MATHEMATICS FOR COMPUTING & ELEMENTS OF COMPUTING

Page rank algorithm and Google Search Engine

Group -   13

Team Members

Adhwaidh K	-	CB.SC.U4AIE24003
Adith S	-	CB.SC.U4AIE24004
Chaitanya Varma	-	CB.SC.U4AIE24017

Introduction

- The PageRank algorithm is a fundamental method for ranking web pages based on their importance in a hyperlink network.

- This project aims to implement and demonstrate the PageRank algorithm using various techniques such as:

    - Gaussian Elimination

    - Inverse Matrix Multiplication

    - Power Iteration Method

    - NetworkX Library

    - Web Crawling and Custom PageRank Calculation

- The project helps understand the significance of link structures in determining web page authority, as seen in search engines like Google.

Objectives

- To implement the PageRank algorithm using multiple approaches.

- To visualize the results through bar graphs and pie charts.

- To develop a web crawler that constructs a hyperlink network and computes PageRank values dynamically.

- To provide insights into how Google's search engine ranks web pages.

Workflow

Data Collection: Crawl web pages and gather links.

Graph Creation: Represent web pages as nodes and links as edges.

Matrix Formulation: Build a matrix with transition probabilities.

PageRank Calculation: Use methods like Gaussian Elimination, Matrix Inversion, or Power Iteration.

Visualization: Display results using bar charts, pie charts, and graphs.

Methodology

1. **Matrix-Based PageRank Calculation**

- The adjacency matrix representing the web structure is constructed.

- Damping factor (usually 0.85) is applied to account for random surfing behavior.

- Two methods are used for calculation:

  - Gaussian Elimination

  - Inverse Matrix Multiplication

- Results are visualized with bar graphs.

2. Power Iteration Method

- This iterative technique calculates PageRank by repeatedly multiplying the rank vector with a modified adjacency matrix.

- Results are visualized using a pie chart.

3. NetworkX Library for Complex Networks

- A directed graph is created using NetworkX.

- The nx.pagerank() function is employed to compute PageRank values efficiently.

- Results are visualized in both circular and pie chart formats.

4.   Web Crawler Implementation

- A web crawler is developed to visit web pages, extract hyperlinks, and build a network graph.
- The custom PageRank algorithm is applied to rank the visited pages.
- Results are visualized using a horizontal bar chart.

Code

Results

- Accurate PageRank values were obtained using Gaussian Elimination, Matrix Inversion, and Power Iteration methods.

- The NetworkX library effectively computed PageRanks for larger networks.

- The web crawler successfully built a hyperlink network and calculated realistic PageRank values.

- Visualizations through bar and pie charts clearly represented the rank distribution.

Thank You!