Hussein Search Algorithm: A Novel Efficient Searching Algorithm in Constant Time Complexity

Omer H Abu El Haijia¹, Arwa H. F. Zabian²
University of Castilla–La Mancha, Toledo-España¹
Jadara University, Jordan-Irbid²

Abstract—Hussein search algorithm focuses on the fundamental concept of searching in computer science and aims to enhance the retrieval of data from various data warehouses. The efficiency of cloud systems is substantially influenced by the manner in which data is saved and retrieved, given the vast quantity of data being generated and stored in the cloud. The act of searching entails the systematic endeavor of locating a particular item within a substantial volume of data, and searching algorithms offer methodical strategies for accomplishing this task. There exists a wide array of searching algorithms, each exhibiting variations in terms of the search procedure, time complexity, and space complexity. The choice of the suitable algorithm is contingent upon various aspects, including the magnitude of the dataset, the distribution of the data, and the desired temporal and spatial intricacy. This study presents a novel prediction-based searching algorithm named the Hussein search algorithm. The system is designed to operate in a straightforward manner and makes use of a simple data structure. This study relies on fundamental mathematical computations and incorporates the interpolation search algorithm, an algorithm that introduces a search by-prediction method for uniformly distributed lists, it forecasts the precise position of the queried object. The cost of prediction remains consistent and, in numerous instances, falls under the O(1) range. Hussein search algorithm exhibits enhanced efficiency in comparison to the binary search and ternary search algorithms, both of which are widely regarded as the best methods for searching sorted data.

Keywords—Binary search; prediction search procedure; prediction cost; constant time complexity

I. INTRODUCTION

On a daily basis, a substantial volume of data is generated across many formats, including photographs, videos, and text. This material is subsequently stored in cloud-based repositories, serving as a collective resource for retrieval from any given database. The increasing complexity of this matter can be attributed to the substantial volume of data that is generated and stored on a daily basis. The act of searching is of significant importance in various contexts, regardless of whether the item being sought is stored within a cloud-based infrastructure or a localized database. In both scenarios, the use of a search algorithm is essential for the successful retrieval of the desired item. The process of searching is commonly employed as a means of problem-solving, whereby the problem is provided as input and a solution is generated in the form of a sequential set of activities. Numerous instances in practical contexts can be classified as searching problems, such as the task of determining the shortest route between two nodes. These types of problems can be effectively addressed through the use of graph search algorithms. The act of searching can be categorized into two main types: sequential search and binary search. Various searching algorithms exist, each employing distinct strategies and exhibiting variations in terms of time and space complexity. Certain algorithms employ an informed approach, while others adopt a uniform approach, and a third category utilizes a partial information strategy for the purpose of item retrieval. The uniformity of the sequential search method arises from its lack of concern for any prior knowledge regarding the distribution of items. The binary search algorithm is considered an informed search algorithm due to its reliance on a sorted array. In recent years, there has been limited progress in enhancing the complexity of the search algorithm. This is primarily due to the satisfactory performance of the binary search algorithm in terms of searching complexity. However, it is important to note that the binary search algorithm does encounter challenges related to sorting, as it can only operate on sorted arrays. Additionally, it faces difficulties when dealing with comparison-based input involving searching for candidate items. Hence, the temporal complexity is intricately linked to the duration required for the sorting process, resulting in a trade-off where the time saved when searching is offset by the time invested in sorting. The interpolation search algorithm has a temporal complexity of O(1) when the items in the list are evenly distributed [1]. Therefore, in some scenarios, an interpolation search can provide an accurate estimation of the closest solution to the search problem. The ternary search method is a variant of the binary search algorithm that has a slower temporal complexity [1, 2]. The meta-binary search algorithm is a variant of the binary search method that iteratively creates the index of the desired value within the array. The approach operates in a way akin to the binary search algorithm, exhibiting a time complexity of O(log n) for locating the desired element. The ternary search technique partitions the array into three segments, utilizing the central point of each segment to locate the desired element. The logarithmic complexity of the search algorithm is determined by the number of steps required to locate the desired element, which is logarithmically proportional to the size of the array, denoted as n. However, the search range, which represents the number of elements that need to be examined during the search, is three times the size of the array (3n). Consequently, the worst-case running time of the algorithm can be expressed as the logarithm of three times the size of the array (log 3n), while the best-case running time may be approximated as being linearly proportional to the size of the array (O(n)). It is important to note that these
running time estimates are valid only when the array is sorted [3]. The jump search algorithm is a searching technique designed for sorted arrays. It involves performing a fixed number of jumps, denoted by k, on a block of data in order to locate the target element. Within each block, linear search operations are conducted to identify the desired element. The algorithm under consideration is superior to the sequential search method, but it falls short of the binary search technique. Specifically, it requires m-1 additional comparisons compared to sequential search, where m represents the size of the block to be traversed [4]. The interpolation search algorithm is a method that generates additional data points within a given range of known data points. Its time complexity is O(log log n) for datasets with uniform distributions, and O(n) in the worst-case scenario. The proposed approach represents an advancement over the binary search technique by employing comparison-based approaches that leverage a mathematical formula to approximate the location of the target element based on its value. Subsequently, the search is conducted in the vicinity of this estimated position. This alternative method has the potential to outperform the binary search algorithm under certain circumstances [5]. The exponential search algorithm operates on a sorted array. It begins by selecting a subarray of size 1, then doubles the size of the subarray in each iteration. The algorithm compares the final element of each subarray until the desired element is found. The algorithm in question is commonly referred to as exponential search, which exhibits a temporal complexity of O(log n) [6]. In the realm of searching for an item within an array of size n, two commonly employed strategies can be identified. The first strategy, known as sequential search, is applicable to unsorted arrays. The second technique, known as binary search, is exclusively applicable to arrays that have been sorted.

