CSC 7300 Fall 2011
Project Description

Write a program to count number of occurrences of a pattern in the text using FM-index. Your program will first read the name of the input text file and will construct FM index using following steps:

1. Build Suffix Array for the input text (see the notes posted on moodle for linear time construction of Suffix arrays)
2. Get BWT(BurrowsWheeler transform) of input text from suffix array
3. Build Wavelet tree for BWT (Code for bit array and rank-select structure will be posted on moodle soon)

On successful construction of FM-index, program will read name of the file containing patterns to be searched and will output number of occurrences of each of the pattern to terminal or to output file. Program should also measure time required for searching the individual patterns.

Objective is to study,

- effect of pattern length on query time
- space required for the index as compared to the original text size