

A Functional Pilot for Certified Freshness-Aware Semantic–Spatial Range Retrieval

Taimoor Ahmed

Abstract—Geographic applications need every object inside a radius that satisfies a semantic threshold, yet embedding indexes return approximate top-ranked lists and may omit qualifying records silently. We present FRESH-GEORANGE, a semantic-spatial range design that separates source-watermark freshness from optional record age. Geographic cells and semantic microblocks provide admissible pruning bounds; a graph proposes verification order but supplies no correctness evidence. Exact mode scans every nonprunable block and the delta overlay. Certified mode may stop early and reports a deterministic query-specific recall lower bound from verified answers and unresolved records. A reproducible CPU pilot uses 2,500 real OpenFlights airport records, a 2,000-record base, and 740 simulated insert, delete, and text-revision events; it evaluates 180 unique queries over five seeds. Exact mode achieved 100.00% set recall on every query. The 95-percent mode achieved 99.91% empirical mean recall with a 99.41% reported mean certificate and no observed bound violation. However, its 7.24 ms median latency was 5.85 times the 1.24 ms spatial-first exact baseline, and full-history delta replay became slower than rebuilding at larger batches. The prototype therefore validates the completeness mechanism, not performance superiority or production freshness. Submission-scale evaluation requires real map diffs, official recent baselines, and truly incremental versioned maintenance.

Index Terms—semantic–spatial retrieval, range reporting, freshness watermark, certified recall, dynamic index, exact verification

I. INTRODUCTION

Location-aware search increasingly depends on meanings beyond literal keyword overlap. A request for “emergency treatment for a severe allergic reaction” should retrieve a nearby facility advertising “anaphylaxis services.” Learned systems support such matches by embedding queries and points of interest (POIs). LIST couples learned routing with embedding relevance [1], while Mesh integrates geographic range constraints with high-dimensional approximate nearest-neighbor (ANN) search [2]. Both target a ranked top- k list. Operational tasks such as emergency-resource discovery and regulatory auditing instead need threshold enumeration—every active object inside a stated radius whose semantic similarity reaches a threshold. The caller must know whether the returned list is exhaustive, rather than merely well ranked; an incomplete top- k list cannot safely represent that set.

For top- k , ANN may stop after collecting enough competitive candidates. Under a threshold, cardinality is unknown and every skipped partition could hide an answer. Exact rescoring removes false positives but cannot recover unseen positives; it certifies precision, not completeness. A valid certificate must therefore account for partitions that heuristic traversal

never reaches, not infer completeness from the retrieved candidates. Range-complete foundations include Boolean radius engines [3], WISK’s learned partitions for exact region-and-keyword processing [4], and geographic-radius enumeration under a Jaccard threshold [5]. Word2Vec has also been used for continuous spatial range processing [6]–[8]. These interfaces do not jointly define dense-semantic threshold enumeration, versioned visibility, and a query-specific completeness certificate.

Mutability creates a second gap. POI catalogs receive insertions, deletions, and text revisions continually, so an answer may be correct for an old index but wrong for the promised version. Per-edit rebuilding is costly; periodic rebuilding hides new resources and can resurrect deleted ones between rebuilds. Dynamic spatial-keyword systems maintain queries as objects change [9]–[11], while vector systems support streaming or incremental maintenance [12]–[14]. LIST is not purely static: it supports cluster-assigned insertion, deletion, and index-only retraining after sustained inserts or distribution shift [1], [15]. It nevertheless remains top- k and supplies no radius-plus-threshold completeness contract. We therefore distinguish an update-visibility watermark, a correctness promise about reflected commits, from optional record-age semantics, an application predicate over old observations. A recent observation can be absent behind a stale watermark, while an old observation can be fully visible; treating both as one “freshness” value obscures correctness.

FRESH-GEORANGE addresses this conjunction. Its query specifies a location, radius, semantic vector, similarity threshold, and maximum visibility lag. The index combines geographic cells, semantic microblocks with admissible cosine bounds, an ANN graph used only to order work, and a latest-write delta overlay for inserts, replacements, and tombstones. A watermark identifies the visible snapshot. Exact mode visits every non-prunable spatially relevant block and scans the entire visible delta. Certified mode may stop earlier, but conservatively counts unresolved physical records and derives a deterministic per-query recall lower bound from that count and the verified positives. Tombstones and replacements suppress superseded base versions, and the delta and watermark are published atomically. Approximation therefore changes cost and certificate tightness, not the truth of the bound.

