WEB PERSONALISATION BASED ON ANT COLONY SEARCH

Dev Panpaliya
Student, School of Computer Science and Engineering, Maharashtra Institute of Technology (MIT-WPU), Pune, 411038 , Maharashtra, India.
email: devsp2501@gmail.com

Abstract

This project introduces an innovative approach to word searching within a given text using the Ant Colony Search algorithm. Analogous to ants seeking food, the algorithm navigates through the text, leaving trails (pheromones) that influence subsequent choices. Users can customize parameters for improve results. The implementation has been enhanced to prioritize the exploration of words sharing the starting letter with the target word. The project provides visualizations to illustrate the search process and presents the best path found. Overall, this approach offers an accessible and visually engaging method for word retrieval in textual data.

I. Introduction

This project introduces a unique way to search for specific words in a text using the Ant Colony Search algorithm. Like how ants find food, the algorithm explores the text, leaving trails that guide its path. We can also change settings like evaporation rate to reduce and shorten the path for better results. The project also provides visualizations to showcase the path discovered as well as the word ranking based on the given path. In short, it offers an accessible and visually appealing method for locating words in text.

II. Motivation

The motivation behind this project is to improve the user experience in web searching by introducing a personalized and adaptive search algorithm. Traditional search engines often rely on algorithms that prioritize popularity and relevance, neglecting individual preferences and context. Ant Colony Optimization, known for its ability to find optimal paths in complex systems, offers a unique solution to the challenges associated with web search personalization.

III. Problem definition

The primary objective is to navigate through a body of text to locate a specific word efficiently. The initial step involves web scraping to extract textual information from websites. Subsequent preprocessing of the data to a normal format. Then we perform the implementation of the Ant Colony Search algorithm on this processed text. The algorithm, inspired by the collective foraging behavior of ants, aids in uncovering paths that lead to the target word. Then we proceed with the visualization of these search paths, providing a tangible representation of the algorithm's exploration. Additionally, we identify the most optimal path among the traversed routes. In essence, this problem seeks to address the need for an effective and systematic approach to navigate diverse textual datasets, enabling the efficient discovery of specific words through the innovative application of the Ant Colony Search algorithm.

IV. Objective

a. Personalized Search: Develop a system that learns and adapts to user preferences over time, providing search results tailored to individual needs.

b. Adaptive Optimization: Implement an Ant Colony Algorithm to dynamically optimize search paths based on the relevance of web pages and user interactions.

c. Efficient Crawling: Design an efficient web crawling mechanism to gather relevant data from diverse sources while respecting ethical considerations and privacy standards.

d. User Feedback Integration: Incorporate mechanisms for user feedback to continuously improve the accuracy and relevance of search results.

V. Technology

a. Programming Language: Python

b. Web Crawling: Scrapy, Beautiful Soup

c. Data Processing: Pandas

d. Network Analysis: NetworkX (for graph representation and manipulation)

e. Statistical Analysis: NumPy, SciPy

f. Modules Used: Beautiful Soup, NetworkX, WordCloud

VI. Dataset Description

Source: WEB PAGES

1. https://www.who.int/india/health-topics/diabetes
2. https://www.niddk.nih.gov/health

VII. Data preprocessing

The following are steps involved in the pre-processing:

1. Removing URL from the data
2. Converting the extracted text to lowercase
3. removing stoping words
4. Combining preprocessed text into a single string
5. Tokenizing the text

6. Removing punctuation

7. Filtering out empty strings

8. Join the words back in a single string

9. Updating the preprocessed text column

10. Printing the preprocessed text

VIII. Key components and algorithm

1. Ant Colony Search Algorithm: The algorithm mimicked the foraging behavior of ants, with each ant navigating through words in the text. Pheromone levels on words influenced subsequent choices, promoting the exploration of paths with higher pheromone concentrations.

2. Local Search Logic: The local search logic was incorporated to systematically traverse the path based on the starting letter of the target word. The algorithm followed a sequential search strategy, progressively narrowing down the possibilities by considering prefixes of the target word.

3. Visualization: The visualization component included generating a word cloud of the found path, emphasizing word frequencies. Systematic annotations, directional arrows, and sorted frequency annotations were added to enhance interpretability.

