CS 201: Data Structures II
Enterprise Oracle

The Search Wizards

Spring 2023

1 Group Members

1. Ali Adnan (aa05432)
2. Syed Ibrahim Ali Haider (sh06565)
3. Muhammad Hashim Memon (mm06501)

2 Data Structure

An **Exponential Search** is a searching algorithm that is used to find the position of a searched value in a sorted array of data. This algorithm is a modification of binary search that first finds the range in which our target value may exist, and then uses binary search to narrow down the search to extract the exact location of our searched value.

The algorithm begins by comparing the target value with the first element of the array. If the target value is equal to the first element, the search terminates successfully. Otherwise, the algorithm sets the range of the search to be between the first and second elements of the array. It then doubles the range size by successively multiplying the size by 2 until the range size is greater than or equal to the length of the array or the target value is found.

Once the range has been determined, binary search is used to search within that range for the target value. Binary search is an efficient searching algorithm that works by repeatedly dividing the search range in half until the target value is found or the range is empty.

Exponential search is a useful algorithm for large sorted arrays when the location of the target value is not known. It has a time complexity of O(log n), where n is the size of the array, which is the same as binary search. However, exponential search has a better worst-case complexity than binary search, making it more suitable for certain scenarios.

3 Application

Exponential search is a data structure algorithm used to search for an element in a sorted array. It works by starting at the beginning of the array and doubling the index range in each iteration until the element being searched for is either found or the end of the array is reached. Once the range containing the element is found, a binary search is performed to determine its exact position. Exponential search can be used in various applications such as:

- Searching for a particular item in a large dataset, like in the grocery store app mentioned earlier.
- Searching for a specific word or phrase in a large text document or website.
- Searching for a particular record in a database of millions of records.
- Searching for a particular file in a computer's file system, especially when the file is located in a deep folder hierarchy.

4 Functionality

The project's functionality is to implement the exponential search algorithm to search through the "Sorted 7+ Million Company Dataset" based on user-defined search criteria. The program will prompt the user to select an attribute to search for from a list of available attributes, which include company name, domain, year founded, industry, size range, locality, country, LinkedIn URL, and employee num. The user can select an attribute by entering the corresponding number or name. Once the user selects the attribute, the program will prompt the user to enter the search value for that attribute. The exponential search algorithm will then be implemented to search for the user-defined search value within the chosen attribute, and the execution time of this search will be displayed in milliseconds.

The program's implementation of the exponential search algorithm ensures efficient and fast search results, even with a dataset as large as the "7+ Million Company Dataset". Upon completion of the search, the program will display the matching company's complete row of all attributes in a user-friendly interface, alongside the execution time in milliseconds.

In summary, the project leverages the exponential search algorithm and a large dataset to provide users with a powerful and efficient search tool for finding relevant companies based on their search criteria. Implementing the exponential search makes it an ideal tool for users who need to search through large datasets.

5 Datasets

The "7+ Million Company Dataset" by People Data Labs is a dataset available on Kaggle that contains information on over 7 million companies from around the world. The dataset includes information on the company name, domain, location, industry, employee size, and other company details.

Here are some key features of the dataset:

- CompanyName: the name of each company, which can be used to identify and group the data.
- Domain: the domain name for each company, which can be used to link the data to other sources of information.
- YearFounded: the year each company was founded, which can be used to analyze trends over time.
- Industry: the industry classification for each company, which can be used to group companies by sector or analyze trends within specific industries.
- SizeRange: the range of the number of employees for each company, which can be used to analyze the size and structure of companies.
- Locality: the city or locality for each company, which can be used to analyze regional trends or target specific areas.
- Country: the country for each company, which can be used to analyze trends and differences between countries.
- LinkedinUrl: the URL for each company's LinkedIn page, which can be used to link to additional information or sources.
- EmployeeNum: the number of employees for each company, which can be used to analyze the size and structure of companies.

This dataset can be used for a variety of purposes, such as market research, data analysis, and data visualization.

6 Work Distribution

Item	Activity	ID
1	Implementing the Frontend and GUI	aa0543
2	Implementing Exponential Search	sh06565
3	Icorporating dataset to the system's needs	mm06501

7 Attribution

1. ChatGPT, a large language model trained by OpenAI, based on the GPT-3.5 architecture, provided the response to "What are some practical applications of Exponential Search." (Accessed on April 3, 2023).

References

[1] OpenAI. *ChatGPT*. [Online]. Available: https://openai.com/chat . Last accessed: 26 Jan 2023. Prompt: how do i cite a chatgpt chat in bibtex format?

[2] PeopledatalabsSF. (2021). Free 7 Million Company Dataset. [Data set]. Kaggle. Available online: https://www.kaggle.com/datasets/peopledatalabssf/free-7-million-company-dataset