

PROGRESS REPORT:

An Experimental Analysis for Comparison of Searching Algorithms

Introduction

The following report responds to your request on the progress of our project for BLG 374E course. As you remember, our project is related to searching algorithms and main purpose of this study is to find best searching algorithm among Linear Search, Binary Search Tree, and Hash Table in terms of running times by implementing these algorithms and doing experiment on various sized string sample words on different operating system environments: Windows, Linux and OS-X.

As our method, we would like to implement three searching algorithms that we will compare by using C++ programming language, initially. Then, we will do studies to analyze the performance of algorithms asymptotically in order to foresee the result. After we would set the element numbers of the data groups to 1K, 10K, 100K, 1M respectively and compare the running times we have achieved, we would obtain the results by experimenting with an operating system to determine which algorithm is the best.

Until now, we did studies for asymptotical analyze of three searching algorithms. We implemented linear search algorithms. Binary search tree and hash algorithms are almost finished. Our project is a little bit late due to the team members' busy schedule. However, we consider that experiment process will be shorter than we supposed. Thus, the project is going to finish on time.

New Developments

Implementing of linear search algorithms and study of analysis of algorithms are done. Moreover, virtual machines for Linux Operating System and OS-X have been set on our computer, which is an environment for the experiment. Various sized sample data set is almost ready. However, implementing Binary Search Tree and Hash Table took longer time than we supposed due to the team members' busy schedule and death of one of the team member's grandfather. Nevertheless, there will be no latency for next process because the available environment is prepared well for experiments and expected result is determined. Experiment process is going to take shorter time than shown previous time schedule and project is going to be finished, on date, on time. There are no changes in our requirements.

Project Description

In our project, we will analyze the efficiency of the three algorithms in terms of their running time and compare them to each other. We will examine these three algorithms in terms of efficiency using different data sets and different operating systems. First, we will analyze the search time of the algorithms according to the number of inputs by changing the number of elements in the data sets. We will compare the run times of the data sets by setting the input numbers to 1K, 10K, 100K, 1M respectively. Later, we will determine the running times of these three algorithms on different operating systems. The purpose of conducting experiments on different operating systems is not testing operating systems

efficiency, yet showing the running times are not dependent on environments. We will graph this data and try to understand the results more clearly.

Progress

Completed

- Implementation of the linear search algorithm.
- The theoretical analysis of the running time from source books.
- Environments were provided for the experiment.

In Progress

- The implementations of the binary search tree and hash search algorithms.
- The preparation of the data set for experiments.

Planned

- The “average cases” of the algorithms will be taken as the result of the obtained researches.
- Running the algorithms with prepared data sets.
- Running algorithms on other operating systems.
- Showing results by using graphics.

Preliminary Results

We asymptotically analyze our algorithms, their time complexity is shown below. In particular hash function theoretically, takes $O(1)$ time. However, in practical, it depends on which collision resolution strategy have been chosen. While implementing we prefer to choose double hashing, which experimentally gave best results.

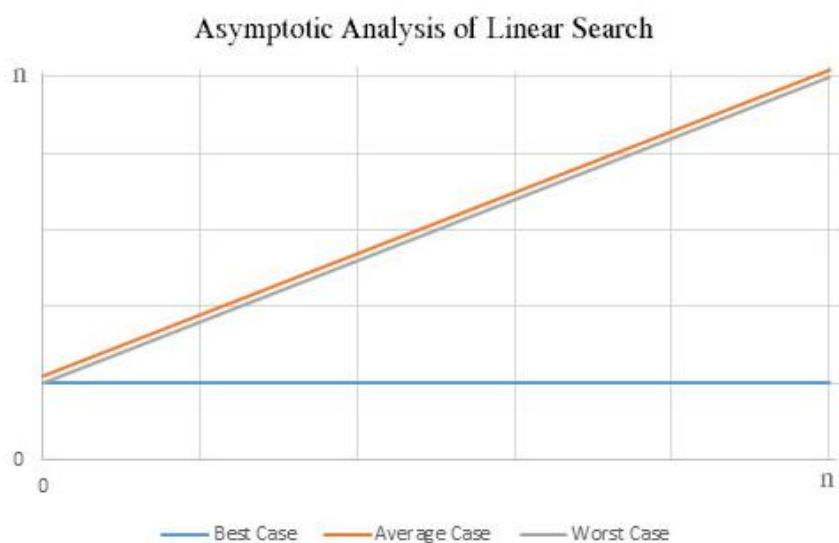


Figure 1: Asymptotic Analysis of Linear Search.