IX. Data Mining

In the context of your ant colony search and the found path data, you are not explicitly performing traditional data mining operations. Instead, you are implementing a heuristic search algorithm inspired by ant colony optimization to find a sequence of words in a given text that leads to a specific target word.

Here are some points related to data mining in a broader sense:

1. Exploratory Data Analysis (EDA): Before implementing the ant colony search, you may have conducted exploratory data analysis on your text data, which is a form of data mining. EDA involves summarizing the main characteristics of the data, visualizing patterns, and gaining insights to inform further analysis.

2. Pattern Recognition: The ant colony search itself involves recognizing patterns in the text data. The algorithm attempts to find a path of words that leads to the target word, using pheromone levels as a heuristic for the decision-making process.

3. Text Mining: The preprocessing steps, such as cleaning the text and extracting words, can be considered as text mining. Text mining involves deriving valuable information from textual data.

X. Output

After training the 3 models and using the models to predict following are the generated outputs

Raw Outputs –

Enter any word: diabetes

Found Path: ['noncommunicable', 'body', 'democratic', 'diabetes']

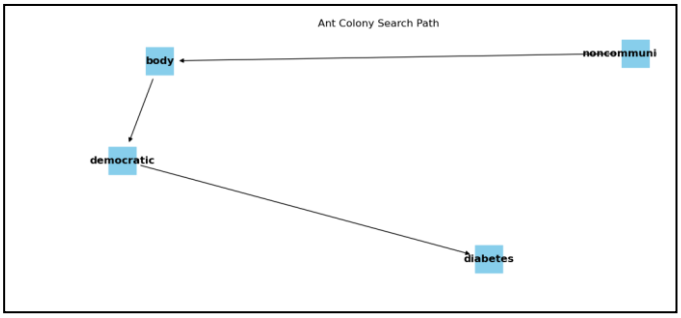

Fig. 1: Node visual representation of Ant Colony Search Path

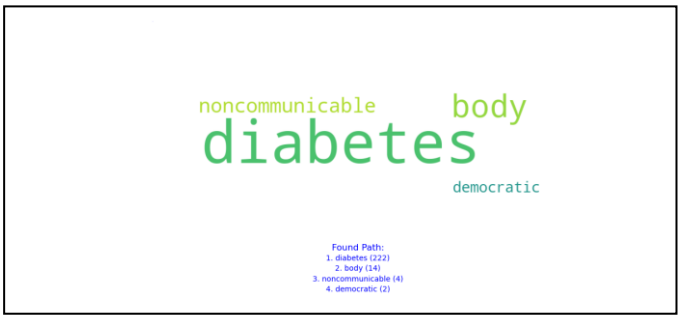

Fig. 2: Visual representation of most ranked words which appear in the path

Fig. 3: Visual representation of most ranked words in the dataset

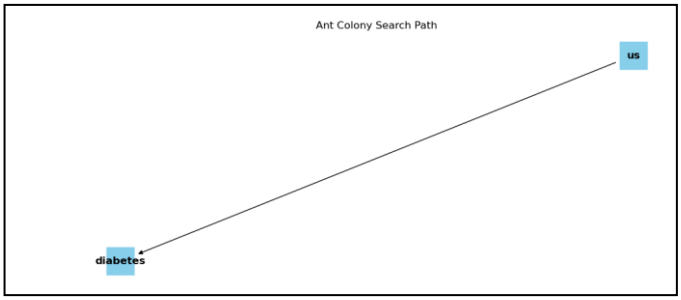

Fig. 4: Node visual representation of the best path

XI. Conclusion

The project aimed to implement an Ant Colony Search algorithm to navigate through preprocessed text and find a specified target word. The process involved simulating ant-like behavior to discover the optimal path by leaving pheromone trails on the words in the text.

XII. References

[1] Nandanwar, A. K., Choudhary, J., & Singh, D. P. (2021). Web search personalization based on the principle of the ant colony. *Procedia Computer Science*, *189*, 100-107.

[2] Konz, M. K. L. (1991). Comparisons between scholarly and collegiate living environments and academic success.