Wikipedia Maps

Rylan Mahany

WIKIPEDIA Maps

I first got the idea of this project when playing “The Wikipedia Game”. For those who haven’t played, You start on one Wikipedia page, and try to get to a target page using the related links provided on each page. Project file in GitHub.

Defining the problem

An example of a simple “route” on WikipediaMaps is Civil War to Abraham Lincoln. Since Abraham Lincoln was the leader of the Union, he is listed on the page. This gives the route a length of 1. For Basketball to Baseball, the route could be Basketball to Michael Jordan to Michael Jordan’s short-lived baseball career. This gives the route a length of 2.

Approach to the Problem

The solution I came to was to use a webcrawler and a depth-first-search. The webcrawler would first grab every page on the starting website.

Example: Basketball -> Michael Jordan, Kobe Bryant, Spalding

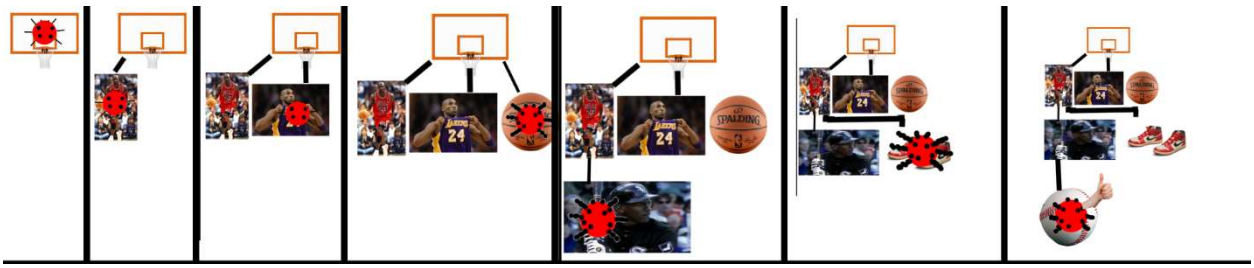
It would then move to Michael Jordan, the first link to look for baseball.

Michael Jordan-> Michael Jordan Baseball Career, Michael Jordan Shoes

After not finding baseball, the first link is taken again.

Michael Jordan Baseball Career -> Baseball

Now that the target is found, the path can be returned.



Here is a visual of the process of how WikipediaMaps would go from Basketball to Baseball, through MJ and his baseball career.

Problems and limitations

This algorithm will absolutely work for "easier" examples of the problem. Breadth-First-Search(BFS) is a useful search algorithm because it can find an item in $O(b^d)$ Time where b is the maximum links per page, and d is the depth of the target. If each page has 5 links and we go for 3 pages, the maximum number of operations to find the target is 5^3=125 operations. The issues lie within two

things, the first is the size of Wikipedia. b on each page can be in the hundreds, and d could be very far away. The other lies in that we need to recreate our path, this makes the Space complexity $O(b^m)$. Where m is the maximum depth, this is the page that is the furthest away from the starting. This could be The Stone Age and the Apple II Computer, literally the furthest away. To take this past a few page jumps, further I would need either a lot stronger computer or a smarter algorithm.