Comprehensive Benchmark Report: File Search Algorithms

Generated: 2025-11-01 07:12:10
Tested file sizes: 10000, 50000, 100000, 250000, 500000, 1000000
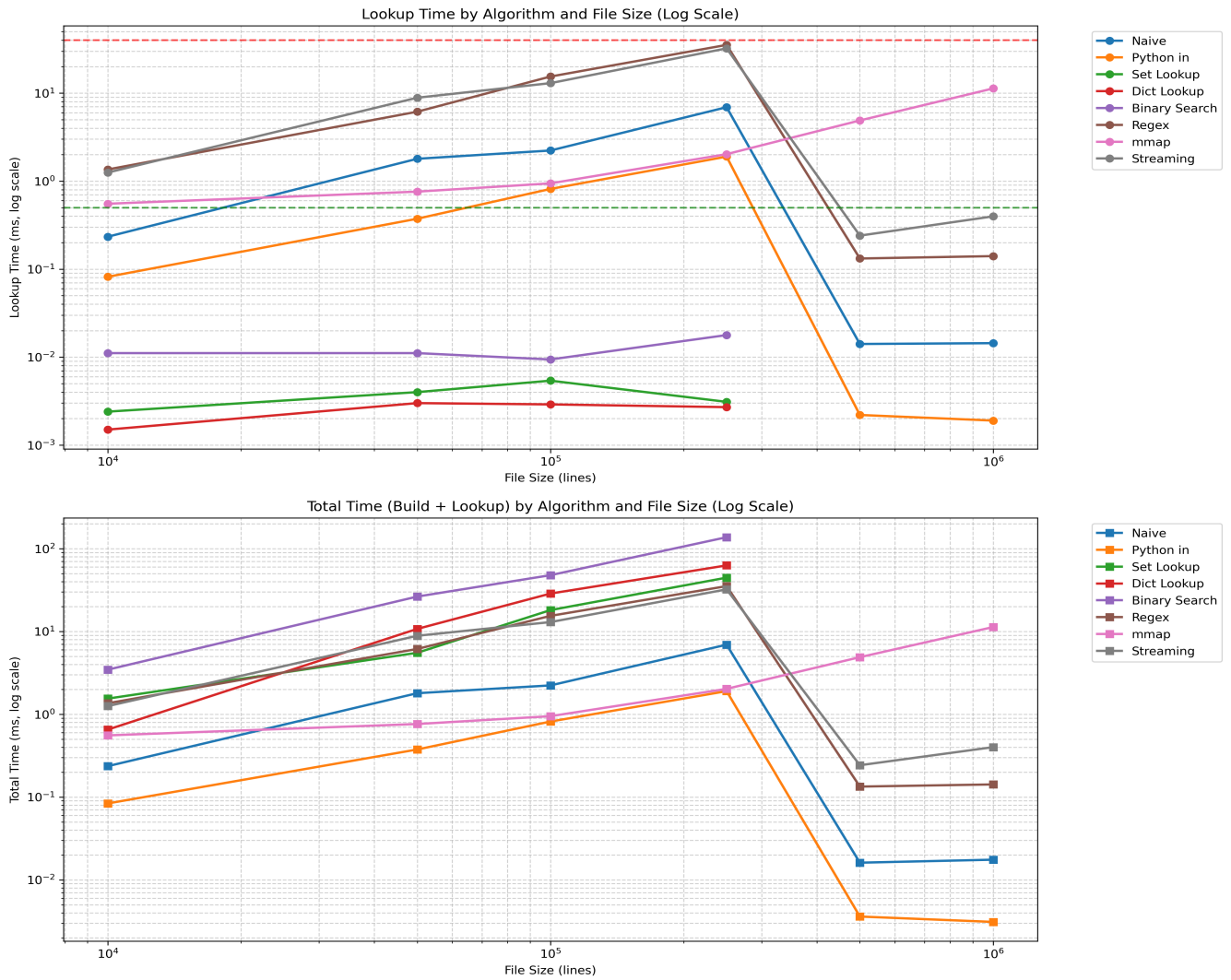
Executive Summary

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	8.943	40ms	PASS
Python in	REREAD_ON_QUERY=True	2.304	40ms	PASS
Set Lookup	REREAD_ON_QUERY=False	0.009	0.5ms	PASS
Binary Search	REREAD_ON_QUERY=False	0.020	0.5ms	PASS
mmap	REREAD_ON_QUERY=True	2.308	40ms	PASS
Streaming	REREAD_ON_QUERY=True	25.233	40ms	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	Avg Lookup (ms)	Max File Size	Meets 0.5ms	Scalability
Naive	1.867	1,000,000	NO	Good
Python in	0.529	1,000,000	NO	Good
Set Lookup	0.004	250,000	YES	N/A
Dict Lookup	0.003	250,000	YES	N/A
Binary Search	0.012	250,000	YES	N/A
Regex	9.750	1,000,000	NO	Good
mmap	3.412	1,000,000	NO	Moderate
Streaming	9.330	1,000,000	NO	Good

Detailed Results (Sample)

Algorithm	File Size	Found	Build Time (ms)	Lookup Time (ms)	Total Time (ms)
Naive	10000	1	0.0026	0.2335	0.2361
Python in	10000	1	0.0019	0.0818	0.0837
Set Lookup	10000	1	1.5428	0.0024	1.5452
Dict Lookup	10000	1	0.6512	0.0015	0.6527
Binary Search	10000	1	3.4446	0.0111	3.4557
Regex	10000	1	0.002	1.3565	1.3585
mmap	10000	1	0.0024	0.5531	0.5555
Streaming	10000	1	0.0025	1.2515	1.254
Naive	50000	1	0.0026	1.7932	1.7958
Python in	50000	1	0.0017	0.3735	0.3752
Set Lookup	50000	1	5.5241	0.004	5.5281
Dict Lookup	50000	1	10.7336	0.003	10.7366
Binary Search	50000	1	26.3858	0.0111	26.3969
Regex	50000	1	0.0033	6.1449	6.1482
mmap	50000	1	0.0027	0.7592	0.7619