Comprehensive Benchmark Report: File Search Algorithms

Generated: 2025-11-03 06:44:05
Tested file sizes: 10000, 50000, 100000, 250000, 500000, 1000000
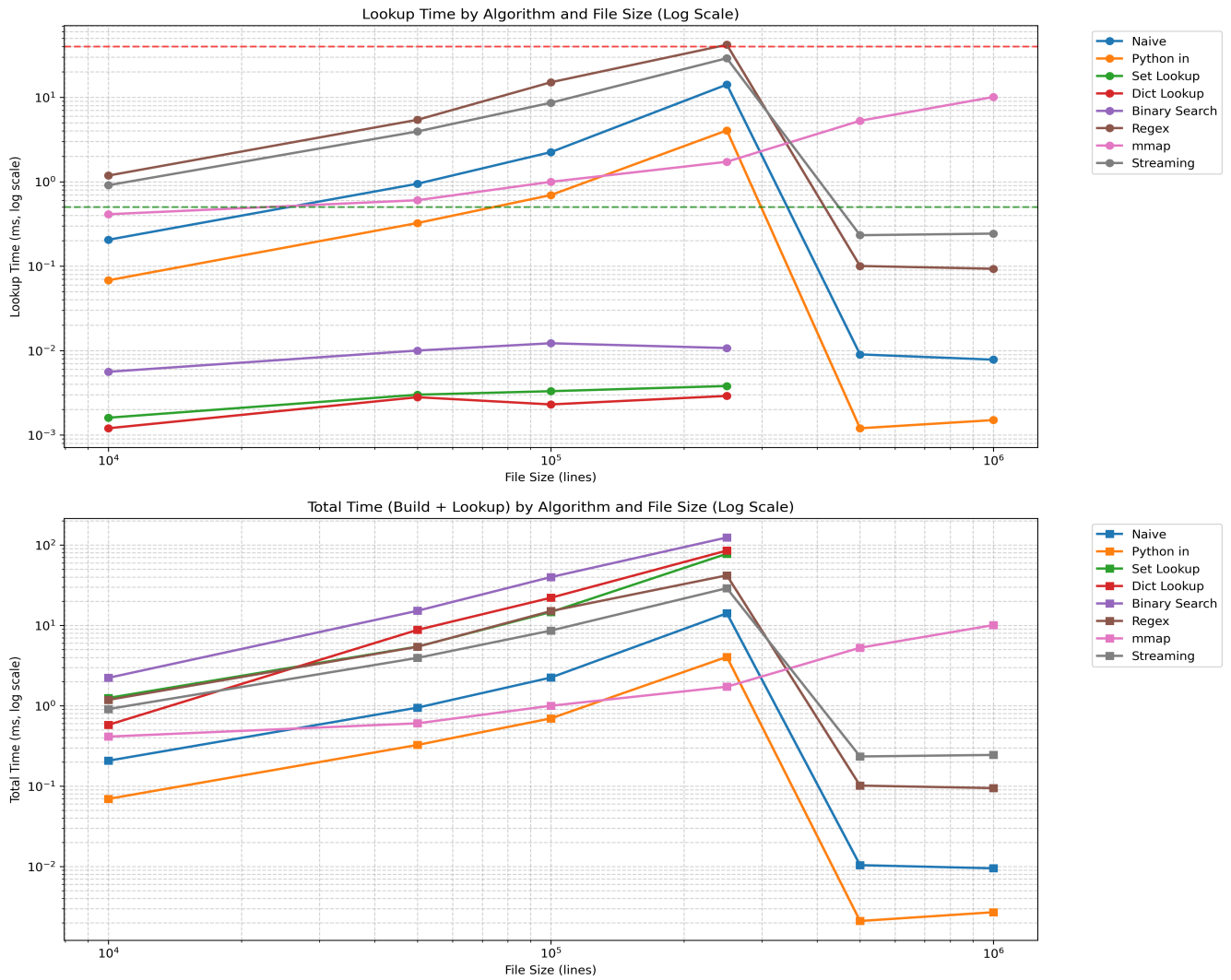
Executive Summary

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	6.805	40ms	PASS
Python in	REREAD_ON_QUERY=True	3.333	40ms	PASS
Set Lookup	REREAD_ON_QUERY=False	0.005	0.5ms	PASS
Binary Search	REREAD_ON_QUERY=False	0.016	0.5ms	PASS
mmap	REREAD_ON_QUERY=True	1.630	40ms	PASS
Streaming	REREAD_ON_QUERY=True	23.137	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	2.926	1,000,000	NO	Good
Python in	0.857	1,000,000	NO	Good
Set Lookup	0.003	250,000	YES	N/A
Dict Lookup	0.002	250,000	YES	N/A
Binary Search	0.010	250,000	YES	N/A
Regex	10.630	1,000,000	NO	Good
mmap	3.182	1,000,000	NO	Moderate
Streaming	7.162	1,000,000	NO	Good

Detailed Results (Sample)

Algorithm	File Size	Found	Build Time (ms)	Lookup Time (ms)	Total Time (ms)
Naive	10000	1	0.0019	0.2053	0.2072
Python in	10000	1	0.0015	0.0679	0.0694
Set Lookup	10000	1	1.249	0.0016	1.2506
Dict Lookup	10000	1	0.5774	0.0012	0.5786
Binary Search	10000	1	2.2198	0.0056	2.2254
Regex	10000	1	0.0015	1.1824	1.1839
mmap	10000	1	0.0015	0.4118	0.4133
Streaming	10000	1	0.0015	0.9086	0.9101
Naive	50000	1	0.0019	0.9465	0.9484
Python in	50000	1	0.0014	0.3242	0.3256
Set Lookup	50000	1	5.4484	0.003	5.4514
Dict Lookup	50000	1	8.7945	0.0028	8.7973
Binary Search	50000	1	15.1944	0.01	15.2044
Regex	50000	1	0.0031	5.4169	5.42
mmap	50000	1	0.0013	0.6038	0.6051