The closest adjacent field is range-filtered ANN (RFANN), which restricts dense-vector top- k search by an interval on one ordered attribute. SeRF and iRangeGraph optimize static filters [16]–[18]; Dynamic RFANNS, DIGRA, RangePQ, and WoW add dynamic or incremental maintenance [19]–[22]. Yet an interval is not a two-dimensional geodesic disk, and approximate top- k is not threshold enumeration. Mapping a disk to

longitude intervals or a space-filling key introduces different selectivity and boundary behavior. Vector-radius retrieval approximately reports all vectors within an embedding distance but lacks a geographic predicate and mutable snapshot [23]. ConANN provides a distribution-free conformal risk guarantee for ANN [24]; our target is instead deterministic, bound-based accounting for one declared snapshot. We claim novelty only for this conjunction, not for any ingredient alone.

In our CPU pilot, exact mode attains 100.00% mean recall and certified mode attains 99.91% with a 99.41% mean lower bound; however, certified retrieval costs $5.85\times$ the median latency of spatial-first exact scanning. This negative result supports feasibility of the guarantee, not deployment superiority.

This paper makes three contributions:

- We formalize *fresh semantic-spatial range enumeration*: all visible POIs inside a geodesic radius and above a dense-similarity threshold, under an explicit update-watermark contract that distinguishes visibility lag from record age.
- We design a dual spatial-semantic index with a latest-write delta overlay, atomic watermark publication, and safe geometric and cosine bounds. Its ANN component only prioritizes verification; exact mode exhausts every non-prunable block, while certified mode returns a sound query-specific recall lower bound after exact delta scanning.
- We provide a reproducible CPU pilot that evaluates set recall and precision, certificate validity and tightness, tail latency, complete-history replay cost, staleness, scripted drift instrumentation, selectivity, and ablations under deterministic insertion, deletion, and text-revision streams, and we state the evidence still required for a submission-scale claim.

The remainder of the paper is organized as follows. Section II positions the problem against spatial-textual, filtered-vector, and dynamic indexes. Section III formulates the query and presents the index, update contract, and certification argument. Section IV describes the reproducible pilot and reports its results and limitations. Section V concludes.

II. RELATED WORK

Spatial-textual retrieval. Classical structures combine a spatial hierarchy with an inverted text index. The IR-tree is a representative design for geographic document search [25]; later main-memory processing supports point-plus-radius Boolean queries with perfect result recall [3]. WISK learns workload-aware spatial partitions while preserving exact keyword-query processing and buffers insertions before localized retraining [4]. Tampakis et al. are especially close because their filter-refine index returns every object inside a radius whose keyword-set Jaccard similarity exceeds a threshold [5]. FRESH-GEORANGE changes the semantic predicate from a sparse token-set measure to dense cosine similarity and couples it to a mutable-snapshot certificate; it does not treat threshold range completeness itself as new.

Embedding-aware spatial retrieval has mainly evolved around ranking. The S^2R -tree uses semantic pivots for spatial

keyword search [26], DrW develops neural relevance modeling [27], and LIST jointly learns object/query representations, cluster routing, and spatial relevance for embedding-based top- k spatial keyword queries [1]. LIST explicitly supports insertion, deletion, and index-only retraining; calling it static would erase a material capability. What it does not offer is all-object output under both a geographic radius and a similarity threshold, nor a per-query lower bound that accounts for work not visited. Mesh directly combines a true spatial range with vector ANN [2], making it the closest integrated vector baseline, but still returns approximate top- k results. Oh et al. maintain Word2Vec-based range results over moving objects [6]; FAST, Salgado et al., and Dong et al. likewise establish that continuous and dynamic spatial-keyword processing is mature [9]–[11]. Their motion, Boolean, or top- k semantics differ from versioned POI create/delete/text-replace operations with a declared visibility watermark.

