Abstract:
The CiteSeer database at http://citeseer.ist.psu.edu contains hundreds of thousands of scientific papers.
All of the papers are connected through citations:

- Red cites Green
- Green cites Blue and Yellow
- Blue cites Yellow

Searching on the database should take advantage of this. For example, finding a paper because it shares citations with other papers: If Red cites Green, they probably share topics, and we should look at Blue and Yellow because they likely do too.

CiteSeer does not allow searching through graph commonality.
My project: Write a web application with algorithms to act on CiteSeer graph, to improve searching:
- Full-text search of titles, authors, excerpts
- Shortest path from source to target
- Local network, by degree
- Find similar papers by common citation

Uses MySQL to store data in tables; uses PHP/HTML to access.

Examples:
Path-Finding

Shortest path from Red to Orange is that of Red lines. The path of Green lines is longer than that of Red Lines, so we examine all of Red's children first, with a Breadth-First Search algorithm.
If we use two-way searching (both inward and outward), then we can use Dark Blue line.

Similarity

Light Blue shares a parent (Yellow) and a child (Dark Blue) with Red. It therefore is similar to Red. Green shares a parent (Gray) with Red. It is also similar to Red, but less so than Light Blue.