Project Goals

Members: Harsh Patel, Vivian Gray, Catherine Liu, Vincent Hammers

DataSet: http://www.imdb.com/interfaces (title.akas.tsv.gz)

The dataset we are going to use for this project is the IMDB movie dataset. The format of the document was in tsv.gz. Therefore, we had to decompress it and were able to turn it into a .log file that consists of raw data which includes:

- titleId (string) - a tconst, an alphanumeric unique identifier of the title
- ordering (integer) – a number to uniquely identify rows for a given titleId
- title (string) – the localized title
- region (string) - the region for this version of the title
- language (string) - the language of the title
- types (array) - Enumerated set of attributes for this alternative title. One or more of the following: "alternative", "dvd", "festival", "tv", "video", "working", "original", "imdbDisplay".
- attributes (array) - Additional terms to describe this alternative title, not enumerated
- isOriginalTitle (boolean) – 0: not original title; 1: original title

Algorithms

- Breadth-First Search Traversal
- Dijkstra's Algorithm (Shortest Path)
- Landmark Path algorithm. (Shortest path from a to b through c)

Goal

We wish to implement a self-sufficient tree that organizes itself depending on the fastest speed to a certain movie given a certain input from the user. In other words, if the user types in a certain movie name, the algorithm will organize the tree in the quickest way possible given the type of data to allow for the fastest search time.