AI Project Proposal

Stephen Olsen
saolsen@uwm.edu

March 27, 2011

Introduction

A Wikipedia race is a competition between two or more people where given a start page and a goal page you try to get to get to the end page as quickly as possible by clicking on internal wikipedia links. For an example and to try it out see http://www.wikipediarace.com/

This project is my attempt at an application that tries to find the shortest path between two wikipedia pages via links, as length of path is probably the greatest determinate of how quickly one can move from the start page to the goal.

AI Algorithm: A*

Input: A starting page and a goal page.

Output: A List of the links in order to get from the start to the end.

Programming Language: Python (and a little JavaScript)

Environment: It will run on both Windows and Unix

Finding the Shortest Path

The English Wikipedia site has over 3.5 million pages with many more interconnected links between them. This makes any simple search algorithms such as BFS and DFS unusable as they would take too long. My plan is to develop a heuristic to be used with A* search that is based on how closely related wikipedia article categories are, and then use this to guide the algorithm to finding a path to the goal node.

Bonus Points

For bonus points I'll be writing a web front end where you can put in the two pages (start and goal) and then it'll show you a cool graphical representation of the path between the two pages.