Assignment 1: Abstract:

- The problem of locating all occurrences of a character pattern in a text is known as string matching. This paper presents an overview of various string matching algorithms, as well as a comparison of these algorithms and assessing complexity and performance

- Examples: such as Naïve search algorithm, Skip algorithm, Boyer–Moore–Horspool (BMH), Boyer–Moore–Smith (BMS), Knuth Morris-Pratt algorithm.

- Pattern matching is the type of research we're interested in, We'd like to look for a particular pattern, which may be something from a specific disease or mutation.

- also In this paper, we present a new idea for a single pattern matching in strings. The idea is named skip search (SS). Even in the worst-case scenario, the pattern is checked for presence in the provided text by accessing only half of the memory locations.

- the skip search is a modified version of the Naive search algorithm, By accessing only half of the characters in the given string and checking whether those characters are the beginning or ending character of the pattern, By doing this, we could obtain the skip length of the pattern size and hence reduce the search time

- The algorithm has a low space complexity since it does not require any preprocessing or additional memory space, When the skip algorithm was compared to the Nave algorithm and the Knuth–Morris–Pratt (KMP) algorithm, the(SS) performed better in most of the test cases

- When the pattern appears at the end of the text or when the pattern does not appear in the text, the algorithm takes significantly longer to execute.

Introduction:

- the pattern searching algorithm is a technique to find one or all occurrences of a pattern in the given text. For example, consider the pattern P and the length of it should be equal or shorter than text T and we need to find one or all occurrences of P in T to obtain useful information.

- We suggested the skip algorithm, which is based on a Naïve search algorithmwith modifications in the skip positions, and The algorithm returns the location of the pattern's first appearance in the text

- **In this paper, we compare the results of the skip algorithm with the Naïve algorithm and Knuth–Morris–Pratt algorithm**

- **The majority of the literature review is devoted to minimising the amount of comparisons and processing time.**

Assignment 2: Related Works:

- **The naive search, also known as the brute-force algorithm, compares the pattern to any text possibility and try all possible alignment.by moving the pattern p over the text T one by one in one direction from one end to other.it has no preprocessing step and The time complexity of it is O(mn) where m is the text size and n is the pattern size.**

- **The Knuth–Morris–Pratt (KMP) algorithm works by improving the length of shift.This shortens the process by depending on previous information obtained from comparisons.This requires preprocessing of the pattern** and **constructing an auxiliary lps[] of size n (same size of pattern) . This works well when there is similarity within the pattern.The time complexity of it is O(m+n) .**
- **Boyer–Moore published one of the most efficient strings matching algorithms in 1997 that performs pattern comparisons from right to left .The algorithm had a preprocessing step that, in the worst case required more time. The preprocessing here has the time complexity of O(m + $|\Sigma|$) and the worst case time is O(nm + $|\Sigma|$) .**
- **The Boyer–Moore–Horspool (BMH) algorithm uses the occurrence heuristic to increase the shift length in the text. The preprocessing here has the time complexity of O$(m + |\Sigma|)$ and the worst case time is O(nm) .**

The quick search (QS) algorithm is another common search algorithm.The shifting requirements are based on the bad character shift rule, and the searching is done from left to right. it has the worst Case complexity is O(nm) .The algorithm by Boyer–Moore–Smith (BMS) is derived from BMH as the reference and the property of bad character shifting. it has the same preprocessing complexity and search time of (BMH) .

- Karp et al. proposed an algorithm for finding approximate matches to a pattern in a string with the aid of mathematical calculations. Ziad et al. suggested a skip algorithm for multiple pattern matching, which minimised the number of pattern comparisons. Walter proposed an algorithm for matching different patterns. This is a modified Boyer–Moore algorithm that showed a significant increase in searching speed .

Rami et al. proposed three algorithms that shared the principle of occurrence search. They used pattern matching to preprocess the given pattern and construct an

occurrence list. Hussain et al. suggested a bidirectional search algorithm.The idea of enhancing the shift decision was used by the writers. The rightmost and mismatched character of the partial text window is compared to the left of the pattern at the same shift length for matching. The algorithm's complexity is O(mn)/2. However, it necessitates a longer preprocessing step

The faster method was suggested by preprocessing and modifying the Naïve search algorithm. Mohammed et al suggested the occurrence algorithm. The algorithm is an improvement over brute force, and this includes preprocessing of the pattern as well as the text.A comparison is done based on the preprocessed input in the algorithm by Reverse Colussi (RC). The preprocessing step has a complexity of O(m) while the searching phase is O(n).

Assignment 3: method:

- **The goal is to make searching more efficient. The suggested skip search, like most commonly used searching algorithms like KMP, BMH, BMS, and others, it searches the pattern from left to right .The index is incremented by skipping undesired elements, which makes this technique unique .**
- **It is accomplished by dividing the text into three groups :-**

1. **The part of the string which contains elements that can be only the starting point of the pattern .**
2. **The part of the string which contains elements that can be the starting as well as the ending point of the pattern .**
3. **The part of the string that can be only the ending point .**

Result:

- **The number of memory locations accessed determines how long the algorithm takes to execute.the KMP algorithm performs faster than this proposed algorithm. If the text that has the same subpatterns appear more than once,the Skip algorithm works better where the input characters are randomly placed.**
- **The result is compared with search algorithms namely:-**

1. **Naive search algorithm**
2. **Knuth–Morris–Pratt (KMP) algorithm.**

- **Naïve search algorithm is chosen for comparison because it does not require any extra space as well as it has no preprocessing phase.**

- **Knuth–Morris–Pratt algorithm is chosen for comparison because it has a minimal preprocessing phase complexity compared to other algorithms, i.e., O(m) and the searching phase is O(n).**
- **The skip algorithm performed better than the Naïve search as well as the KMP algorithm. The time taken for execution decreased gradually as the search position was longer. Hence, the best performance of the skip search algorithm is when the search pattern is not found in the given text.**
- **The skip search converges quicker than the other algorithms. The m — 1 elements before the last m — 1 could be skipped in some cases. and further time can be saved. The algorithm can be further developed for multiple patterns multiple search problems. Conceptual graphs are used here because they enable humans to quickly understand what they should do.**