The study presented in this paper aims to introduce a novel searching algorithm that operates on a sorted array, relying solely on mathematical operations and a computed prediction approach implemented through a straightforward data structure. The search process in question exhibits a constant time complexity. While the space complexity may exceed that of sequential search, the time complexity can be lowered to O(1) in numerous scenarios and to O(constant) in the worst-case scenario. The algorithm relies on the computation of the array’s average, operating under the assumption of a uniform distribution of items within the list. Additionally, it generates supplementary arrays, one of which records the frequency of successful matches, while the other stores the locations where the sought-after items can be located. Part of process is similar to the Knuth-Pratt-Morris algorithm, which is commonly used for pattern matching [7]. The approach under consideration aims to decrease the time complexity by minimizing the number of comparisons and implementing a straightforward prediction system that ensures the absence of collisions through the utilization of error-free lookup tables and basic arithmetic operations.

The primary objectives underlying this research endeavor are to provide a straightforward predictive approach utilizing search techniques and to execute a basic computation with constant time complexity.

The subsequent sections of this study are structured as follows: Section II presents a review of relevant literature pertaining to searching algorithms. It is worth noting that the process of locating sufficient recent works on searching algorithms proved to be challenging. The majority of the literature discovered consisted of dated publications or encompassed broader discussions on binary search algorithms. In Section 3, the proposed algorithm is introduced. In Section IV are shown the results obtained from the proposed study, followed by an examination of the algorithm employed. Finally, the paper concludes with findings and discusses potential avenues for further research in Section V.

II. RELATED WORKS

One of the main benefits of searching operations in computer science is their capability to assist in finding particular data from databases. The effectiveness of the search process directly impacts the overall performance of the system. Searching algorithms are widely employed in several computer applications, such as problem-solving [9], data analysis, and information retrieval, due to their ability to efficiently search through extensive datasets within a limited timeframe. Various searching techniques exist, including sequential, binary, hashing, and graph search. The selection of an appropriate algorithm is contingent upon the particular situation at hand and the attributes of the data being queried [10]. The authors of [5] introduce a hybrid search method known as interpolated binary search (IBS), which integrates the interpolation algorithm and the binary search algorithm to accurately determine the precise position of the desired object. The IBS algorithm commences by employing an interpolation algorithm to estimate the approximate location of the item being searched. It subsequently operates as a binary search algorithm to precisely determine the location of the sought-after item. The Inverse Binary Search (IBS) algorithm exhibits a greater computational cost compared to both the binary search algorithm and the interpolation technique. However, when executed on uniformly distributed data, IBS demonstrates a lower temporal complexity cost than the aforementioned algorithms. Specifically, its time complexity is O(log2log2n). On the other hand, when applied to non-uniformly distributed datasets, IBS necessitates O(log2n) operations. In [8], a comparison between different search algorithms is presented, the authors analyze the performance of various search algorithms, including uninformed search algorithms (DFS, uniform cost search) and informed search algorithms (A* and BFS), with a focus on their time complexity and space complexity.