Filtered and range-reporting vector search. Filtered-DiskANN demonstrates graph search with categorical filters [28]; SeRF, iRangeGraph, UNIFY, and window-filter methods exploit ordered filter structure [16], [17], [29], [30]. More recent work makes the filtered index mutable: Dynamic RFANNS, DIGRA, and RangePQ address changing data [19]–[21], while WoW is a fully incremental insert-from-empty design but leaves frequent deletion and in-place update as open concerns [22]. These methods inform graph routing, segment organization, and update engineering. Their conventional query asks for the nearest k vectors satisfying a one-dimensional range, however; adapting longitude or a space-filling key to that interface does not faithfully represent a circular geographic predicate. Manohar et al. instead study approximate reporting of every vector within a vector-distance radius [23], which is the direct threshold-output precedent, but lacks a geographic constraint, snapshot visibility, and deterministic accounting of missed objects.

Freshness, drift, and guarantees. HNSW supplies the graph-search foundation used by many ANN systems [31], but benchmark recall remains an empirical aggregate. SPFresh performs incremental in-place vector-index updates [12]; VStream addresses distributed streaming search [13]; and Quake adapts partitioning and predicts parameters for requested recall levels under changing workloads [14]. These systems motivate delta maintenance and drift-triggered consolidation, not the certificate itself. ConANN provides an important formal counterpoint by controlling ANN risk with distribution-free conformal methods [24]. Its statistical guarantee and assumptions differ from a deterministic statement about one radius-and-threshold result set. In FRESH-GEORANGE, admissible block bounds and a conservative count of every unresolved visible record yield the certificate; learned routing and ANN traversal may reduce work but are never trusted as proof. Table I summarizes why no individual ingredient is unprecedented and why the proposed contribution must be evaluated as a conjunction.

TABLE I

CLOSEST-WORK LANDSCAPE. “ALL” MEANS THRESHOLD/RANGE ENUMERATION RATHER THAN A FIXED TOP- k LIST; GUARANTEES ARE CHARACTERIZED BY WHAT THE CITED INTERFACE ESTABLISHES. FRESH-GEORANGE’S PROPOSED NOVELTY IS THE CONJUNCTION IN THE LAST ROW.

System	Spatial filter	Semantic model	Output	Update model	Recall/completeness evidence
Lee et al. [3]	2-D point radius	Boolean keywords	All	Not central	Perfect result recall
Tampakis et al. [5]	2-D radius	Jaccard threshold	All	Static	Exact filter-refine
Oh et al. [6]	2-D range	Word2Vec	Range results	Moving objects	No query certificate established
WISK [4]	2-D region	Boolean keywords	All	Buffered inserts; local re-training	Exact keyword processing
Dong et al. [10]	Spatial proximity	Keyword relevance	Top- k	Dynamic objects	Maintained top- k semantics
LIST [1]	Learned spatial relevance	Dense embeddings	Top- k	Insert; delete; index-only retraining	Empirical top- k effectiveness
Mesh [2]	2-D spatial range	Dense vectors	Approximate top- k	Static	Empirical ANN recall
Vector-radius retrieval [23]	None	Vector-distance radius	Approximate all	Static	Empirical reporting recall
SeRF [16]	Scalar interval	Dense vectors	Approximate top- k	Static	Empirical ANN recall
iRangeGraph [17]	Scalar interval	Dense vectors	Approximate top- k	Static	Empirical ANN recall
Dynamic RFANNS [19]	Scalar interval	Dense vectors	Approximate top- k	Dynamic	Empirical ANN recall
DIGRA [20]	Scalar interval	Dense vectors	Approximate top- k	Dynamic graph	Empirical ANN recall
RangePQ [21]	Scalar interval	Dense vectors	Approximate top- k	Dynamic quantization index	Empirical ANN recall
WoW [22]	Scalar window	Dense vectors	Approximate top- k	Incremental insert-from-empty	Empirical ANN recall
FRESH-GEORANGE	2-D geodesic radius	Dense threshold	All / certified	Versioned insert, delete, replace	Exact mode or deterministic per-query lower bound

