I will be parsing sample research papers by words, and associating several features with these words, such as the frequency with which they appear, the section of the paper they appear in, and the positioning of the word (sentence/paragraph-wise). Commonly-used words will be disregarded. An SVM classifier will be used to create a scoring system to determine which words are most likely to be in the "class" of significant words that are found in the summary of the article.

In terms of generating the summaries, a subset of sentences from the original paper will be chosen, based on the average word score of the sentence.

I haven't figured out the exact details of how the features will be recorded and how best to calculate the sentence scores based on the word scores, but this is a rough overview of the process I wish to employ.

Patrick Paczkowski