Graph search algorithms typically construct a graph based on the given input data and traverse the nodes of the graph using various strategies in order to locate the desired objects [13]. The algorithms that were examined all exhibited a time complexity of O(m), where m represents the number of offspring for each node (also known as the branching factor) and b represents the solution depth, which is the length of the path. The binary search algorithm utilizes a value of m equal to 2 and assigns b as the logarithm base 2 of n.

The act of searching is a crucial and widely employed process in several contexts. In order to obtain data of various
In the event of the most unfavorable scenario, the computational complexity for locating two items in the BSS (Binary Search Structure) is $N$, as opposed to $2N$ in sequential search or $(n^2 + \log n)$ in the binary search algorithm. The Bubble Sort algorithm demonstrates superior performance in terms of comparison count in both the best and worst case scenarios, when compared to the Binary Search algorithm and the Sequential Search method. A limitation of this algorithm is its inability to perform a search for a single item, as it is designed to search for about two objects simultaneously.

Hashing is a technique employed in the search of extensive databases, wherein a distinct key is generated for each record or item. This key serves to designate the specific area within the database where the item is potentially kept. The time complexity for retrieving an item from a big database using hashing can range from constant time ($O(1)$) to linear time ($O(n)$) in the worst-case scenario. The variable $n$ represents the size of the linked list, which is utilized for the purpose of storing colliding data within a single slot, denoted as $[12]$. The objective is to verify the presence of item $X$ inside a given dataset. As the volume of data expands, the intricacy of the situation escalates. The problem of searching has been introduced and elucidated by Levin and Solomonoff throughout the time span of 1973 to 1984 $[14,15,16, and 17]$. In reference $[18]$, a novel quantum technique is presented for the search problem, demonstrating a polynomial time complexity. This algorithm utilizes XOR logical gates to convert the data into a polynomial form, following the amplification process provided in reference $[19]$. Consequently, the search operation may be executed within a polynomial time frame relative to the input size, denoted as $n$.

In the cited work $[20, 21]$, the authors introduce a fractional cascade algorithm, which is a method aimed at enhancing the efficiency of binary search algorithms. This methodology achieves a reduction in the time complexity of binary search algorithms to $O(k + \log n)$ while searching for $k$ elements within a sorted array of size $n$.

### III. Hussein Search Algorithm

The Hussein search algorithm is designed to efficiently locate an element in a sorted array. It achieves this by utilizing a prediction table structure and a sequential search algorithm. This approach reduces the time complexity from $O(n\log n)$ in sequential search or $O(\log n)$ in binary search to $O(1)$ $[7]$. The algorithm achieves this improvement by employing only arithmetic operations and a technique inspired by the interpolation method for searching sorted lists $[1]$. It is important to note that the algorithm assumes a uniform distribution of elements in the list. The algorithm operates in two distinct phases: the preprocessing phase and the searching phase. During the preparation step, the entirety of the array undergoes arithmetic operations in order to ready the data for the subsequent prediction finding procedure. During the preprocessing phase, an array of integers denoted as $A$, which has a size of $n$ is examined. The data within this array is created in a random manner. During this phase, the operations conducted involve the calculation of the average value of variable $A$. The average can be determined by calculating the mean of a set of values. To obtain the mean, each element in array $A$ should be divided by a given value $t$, and the resulting values should be stored in a new array $B$. The size of array $B$ is denoted as $n$. Generate a novel array $C$, whose size is determined by the floor function applied to the value of $t$. To iterate through a set of counters starting from 0 to size $(c)$, the ceiling of each element $B[i]$ is compared with the indices of $C$. The number of matching values is determined, and the resulting matching score is stored in a new list $C1$ at the corresponding index of the matching item. The elements contained within the array $C1$ will represent the respective indices of the desired item within the original array. During the searching phase, the following operations are performed to determine whether an item $X$ is present in an array: $X$ is divided by $t$, and the resulting value is compared with the indices of $C1$ using the ceiling function. The value found in the corresponding cell of $C1$ represents the index of the searched item $X$ in the original array. If the corresponding cell is empty, it indicates that the item $X$ is not present in the array.