III. METHOD

A. Problem Formulation

Unlike top- k search, our threshold query must return every live object that passes its spatial, semantic, and freshness predicates. Define

$$\begin{aligned} x_i &= (i, \nu_i, p_i, v_i, a_i, o_i), \\ Q &= (\ell, r, q, \tau, t_q, A_{\max}, \Delta_{\max}, \rho_{\text{req}}). \end{aligned} \quad (1)$$

where p_i is a location, $v_i \in \mathbb{R}^d$ a unit embedding, a_i an observation day, and o_i an upsert or delete (whose non-key fields may be absent). Sequence ν_i totally orders an identifier’s events. Query Q specifies center, radius $r \geq 0$, unit vector, inclusive cosine threshold $\tau \in [-1, 1]$, time, maximum record age and watermark lag, and requested certificate $\rho_{\text{req}} \in [0, 1]$ (Table II).

Exact verification uses great-circle distance, cosine similarity, and age; exponential freshness is only a monotone reparameterization:

$$\begin{aligned} d_i &= d_{\text{geo}}(p_i, \ell), & s_i &= q^\top v_i, \\ g_i &= (t_q - a_i)_+, & \phi_i(t_q) &= \exp(-\lambda g_i). \end{aligned} \quad (2)$$

The implementation uses $g_i \leq A_{\max}$, equivalent for $\lambda > 0$ to $\phi_i \geq \exp(-\lambda A_{\max})$, but never multiplies similarity by decay. Record age and source currency obey distinct contracts:

$$0 \leq t_q - W \leq \Delta_{\max}, \quad g_i \leq A_{\max}. \quad (3)$$

The first concerns completeness of the captured source prefix; the second filters a stored observation day. Neither proves recent physical inspection. Pilot observation/event days are deterministic experimental attributes, not historical OSM edits, so freshness claims apply only to the replay.

At watermark W , the range-complete answer is

$$A_W^*(Q) = \{i \in \mathcal{L}_W : d_i \leq r, s_i \geq \tau, t_q - a_i \leq A_{\max}\}. \quad (4)$$

All predicates are inclusive; subsequent ranking cannot alter membership.

B. Index and Current-View Semantics

FRESH-GEORANGE separates an immutable base from a mutable replacement delta. For the general event model, latest-write-wins semantics are

$$\begin{aligned} e_i^W &= \arg \max_{e \in B \cup D_{\leq W} : e.i = i} e.\nu, \\ \mathcal{L}_W &= \{e_i^W : e_i^W.o = \text{upsert}\}. \end{aligned} \quad (5)$$

The pilot is narrower: `apply_events` replays the complete prefix to a private one-value-per-ID dictionary, whose final sequence-ordered value is an upsert or tombstone. It atomically publishes that dictionary and W under a re-entrant lock; queries hold the lock throughout. Dictionary membership hides an old base copy. Building privately leaves readers on the preceding complete state, while publishing W before the dictionary would invalidate the proof. Thus calls are serializable, but there is no retained MVCC, lock-free reading, or durable recovery: $t_q < W$ and $t_q - W > \Delta_{\max}$ are rejected. Past-prefix republication is an evaluation operation, not a historical query. Inputs must have complete, deterministic sequences. The pilot generator produces a unique, gap-free sequence; malformed or conflicting external sequences are outside the implemented contract.

Figure 1 separates the current-view path from scripted drift feedback: the graph orders work; safe summaries and full delta scanning support correctness.

The day-30 base alone fits a deterministic 48-D word/character TF-IDF-SVD encoder. Normalized vectors are grouped by fixed 10° cells and approximately 64-record semantic blocks; dense record arrays remain the verification authority. A brute-built exact- k NN graph is traversed approximately only to propose base records—it is not HNSW or evidence of absence. Every current delta value is scanned, so unseen insertions remain discoverable. A replacement or tombstone hides its physical base slot; this auditable design trades delta latency for safety.

For cell C , let c_C^p be its center and let ρ_C^p cover the complete cell region under the declared great-circle metric.

TABLE II
NOTATION. TIME IS REPRESENTED BY INTEGER REPLAY DAY IN THE PILOT.