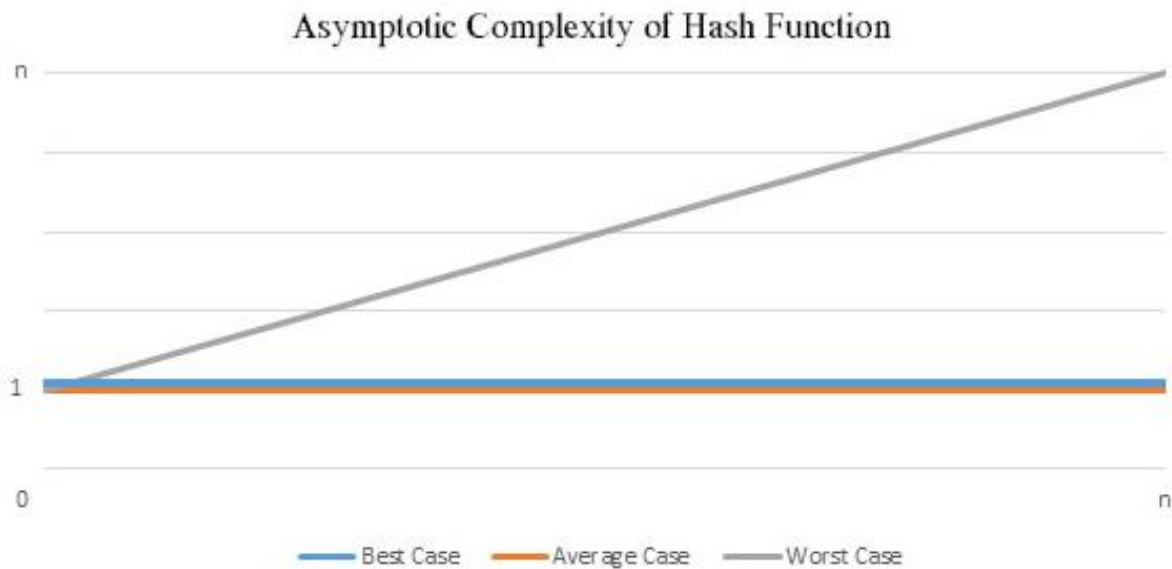


Figure 2: Asymptotic Analysis of Hash Function.