The following example provides a comprehensive elucidation of the functioning of the Hussein search method. In this scenario, array $A$ of size 16 is randomly produced by $[7]$, and the data within the array is uniformly dispersed. The elements in set $A$ are arranged in ascending order, while the outcomes of dividing each element in set $A$ by the average of the average are recorded in set $B$ (refer to Fig. 1(a)). A new array, denoted as $C$, is to be created with a size of 30. This array will contain the floor value of the average values obtained from Fig. 1(b). The quantity of corresponding elements in the array is stored in Fig. 1(c). Now, let us explore the scenario where aim to search for item $X$, which has a value of 310, within the table. Initially, the value of $X$ is divided by 30.89. The resulting quotient is then rounded up to the nearest whole number, denoted as 11. Subsequently, proceed to locate the element within the array by referring to the index position...
11. Upon examination, it is determined that the value at this index is 0, indicating that the desired element is not present inside the array. Let X be equal to 275. Next, perform a division operation on X by t, and subsequently apply the ceiling function. The resulting value, 9, is assigned to the index 9 in array C. Upon further examination, we finally, observe that the element at index 9 in array C is 1, indicating that X is located at index 5 in the original array (as depicted in Fig. 2).

![Figure 1: the preprocessing phase](image1)

![Figure 1b: index matching](image2)

![Figure 1c: the matching score](image3)

**Fig. 1.** Hussein search algorithm simulation on random data generation.

**Fig. 2.** Testing the search operation using Hussein search algorithm.

A. Algorithm Description:

1) Preprocessing phase: Consider we have an array of integers A of size n, where the data is randomly generated, in this phase the following operations are performed:

- Calculate the average of A. average = \( \frac{\text{sum}(A)}{n} \)
- Calculate the average of average: \( t = \frac{\text{average}}{n} \)
- Divide all the elements of A by t and store it in a new array B, the size of B is n
- Create a new prediction array C with size floor (t)
- For I counter that start from 0 to size (c), compare the ceiling of B[i] with the indices of C, and count the matching values and store the matching score in a new list C1 in the corresponding index to the matching item. The values stored in the array C1 will be the corresponding index of the searched item in the original array.

2) Searching phase: For searching an item X, if in the array or not is performed the following operations:

- Divide X by t, \( S = \frac{X}{t} \)
- Compare the ceiling of (S) with the indices of C1; the value found in the corresponding cell will be the index of the searched item (X) in the original array, if the corresponding cell is empty that means the item is not found in the array.

IV. RESULTS AND ANALYSIS

The proposed algorithm is simple and easy to understand and implement. It is implemented and tested in MacBook air i5 processor 1.3 GHz speed, 8 GB ram, using visual studio, C# and in Python, and is tested for large input size list up to 16 MB. Furthermore, we have successfully implemented both the binary search method and the ternary algorithm using the C# programming language. These implementations were carried out in an identical setting, with the input size being consistent with that of the Hussein search algorithm. The objective of the experiment was to conduct N iterations in order to seek a randomly produced list with a size of N. The findings indicate that the speed of the prediction searching method in all evaluated algorithms exhibits a linear relationship with the amount of input. However, it is noteworthy that as the input size increases, the performance of the Hussein search surpasses that of the other algorithms. The Hussein search method exhibits a constant time complexity of O(1) for finding an individual item.

Consequently, the search operation for n items may be accomplished in linear time complexity of O(n). This stands in contrast to the binary search strategy, which necessitates a time complexity of O(n log n) for searching n items. Fig. 3 presents the outcomes achieved by the Hussein search algorithm in contrast to the other algorithms when searching for N items across varying input sizes. While considering input sizes of 8 MB and 16 MB, it has been observed that the Hussein search method exhibits a time requirement that is 20% lower than that of the binary search strategy, and 17.3% lower than that of the ternary algorithm, while searching for all items. Hussein's search method demonstrates a search speed that is approximately 494% greater than that of binary search when applied to a dataset of 16 MB. Table 1 illustrates the speedup, which quantifies the extent to which the Hussein search algorithm outperforms the binary search algorithm across various input sizes. Table 1 illustrates the observed increase in search speed across various input sizes.

In the Hussein search algorithm the searching process about an item requires O (1), which means searching n items requires only O(n) in comparison with the binary search algorithm that requires O(n log n). Fig. 3, show the results obtained by the Hussein search algorithm in comparison to the other algorithms for different input size. For input size (8 MB,
16 M) searching all the items using the Hussein search algorithm requires time that is 20% smaller than the binary search algorithm and 17.3% smaller than the Ternary algorithm. The search speed in the Hussein search algorithm is increased by about 494% than the binary search for 16 M of data. Table I, and Fig. 3 show the speed up that represents how much Hussein's search is faster than the binary search algorithm in searching about all the input sizes.

