# Comprehensive Benchmark Report: File Search Algorithms

Generated: 2025-11-03 22:09:10
Tested file sizes: 10000, 50000, 100000, 250000, 500000, 1000000
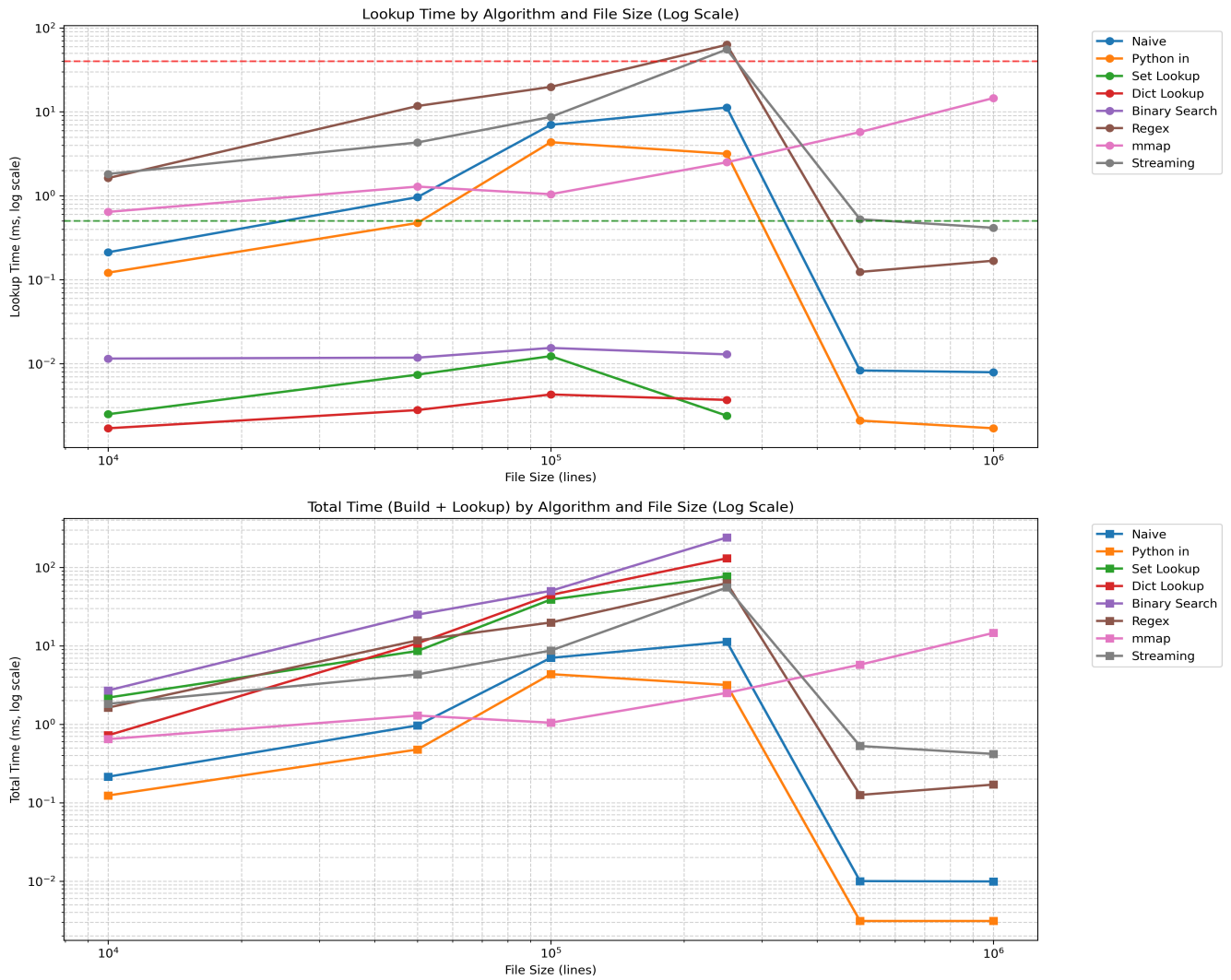
## Executive Summary