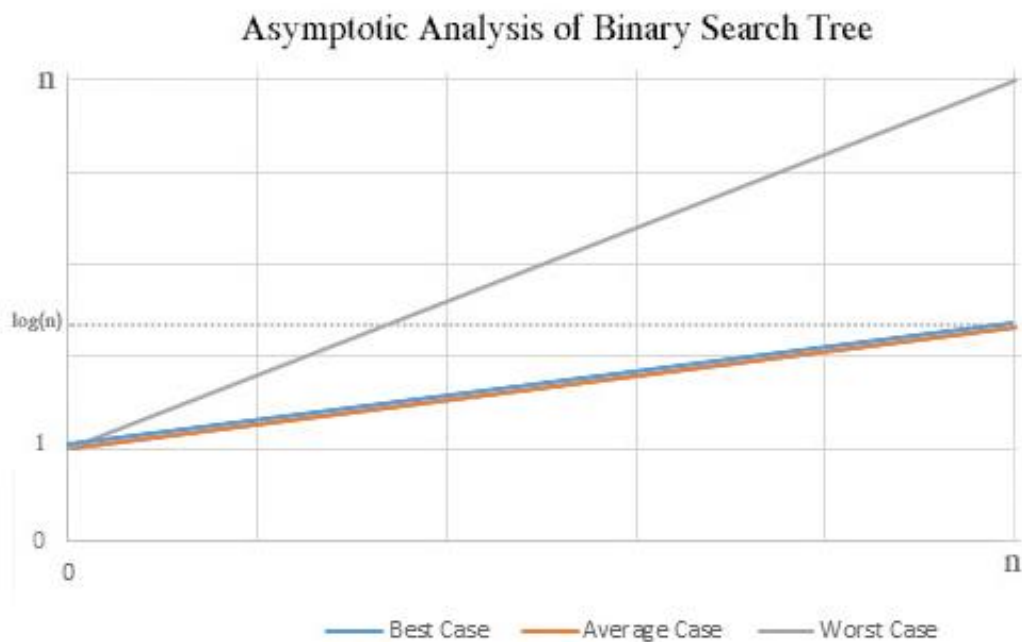


Figure 3: Asymptotic Analysis of Binary Search Tree.

Progress Assessment

As mentioned above we make minor modification on our time schedule. However, it does not affect our financial conditions. Modified time schedule provided below.

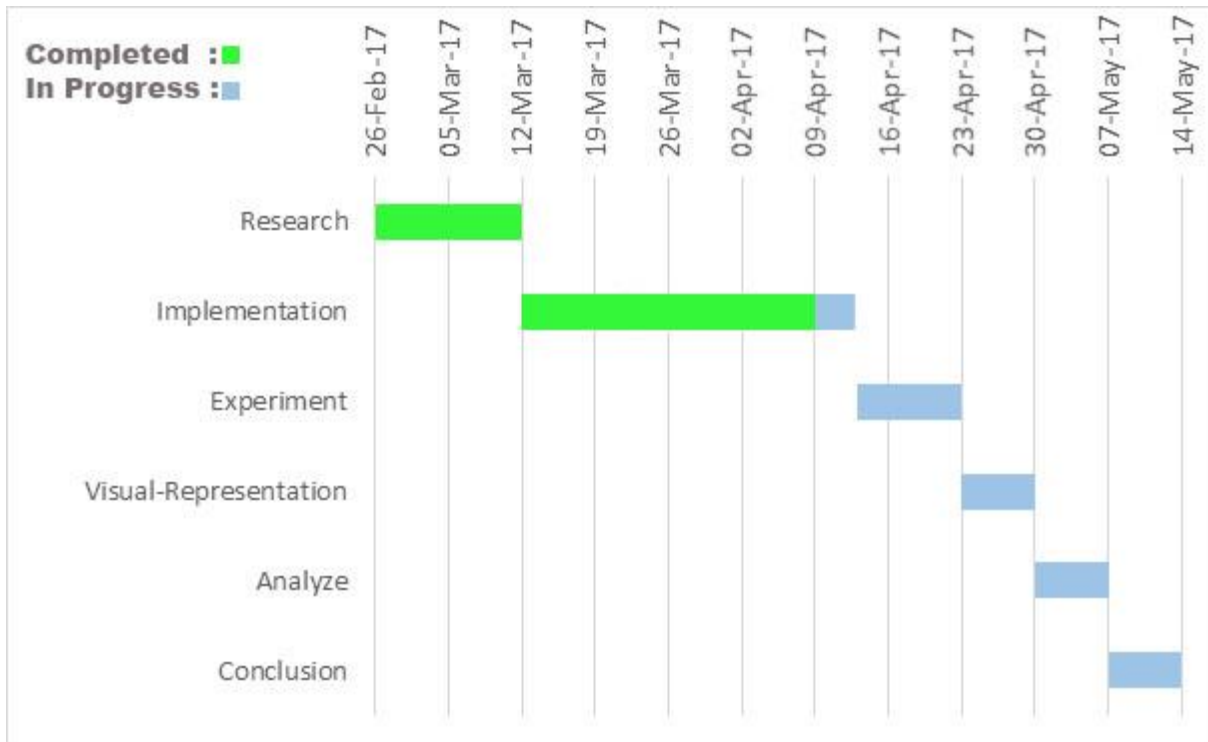


Figure 4: Modified Time Table.

Members of 3-Searchers:

- Okan YILDIRIM 150160537
- Fatih YILMAZ 150160531
- Özgür AKTAŞ 150160530