ISONQ v7.0.1: Performance Evaluation of a Local-First Semantic Search System

ISONQ LLC

Lakewood, Colorado, USA • December 2025

Abstract— We present a quantitative evaluation of ISONQ v7.0.1, a deterministic document retrieval system designed for privacy-sensitive environments requiring zero data egress. Benchmarks conducted on a corpus of n = 1,045 documents demonstrate median query latency P_{50} = 6.25ms (95% CI: 5.76–7.36ms), sustained throughput of 152.43 queries/second, and memory footprint of 18.33 MB. Statistical validation confirms sample adequacy (n = 100 queries, seed = 42). Future work addresses retrieval coverage improvements targeting MRR > 0.80.

1. SYSTEM CONFIGURATION

Parameter	Value
ISONQ Version	7.0.1
Processor	Intel Core i7-1365U (12 cores)
Memory	15,963 MB
Operating System	Windows 11 (Build 26100)
Benchmark Seed	42

Table 1. Test environment specifications.

2. CORPUS STATISTICS

Source Type	Count	Proportion
Files	950	90.91%
Emails	86	8.23%
Teams Messages	9	0.86%
Total	**1,045**	**100.00%**

Table 2. Index composition by source type.

Index size: 2,854,912 bytes. Per-document overhead:

$$\text{Overhead} = 2{,}854{,}912 / 1{,}045 = 2{,}732 \text{ bytes/doc}$$

3. LATENCY ANALYSIS

Statistic	Value (ms)
Minimum	0.86
P_{50} (Median)	6.25
Mean ($\bar{x}$)	6.56
P_{95}	13.89
P_{99}	20.69
Maximum	20.69
Std. Deviation (σ)	4.01
Coefficient of Variation	61.1%

Table 3. Latency distribution (n = 100 queries).

3.1 Confidence Interval

The 95% confidence interval for mean latency is computed using the Student's t-distribution:

$$CI_{95\%} = \bar{x} \pm t_{\alpha/2,\, n-1} \cdot (\sigma / \sqrt{n})$$

Substituting observed values:

$$CI_{95\%} = 6.56 \pm 1.984 \cdot (4.01 / \sqrt{100}) = [5.76, 7.36] \text{ ms} \quad (3)$$

3.2 Sample Adequacy

Required sample size for ±1ms margin of error at 95% confidence:

$$n = (z \cdot \sigma / E)^2 = (1.96 \cdot 4.01 / 1.0)^2 = 61.8 \approx 62 \quad (4)$$

The observed n = 100 exceeds the minimum requirement of 62 samples.

3.3 Tail Latency Ratio

$$P_{99} / P_{50} = 20.69 / 6.25 = 3.31 \quad (1)$$

4. THROUGHPUT

Metric	Value
Queries per Second	152.43
Total Queries	100
Successful	100
Failed	0
Success Rate	100.00%

Table 4. Throughput measurements.

4.1 Daily Capacity

$$C_{daily} = 152.43 \times 86{,}400 = 13.17 \times 10^6 \text{ queries/day} \tag{6}$$

4.2 Enterprise Deployment Headroom

For a 50-user organization averaging 20 queries/user/day:

$$\text{Headroom} = (13.17 \times 10^6) / (50 \times 20) = 13{,}170\times \tag{7}$$

5. MEMORY PROFILE

Metric	Value (MB)
Baseline	17.17
Peak	19.34
Index Resident	18.33
Delta (Peak − Baseline)	2.17

Table 5. Memory consumption during query operations.

6. RETRIEVAL QUALITY

Metric	Value	Definition
MRR	0.560	Mean Reciprocal Rank
P@1	0.560	Precision at rank 1
P@5	0.374	Precision at rank 5
P@10	0.315	Precision at rank 10

Table 6. Retrieval quality metrics.

6.1 Metric Definitions

Mean Reciprocal Rank:

$$MRR = (1/|Q|) \times \Sigma_{i=1}^{|Q|} (1/rank_i) \tag{8}$$

Precision at K:

$$P@K = |\{relevant\} \cap \{retrieved_K\}| / K \tag{9}$$

7. PLANNED IMPROVEMENTS

Enhancement	Current	Target
Mean Reciprocal Rank	0.560	> 0.800
Zero-Result Rate	44.0%	< 10.0%
Recall Coverage	56.0%	> 90.0%

Table 7. Retrieval improvement targets (Q1 2026).

Fuzzy Matching Fallback: Levenshtein distance threshold for approximate string matching when exact signals fail.

$$d(q, d) = Levenshtein(q, d) < \tau \tag{10}$$

Query Suggestion: Maximum likelihood estimation for "did you mean" recommendations.

$$q' = \text{argmax}_{q' \in V} P(q' \mid q)$$

8. SUMMARY

P_{50} Latency	6.25 ms
95% Confidence Interval	[5.76, 7.36] ms
Throughput	152.43 QPS
Memory Footprint	18.33 MB
Index Efficiency	2,732 bytes/doc
Daily Capacity	13.17×10^6 queries

Reproducibility: Benchmark harness v1.0.0, seed = 42, n = 100. All measurements recorded on 2025-12-10T19:15:22Z.
Contact: ISONQ LLC, Lakewood, CO, USA • support@isonq.com