**Algorithms meeting 40ms requirement (REREAD_ON_QUERY=True):**
Python in, Naive, mmap, Streaming, Regex
**Algorithms meeting 0.5ms requirement (REREAD_ON_QUERY=False):**
Dict Lookup, Set Lookup, Binary Search

## Performance Requirement Validation (250k lines)

| Algorithm | Scenario | Avg Time (ms) | Requirement | Status |
|---|---|---|---|---|
| Naive | REREAD_ON_QUERY=True | 9.539 | 40ms | PASS |
| Python in | REREAD_ON_QUERY=True | 3.279 | 40ms | PASS |
| mmap | REREAD_ON_QUERY=True | 2.027 | 40ms | PASS |
| Streaming | REREAD_ON_QUERY=True | 28.401 | 40ms | PASS |
| Regex | REREAD_ON_QUERY=True | 34.332 | 40ms | PASS |
| Set Lookup | REREAD_ON_QUERY=False | 0.002400 | 0.5ms | PASS |
| Binary Search | REREAD_ON_QUERY=False | 0.005680 | 0.5ms | PASS |
| Dict Lookup | REREAD_ON_QUERY=False | 0.001520 | 0.5ms | PASS |

## Performance Analysis Charts

Lookup Time by Algorithm and File Size (Log Scale)



Total Time (Build + Lookup) by Algorithm and File Size (Log Scale)

## Algorithm Performance Summary

| Algorithm | Scenario | Avg Lookup (ms) | Max File Size | Status |
|---|---|---|---|---|
| Naive | REREAD_ON_QUERY=True | 3.242 | 1,000,000 | PASS |
| Python in | REREAD_ON_QUERY=True | 1.351 | 1,000,000 | PASS |
| Set Lookup | REREAD_ON_QUERY=False | 0.006 | 250,000 | PASS |
| Dict Lookup | REREAD_ON_QUERY=False | 0.003 | 250,000 | PASS |
| Binary Search | REREAD_ON_QUERY=False | 0.013 | 250,000 | PASS |
| Regex | REREAD_ON_QUERY=True | 16.066 | 1,000,000 | PASS |
| mmap | REREAD_ON_QUERY=True | 4.299 | 1,000,000 | PASS |
| Streaming | REREAD_ON_QUERY=True | 11.855 | 1,000,000 | PASS |

## Detailed Results (Sample)

| Algorithm | File Size | Found | Build Time (ms) | Lookup Time (ms) | Total Time (ms) |
|---|---|---|---|---|---|
| Naive | 10000 | 1 | 0.002 | 0.2124 | 0.2144 |
| Python in | 10000 | 1 | 0.0021 | 0.1214 | 0.1235 |
| Set Lookup | 10000 | 1 | 2.1839 | 0.0025 | 2.1864 |
| Dict Lookup | 10000 | 1 | 0.7216 | 0.0017 | 0.7233 |
| Binary Search | 10000 | 1 | 2.6873 | 0.0115 | 2.6988 |
| Regex | 10000 | 1 | 0.0027 | 1.6261 | 1.6288 |
| mmap | 10000 | 1 | 0.0027 | 0.6429 | 0.6456 |
| Streaming | 10000 | 1 | 0.0032 | 1.8191 | 1.8223 |
| Naive | 50000 | 1 | 0.002 | 0.9631 | 0.9651 |
| Python in | 50000 | 1 | 0.002 | 0.4743 | 0.4763 |
| Set Lookup | 50000 | 1 | 8.5844 | 0.0074 | 8.5918 |
| Dict Lookup | 50000 | 1 | 10.7566 | 0.0028 | 10.7594 |
| Binary Search | 50000 | 1 | 24.9298 | 0.0118 | 24.9416 |
| Regex | 50000 | 1 | 0.0044 | 11.7303 | 11.7347 |
| mmap | 50000 | 1 | 0.0039 | 1.2864 | 1.2903 |