Comprehensive Benchmark Report: File Search Algorithms

Generated: 2025-11-02 16:55:07
Tested file sizes: 10000, 50000, 100000, 250000, 500000, 1000000
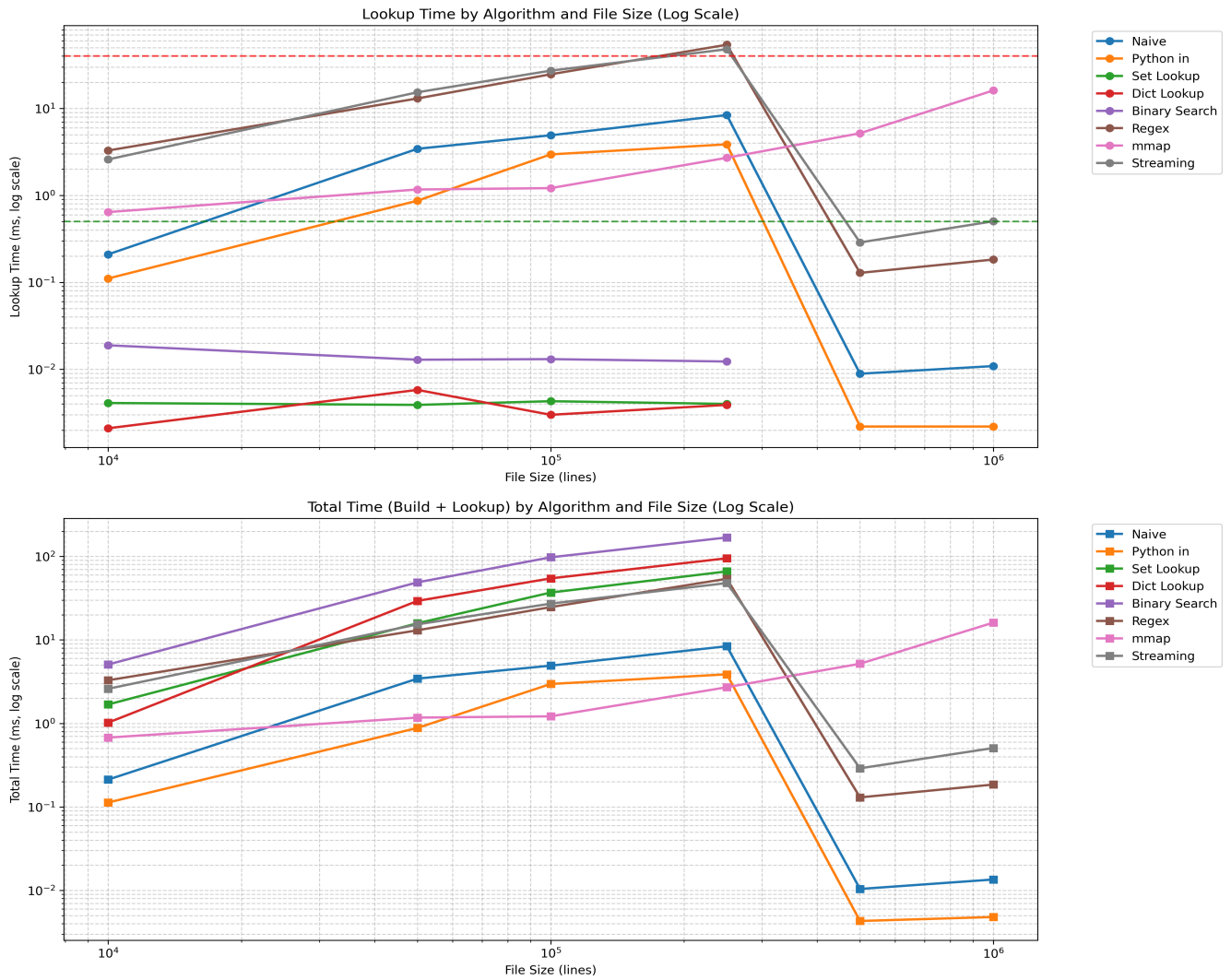
Executive Summary

Recommended algorithms (meet 0.5ms requirement): Dict Lookup, Set Lookup, Binary Search

Performance Requirement Validation (250k lines)

Algorithm	Scenario	Avg Time (ms)	Requirement	Status
Naive	REREAD_ON_QUERY=True	8.848	40ms	PASS
Python in	REREAD_ON_QUERY=True	3.042	40ms	PASS
Set Lookup	REREAD_ON_QUERY=False	0.005	0.5ms	PASS
Binary Search	REREAD_ON_QUERY=False	0.019	0.5ms	PASS
mmap	REREAD_ON_QUERY=True	2.445	40ms	PASS
Streaming	REREAD_ON_QUERY=True	57.770	40ms	FAIL

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	Avg Lookup (ms)	Max File Size	Meets 0.5ms	Scalability
Naive	2.829	1,000,000	NO	Good
Python in	1.302	1,000,000	NO	Good
Set Lookup	0.004	250,000	YES	N/A
Dict Lookup	0.004	250,000	YES	N/A
Binary Search	0.014	250,000	YES	N/A
Regex	15.889	1,000,000	NO	Good
mmap	4.503	1,000,000	NO	Moderate
Streaming	15.649	1,000,000	NO	Good

Detailed Results (Sample)

Algorithm	File Size	Found	Build Time (ms)	Lookup Time (ms)	Total Time (ms)
Naive	10000	1	0.0025	0.21	0.2125
Python in	10000	1	0.002	0.1109	0.1129
Set Lookup	10000	1	1.6837	0.0041	1.6878
Dict Lookup	10000	1	1.0202	0.0021	1.0223
Binary Search	10000	1	5.0612	0.0189	5.0801
Regex	10000	1	0.0033	3.2838	3.2871
mmap	10000	1	0.0335	0.6432	0.6767
Streaming	10000	1	0.0055	2.5939	2.5994
Naive	50000	1	0.0039	3.4367	3.4406
Python in	50000	1	0.0133	0.8694	0.8827
Set Lookup	50000	1	15.8395	0.0039	15.8434
Dict Lookup	50000	1	29.3409	0.0058	29.3467
Binary Search	50000	1	48.9144	0.0129	48.9273
Regex	50000	1	0.004	13.0353	13.0393
mmap	50000	1	0.0042	1.1696	1.1738