Symbol	Meaning	Symbol	Meaning
i	stable object identifier	ν_i	totally ordered event sequence
p_i, ℓ	object and query locations	v_i, q	unit object and query vectors
d	embedding dimension	r	inclusive range radius
τ	inclusive cosine threshold	t_q	query day
a_i	object observation day	$A_{\max}$	maximum object age
W	committed source watermark	$\Delta_{\max}$	maximum watermark lag
B, D_W	immutable base and current delta	$\mathcal{L}_W$	live view at W
C, b	spatial cell and semantic block	L_C^p	spatial lower bound
U_b^s, U_b^a	semantic and age upper bounds	$\mathcal{U}$	unexamined, possible blocks
$\hat{A}$	verified returned identifiers	M_{rem}	residual physical-slot count
ρ_{req}	requested certified recall	D_t	drift statistic

The implementation uses a meridian-then-parallel path with the parallel length evaluated at the cell latitude closest to the equator. The triangle inequality yields

$$\begin{aligned} L_C^p(\ell) &= \max\{0, d_{\text{geo}}(\ell, c_C^p) - \rho_C^p\}, \\ U_C^p(\ell) &= d_{\text{geo}}(\ell, c_C^p) + \rho_C^p. \end{aligned} \quad (6)$$

Hence $L_C^p > r$ safely eliminates a cell, including near poles or the date line.

For each base block b , store the arithmetic center c_b , radius $\rho_b = \max_{i \in b} \|v_i - c_b\|_2$, normalized center $\mu_b = c_b / \|c_b\|_2$ when the norm is nonzero (one member vector is used otherwise), and angular radius $\theta_b = \max_{i \in b} \arccos(\mu_b^\top v_i)$. Cauchy–Schwarz and the spherical triangle inequality give two independent safe upper bounds:

$$\begin{aligned} U_b^{\text{ball}}(q) &= \min\{1, q^\top c_b + \rho_b\}, \\ \eta_b(q) &= \max\{0, \arccos(\text{clip}(q^\top \mu_b, -1, 1)) - \theta_b\}, \\ U_b^{\text{ang}}(q) &= \cos(\eta_b(q)), \\ U_b^s(q) &= \min\{U_b^{\text{ball}}(q), U_b^{\text{ang}}(q)\}. \end{aligned} \quad (7)$$

Clip $\arccos$ arguments to $[-1, 1]$. Both components upper-bound every block dot product, so their minimum is safe. For $a_b^{\max} = \max_{i \in b} a_i$, age requires

$$U_b^a = a_b^{\max}, \quad U_b^a \geq t_q - A_{\max}. \quad (8)$$

A fixed outward margin protects bounds; verification uses float64. Pruning is only a disjunction of necessary-condition failures:

$$\begin{aligned} \text{prune}(b, Q) &\iff L_{C(b)}^p(\ell) > r \vee U_b^s(q) < \tau \\ &\vee U_b^a < t_q - A_{\max}. \end{aligned} \quad (9)$$

No learned score prunes. Immutable base blocks and a full delta scan avoid dynamic radius expansion in this pilot.

For any $i \in b$, $q^\top v_i = q^\top c_b + q^\top (v_i - c_b) \leq q^\top c_b + \rho_b$ because $\|q\|_2 = 1$. Spherical triangle inequality gives angle at least $\angle(q, \mu_b) - \theta_b$; cosine decreases on $[0, \pi]$. Metric triangle inequality proves (6). Therefore loose summaries admit false positives but cannot prune true members, assuming normalized nonzero vectors and the outward margin. Sampled bound audits check implementation maxima but do not replace this proof.

Algorithm 1 EXACT-RANGE(Q)

```

1: acquire the index lock and capture  $(B, D_W, W)$ 
Require:  $t_q \geq W$  and  $t_q - W \leq \Delta_{\max}$ 
2: encode and normalize  $q$ ; compute non-pruned base blocks  $\mathcal{S}$ 
3:  $\hat{A} \leftarrow \emptyset$ ;  $E \leftarrow \emptyset$ 
4: for each value in  $D_W$  do
5:   if it is an upsert and VERIFY passes then
6:     add its identifier to  $\hat{A}$ 
7:   end if
8: end for
9: for each graph-proposed base slot in  $\mathcal{S}$  do
10:  mark the slot in  $E$  and exactly verify it, including visibility
11: end for
12: for each base slot in  $\mathcal{S} \setminus E$  do
13:  mark and exactly verify the slot; add every passing identifier
14: end for
15: return sorted  $\hat{A}$ , certificate 1, and watermark  $W$ 

```

C. Exact and Certified Execution

Algorithm 1 describes exact mode as implemented. Graph traversal only changes the order of some checks; completion scans every unexamined slot in every non-pruned base block. Each candidate is checked for delta shadowing, hard age, exact dot product, and exact great-circle distance.

