Search Algorithms Report

1. Linear Search

- Scans each line one by one to find the query.
- **Pros**: Simple, works on unsorted data.
- **Cons**: Slow for large files.
- **Time complexity**: O(n)

2. Binary Search

- Searches for the query in a sorted list using divide-and-conquer.
- **Pros**: Very fast for sorted data.
- **Time complexity**: O(log n)
- **Cons**: Requires pre-sorting the data.

3. Hash Set Lookup

- Loads lines into a hash set and checks if the query exists.
- **Time complexity**: O(1) average case.
- **Pros**: Very fast.
- **Cons**: High memory usage.

4. Regex Search:

- Uses Python’s regex engine to search for patterns in lines.
- **Time complexity**: O(n * m)
- **Pros**: Flexible and powerful for partial or pattern matching.
- **Cons**: Can be slower than direct methods for simple queries.

5. Rabin-Karp Algorithm

- Uses hashing to compare substring hash values.
- **Time complexity**: O(n) average, O(nm) worst.
- **Pros**: Efficient for multiple pattern matches.
- **Cons**: Hash collisions can occur.

6. Knuth-Morris-Pratt (KMP) Algorithm

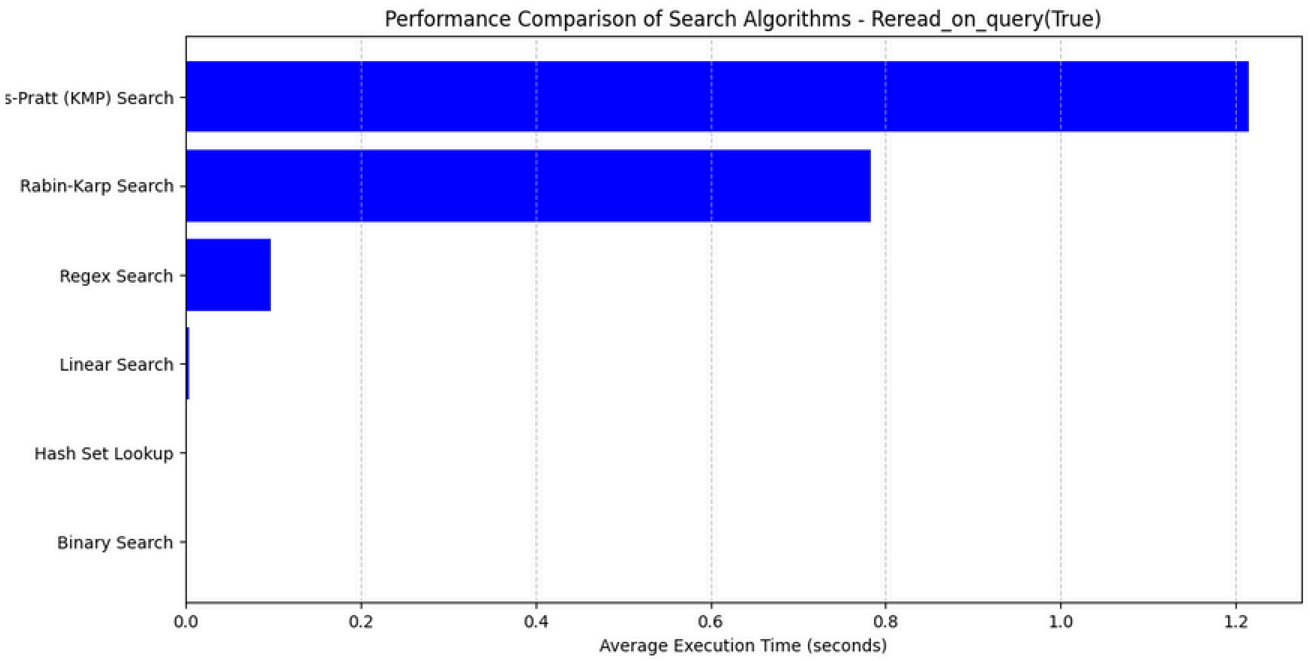
- Preprocesses the pattern to skip redundant comparisons.
- **Time complexity**: O(n + m)
- **Pros**: Efficient and deterministic.
- **Cons**: More complex to implement.