### Table I. The Speed Up in Searching Different Input Size

<table>
<thead>
<tr>
<th>Input Size</th>
<th>Binary</th>
<th>Ternary</th>
<th>Hussein</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 k</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>50%</td>
</tr>
<tr>
<td>8k</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>250%</td>
</tr>
<tr>
<td>16k</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>200%</td>
</tr>
<tr>
<td>32k</td>
<td>12</td>
<td>13</td>
<td>3</td>
<td>400%</td>
</tr>
<tr>
<td>64k</td>
<td>27</td>
<td>24</td>
<td>5</td>
<td>540%</td>
</tr>
<tr>
<td>128k</td>
<td>43</td>
<td>22</td>
<td>10</td>
<td>430%</td>
</tr>
<tr>
<td>256k</td>
<td>89</td>
<td>109</td>
<td>20</td>
<td>445%</td>
</tr>
<tr>
<td>512k</td>
<td>189</td>
<td>222</td>
<td>39</td>
<td>485%</td>
</tr>
<tr>
<td>1 MB</td>
<td>349</td>
<td>419</td>
<td>69</td>
<td>506%</td>
</tr>
<tr>
<td>2MB</td>
<td>703</td>
<td>825</td>
<td>140</td>
<td>502%</td>
</tr>
<tr>
<td>4MB</td>
<td>1398</td>
<td>1606</td>
<td>259</td>
<td>540%</td>
</tr>
<tr>
<td>8MB</td>
<td>3343</td>
<td>3568</td>
<td>629</td>
<td>531%</td>
</tr>
<tr>
<td>16MB</td>
<td>5813</td>
<td>6784</td>
<td>1176</td>
<td>494%</td>
</tr>
</tbody>
</table>

Implementation of Hussein Search Algorithm: as mentioned previously Hussein Search algorithm is implemented in C#, with binary search algorithm and ternary algorithm on the same environment and the same data. Fig. 4 and 5 represent the implementation of the main functions of the Hussein search algorithm.

![Fig. 3. Hussein search algorithm speed up.](image-url)

![Fig. 4. The implementation of mathematical operations in Hussein search algorithm.](image-url)

Fig. 3. Hussein search algorithm speed up.

### Fig. 4.

```csharp
static float HusseinSearch(int x, float avg, int[] arr, int[] arrb, uint[] arrc)
{
    uint i = (uint)x / avg;
    if ((arrb.Length) > i)
    {
        uint m = arrc[i];
        for(uint i = m; i < arrc[i+1]; i++)
        {
            if (arrc[i] == x)
                return i;
        }
    }
    return -1;
}
```

V. CONCLUSION AND FUTURE WORKS

In this study, we introduce the Hussein search algorithm, a novel informed search approach that leverages a straightforward prediction method, basic arithmetic operations, and a simple data structure. The findings demonstrate that the Hussein search algorithm outperforms previous search algorithms in terms of time complexity, particularly when dealing with substantial data sets. Moving forward, there are several avenues for future research. Firstly, it would be valuable to explore the algorithm’s performance under different search scenarios and input distributions. Additionally, investigating potential optimizations and further enhancements to the algorithm could yield even more efficient search capabilities. Finally, conducting comparative studies with other state-of-the-art search algorithms would provide a comprehensive evaluation of the Hussein search algorithm’s effectiveness. The procedure operates on an array that has been sorted, with the underlying assumption that the data is spread equally. The Hussein search algorithm has a time complexity of $O(n)$ for finding $n$ items. In contrast, the binary search algorithm has a time complexity of $O(n \log n)$, while the sequential search technique requires $O(nn)$ in the worst case. Given the assumption of a sorted array, the proposed technique offers a notable advantage in terms of computational simplicity and a reduction in the number of comparisons required for item search. Specifically, the algorithm achieves a time complexity of $O(1)$ when searching for an item by index rather than by value. Fig. 6 presents a comparison of the running times for various input sizes. The future objective is to enhance the algorithm based on simple prediction methods to operate with the same time complexity for an unsorted array. Additionally, we aim to offer a sorting algorithm that utilizes the same mathematical processes and achieves linear time complexity in the worst-case scenario.
Fig. 6. Running time comparison for different input size.

REFERENCES


