

Comprehensive Benchmark Report: File Search Algorithms

Generated: 2025-11-02 16:43:18

Tested file sizes: 10000, 50000, 100000, 250000, 500000, 1000000

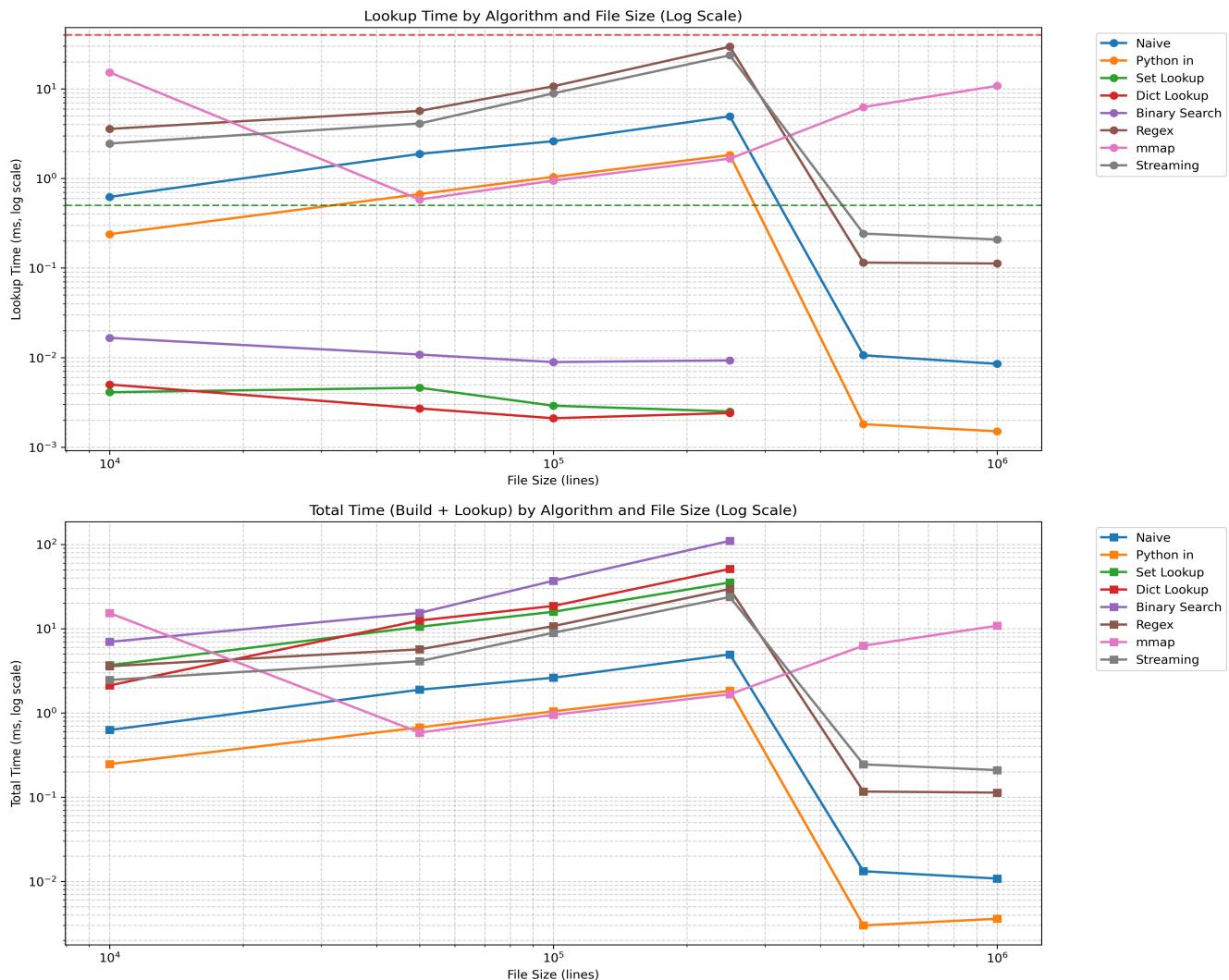
Executive Summary

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	6.278	40ms	PASS
Python in	REREAD_ON_QUERY=True	2.532	40ms	PASS
Set Lookup	REREAD_ON_QUERY=False	0.006	0.5ms	PASS
Binary Search	REREAD_ON_QUERY=False	0.023	0.5ms	PASS
mmap	REREAD_ON_QUERY=True	1.953	40ms	PASS
Streaming	REREAD_ON_QUERY=True	27.491	40ms	PASS

Performance Analysis Charts



Algorithm Performance Summary

Algorithm	Avg Lookup (ms)	Max File Size	Meets 0.5ms	Scalability
Naive	1.677	1,000,000	NO	Good
Python in	0.629	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	0.011	250,000	YES	N/A
Regex	8.286	1,000,000	NO	Good
mmap	5.927	1,000,000	NO	Good
Streaming	6.605	1,000,000	NO	Good

Detailed Results (Sample)

Algorithm	File Size	Found	Build Time (ms)	Lookup Time (ms)	Total Time (ms)
Naive	10000	1	0.0048	0.6216	0.6264
Python in	10000	1	0.0081	0.2386	0.2467
Set Lookup	10000	1	3.6699	0.0041	3.674
Dict Lookup	10000	1	2.1094	0.005	2.1144
Binary Search	10000	1	6.9449	0.0166	6.9615
Regex	10000	1	0.0046	3.5723	3.5769
mmap	10000	1	0.0035	15.2932	15.2967
Streaming	10000	1	0.0034	2.4535	2.4569
Naive	50000	1	0.0036	1.8795	1.8831
Python in	50000	1	0.0026	0.6687	0.6713
Set Lookup	50000	1	10.5195	0.0046	10.5241
Dict Lookup	50000	1	12.4819	0.0027	12.4846
Binary Search	50000	1	15.3524	0.0108	15.3632
Regex	50000	1	0.0029	5.6749	5.6778
mmap	50000	1	0.0017	0.5818	0.5835