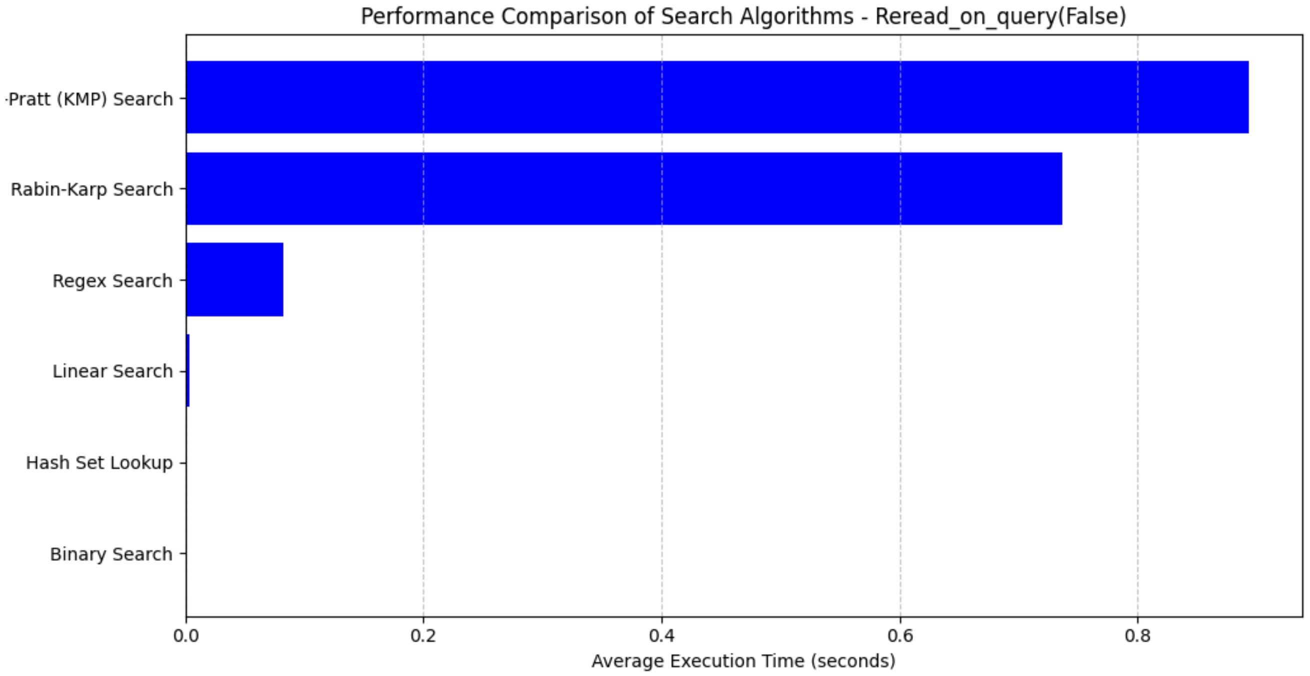
Performance Table Reread_on_query(True) - Time(seconds):

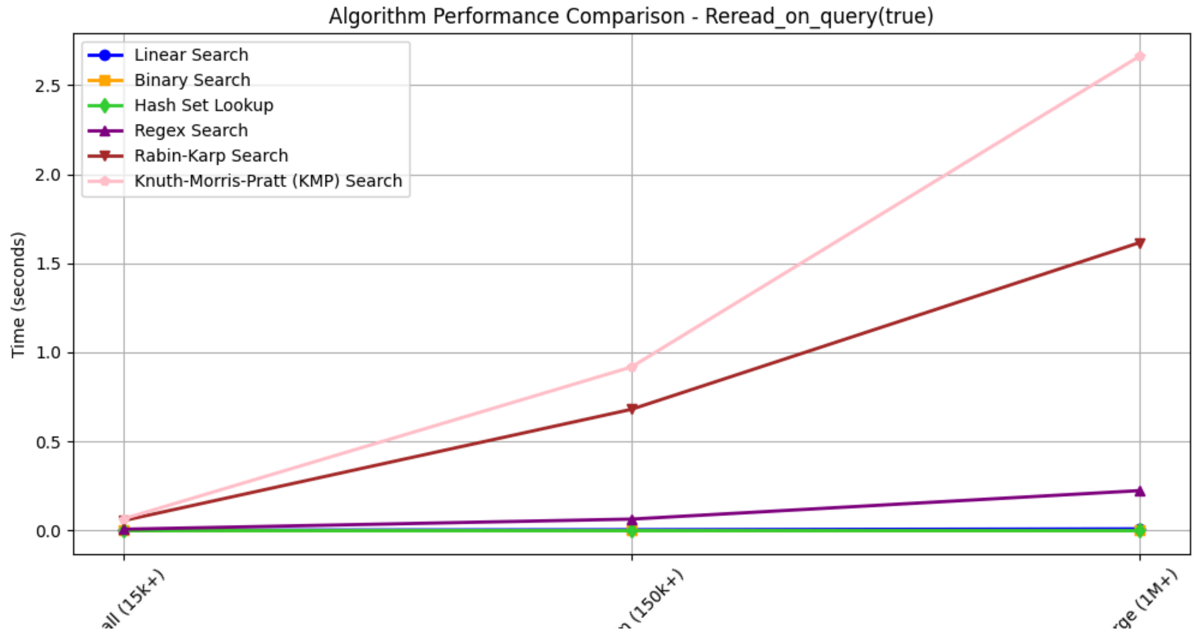
File Size(lines)	Large(1 million+)	Medium(150,000+)	Small(15,000+)
Algorithm			
Hash Set Lookup	0.000000	0.000000	0.000000
Binary Search	0.000000	0.000000	0.000012
Knuth-Morris-Pratt (KMP) Search	2.661365	0.917436	0.064959
Linear Search	0.008011	0.003018	0.000000
Rabin-Karp Search	1.615010	0.679579	0.054965
Regex Search	0.222860	0.062960	0.005996

