Final Project Progress Report

1. COMPLETED: Abstracts have been scraped from the ACM Digital Library for two prominent data mining researchers: Jiawei Han (UIUC) and Philip S. Yu (UIC). To match the temporal time frame, abstracts were only obtained up until the end of 2005. It is likely that Han's abstracts were used in the actual paper, but I am not so sure about Yu.

Label	Author	School	Link
Author A	Jiawei Han	UIUC	ACM DL
Author B	Philip S. Yu	UIC	ACM DL

An initial PLSA model was repurposed using the model created in MP3. A "frequent pattern mining" topic was found withing Han's abstracts, including a good match of coverage, but not for Yu's. I will attempt to contact the author to determine see if I can find the name of the second author. Otherwise, I will proceed with the two authors named above. (The scraper and initial data can be found here.)

2. PENDING: I need to implement the Fixed Coverage CPLSA model described in the paper. At this point I feel I understand the mechanics behind it, it is just a matter of coding it. Following that, I will compile the documentation and compose the video tutorial.

3. CHALLENGES: The main challenge is duplicating the data set used by the authors. While this is not entirely necessary, I feel that it would provide a good sanity check for my work if the data sets match.