

Comprehensive Benchmark Report: File Search Algorithms

Generated: 2025-11-01 07:09:26

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

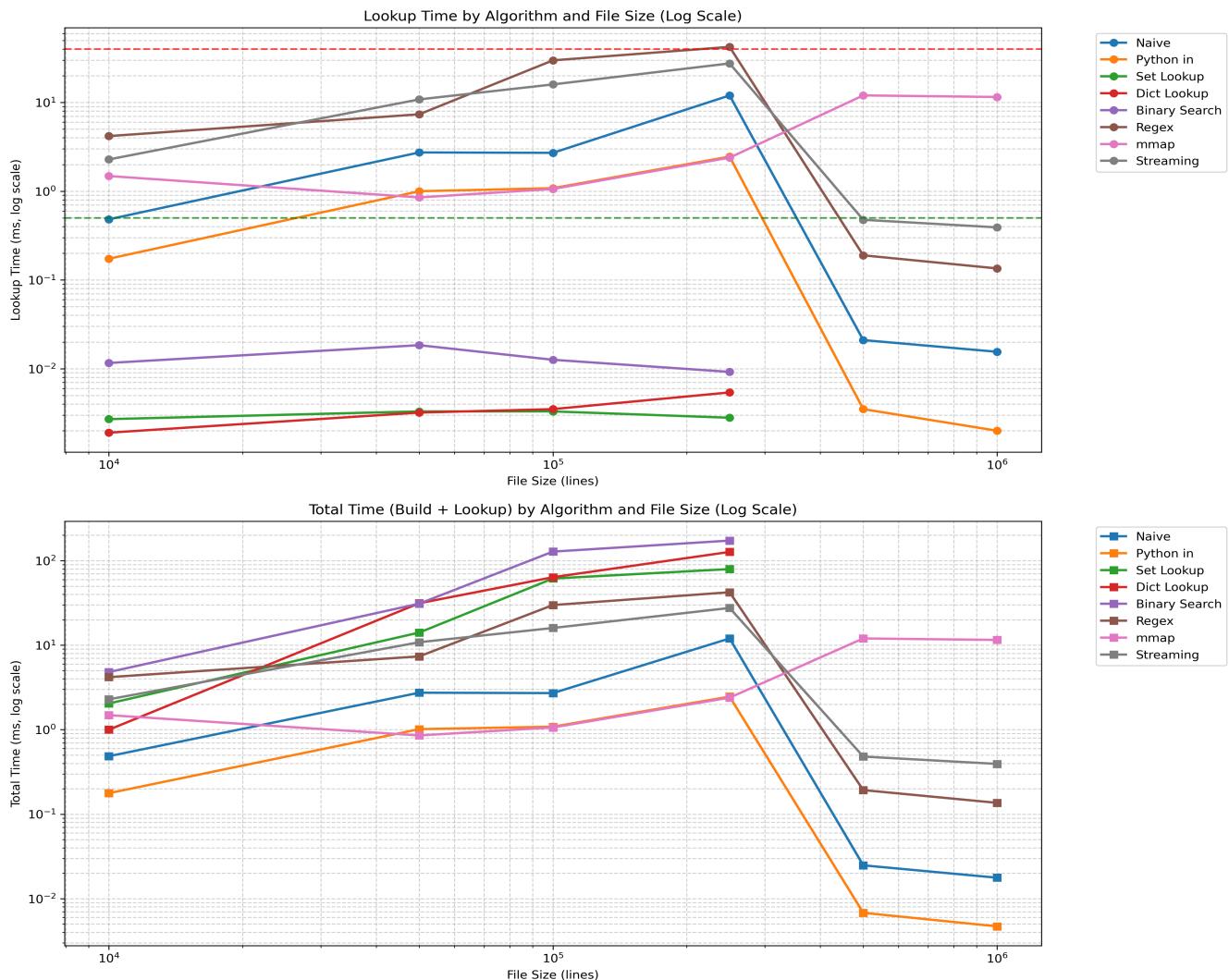
Executive Summary

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

Performance Requirement Validation (250k lines)

Algorithm	Scenario	Avg Time (ms)	Requirement	Status
Naive	REREAD_ON_QUERY=True	14.093	40ms	PASS
Python in	REREAD_ON_QUERY=True	5.279	40ms	PASS
Set Lookup	REREAD_ON_QUERY=False	0.010	0.5ms	PASS
Binary Search	REREAD_ON_QUERY=False	0.019	0.5ms	PASS
mmap	REREAD_ON_QUERY=True	2.661	40ms	PASS
Streaming	REREAD_ON_QUERY=True	49.656	40ms	FAIL

Performance Analysis Charts



Algorithm Performance Summary

Algorithm	Avg Lookup (ms)	Max File Size	Meets 0.5ms	Scalability
Naive	2.994	1,000,000	NO	Good
Python in	0.786	1,000,000	NO	Good
Set Lookup	0.003	250,000	YES	N/A
Dict Lookup	0.004	250,000	YES	N/A
Binary Search	0.013	250,000	YES	N/A
Regex	14.006	1,000,000	NO	Good
mmap	4.889	1,000,000	NO	Moderate
Streaming	9.588	1,000,000	NO	Good

Detailed Results (Sample)

Algorithm	File Size	Found	Build Time (ms)	Lookup Time (ms)	Total Time (ms)
Naive	10000	1	0.003	0.4825	0.4855
Python in	10000	1	0.0037	0.1735	0.1772
Set Lookup	10000	1	2.0396	0.0027	2.0423
Dict Lookup	10000	1	0.9978	0.0019	0.9997
Binary Search	10000	1	4.7939	0.0116	4.8055
Regex	10000	1	0.0037	4.1809	4.1846
mmap	10000	1	0.0035	1.4825	1.486
Streaming	10000	1	0.006	2.282	2.288
Naive	50000	1	0.003	2.7354	2.7384
Python in	50000	1	0.0147	0.9997	1.0144
Set Lookup	50000	1	14.1066	0.0033	14.1099
Dict Lookup	50000	1	31.357	0.0032	31.3602
Binary Search	50000	1	31.0252	0.0184	31.0436
Regex	50000	1	0.005	7.362	7.367
mmap	50000	1	0.0035	0.8509	0.8544