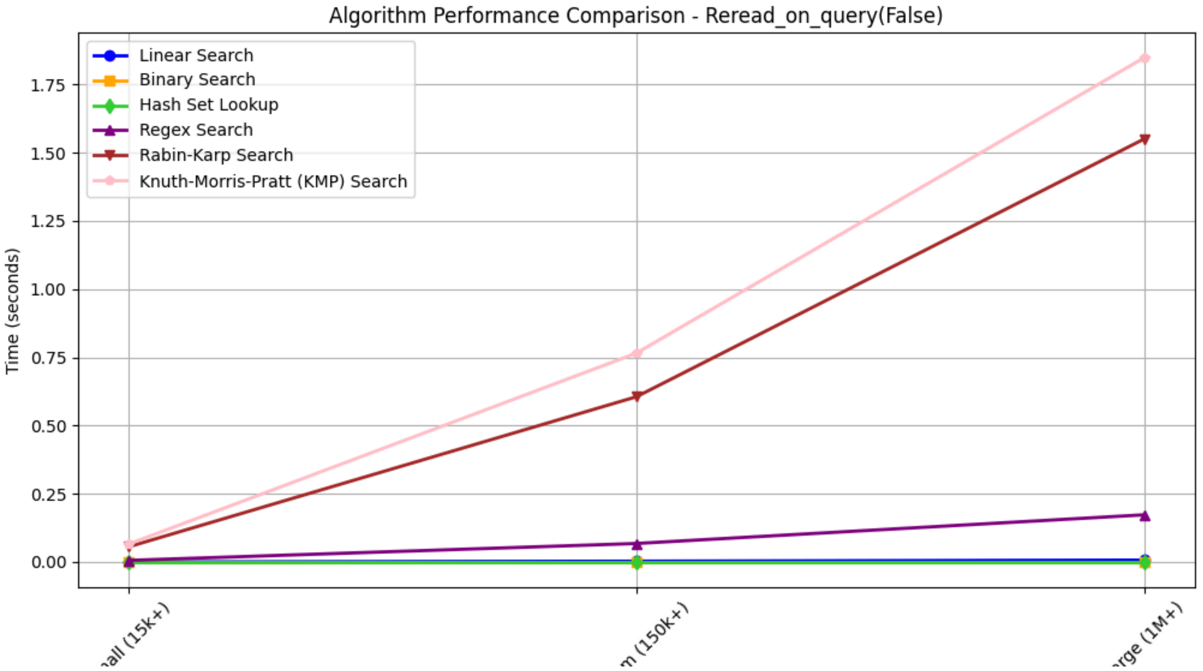
Performance Table Reread_on_query(False) - Time(seconds):

File Size(Lines)	Large(1 million+)	Medium(150,000+)	Small(15,000+)
Algorithm			
Binary Search	0.000000	0.000000	0.000000
Hash Set Lookup	0.000000	0.000000	0.000000
Knuth-Morris-Pratt (KMP) Search	1.847867	0.765531	0.065959
Linear Search	0.006999	0.003002	0.000000
Rabin-Karp Search	1.550042	0.605626	0.055966
Regex Search	0.172893	0.067954	0.005996


Performance Comparison of Search Algorithms - Reread_on_query(True)
s-Pratt (KMP) Search
Rabin-Karp Search
Regex Search
Linear Search
Hash Set Lookup
Binary Search
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Average Execution Time (seconds)


Performance Comparison of Search Algorithms - Reread_on_query(False)
-Pratt (KMP) Search
Rabin-Karp Search
Regex Search
Linear Search
Hash Set Lookup
Binary Search
0.0
0.2
0.4
0.6
0.8
Average Execution Time (seconds)


Algorithm Performance Comparison - Reread_on_query(true)
Linear Search
Binary Search
Hash Set Lookup
Regex Search
Rabin-Karp Search
Knuth-Morris-Pratt (KMP) Search
Time (seconds)
2.5
2.0
1.5
1.0
0.5
0.0
all (15k+)
n (150k+)
rge (1M+)


Algorithm Performance Comparison - Reread_on_query(False)
Linear Search
Binary Search
Hash Set Lookup
Regex Search
Rabin-Karp Search
Knuth-Morris-Pratt (KMP) Search
Time (seconds)
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
nall (15k+)
m (150k+)
arge (1M+)