Certified mode scans all deltas, takes graph proposals, then visits surviving blocks by decreasing semantic upper bound. For examined base slots E and non-pruned blocks with unexamined slots $\mathcal{U}$, define

$$M_{\text{rem}} = \sum_{b \in \mathcal{U}} |\{j \in b : j \notin E\}|. \quad (10)$$

This physical count includes shadowed and predicate-failing records; such overcounting only weakens the bound. For exactly verified IDs $\hat{A}$,

$$\begin{aligned} \underline{R} &= \begin{cases} 1, & |\hat{A}| = 0 \wedge M_{\text{rem}} = 0, \\ |\hat{A}| / (|\hat{A}| + M_{\text{rem}}), & \text{otherwise,} \end{cases} \\ &\text{stop if } M_{\text{rem}} = 0 \vee \underline{R} \geq \rho_{\text{req}}. \end{aligned} \quad (11)$$

The certificate assumes the locked snapshot, safe bounds, exact verification, and normalized vectors, but no graph recall. For missed answers $\mathcal{M} = A_W^*(Q) \setminus \hat{A}$, all deltas were checked

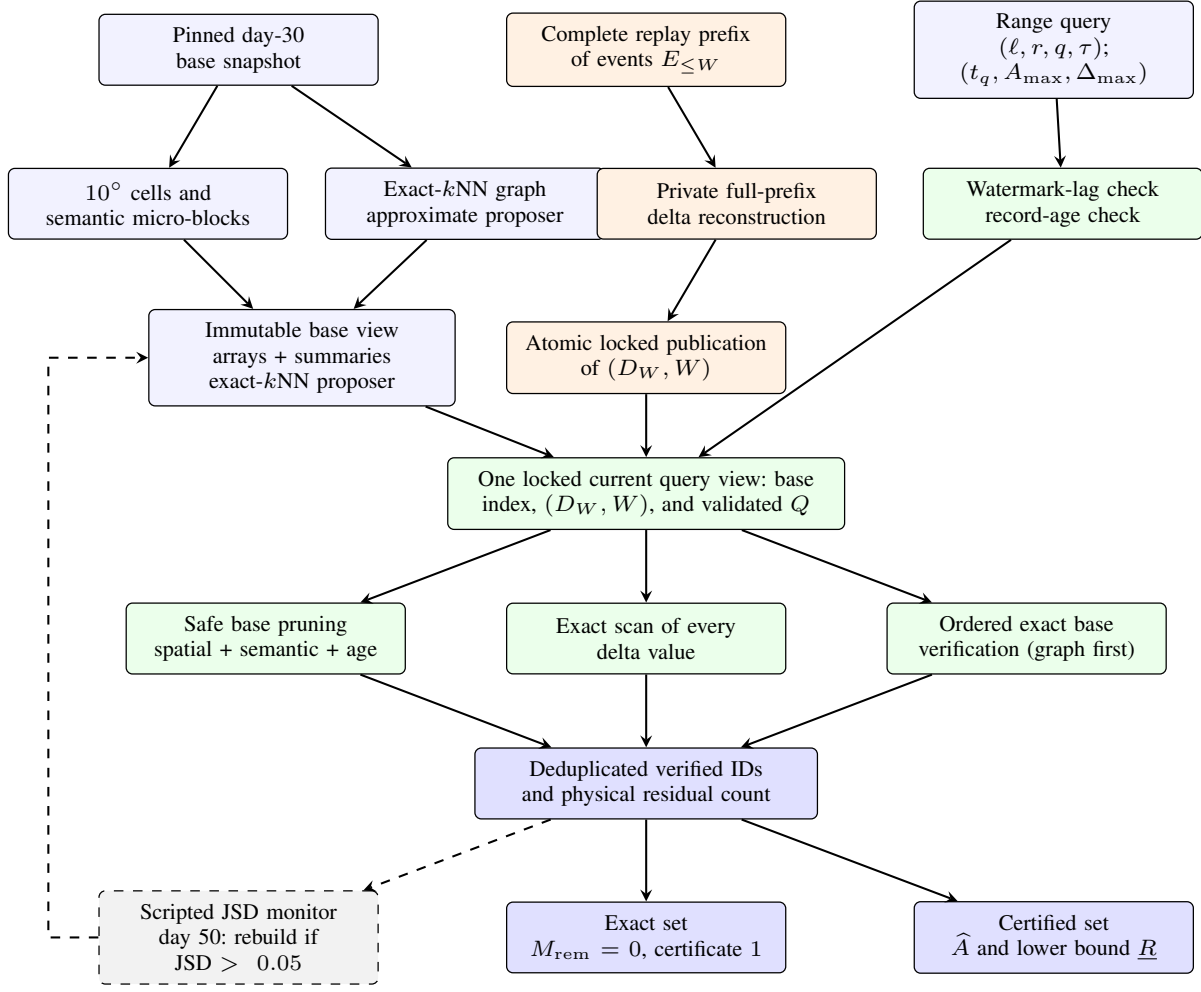


Fig. 1. FRESH-GEORANGE pilot architecture. Solid arrows are implemented data or query paths. The dashed path is the scripted drift experiment: it performs a full rebuild at a fixed checkpoint, not continuous retraining, partial repartitioning, or the benefit-aware policy in Eq. (14).

Algorithm 2 CERTIFIED-RANGE($Q, \rho_{\text{req}}, b, e_f$)

- 1: capture and validate the locked current view
- 2: prune base blocks safely; exactly scan every current delta value
- 3: request at most e_f graph proposals; exactly verify allowed slots
- 4: compute M_{rem} and $\underline{R}$
- 5: **while** $\underline{R} < \rho_{\text{req}}$ and fewer than $b + e_f$ base slots are examined **do**
- 6: choose the remaining block with largest $U_b^s(q)$
- 7: scan its unexamined slots until the block or cap ends
- 8: update $\hat{A}$, M_{rem} , and $\underline{R}$
- 9: **end while**
- 10: **return** sorted $\hat{A}$, $\underline{R}$, M_{rem} , and W

and Eq. (9) excludes pruned blocks; each miss therefore maps to an unexamined physical slot in (10):

$$|\mathcal{M}| \leq M_{\text{rem}} \implies \text{Recall}(\hat{A}) = \frac{|\hat{A}|}{|\hat{A}| + |\mathcal{M}|} \geq \underline{R}. \quad (12)$$

Empty truth has recall one; the bound reaches one only when the residual is zero, as exact mode ensures. An unexamined shadowed slot conservatively remains in M_{rem} ; examining it removes the slot without adding an answer. Delta answers may

enter the numerator because they were verified, and never enter the residual because all deltas were scanned. ID deduplication prevents double counting. Thus the certificate is a lower bound, not empirical-recall estimation; tightness reflects how quickly pruning/scanning removes slots.

D. Drift, Maintenance, and Cost Policy

For a full deployment, fixed, smoothed histograms can track spatial occupancy, semantic-cluster occupancy, query location, radius, and threshold. With reference distribution $P_{j,0}$, recent distribution $P_{j,t}$, and $M_j = (P_{j,0} + P_{j,t})/2$, define

$$D_t = \sum_j \alpha_j \text{JSD}(P_{j,t}, P_{j,0}), \quad \sum_j \alpha_j = 1, \quad (13)$$

$$\text{JSD}(P_{j,t}, P_{j,0}) = \frac{1}{2} \text{KL}(P_{j,t} \| M_j) + \frac{1}{2} \text{KL}(P_{j,0} \| M_j).$$

A defensible production decision would require persistent drift and amortized benefit:

$$\text{REBUILD}_t = \mathbf{1} \left[\min_{0 \leq h < k} D_{t-h} > \gamma_D \right] \cdot \mathbf{1} \left[H \widehat{\Delta L}_{p99} > C_{\text{rebuild}} \right]. \quad (14)$$

Here H is forecast query count and $\widehat{\Delta L}_{p99}$ conservative per-query benefit in rebuild-cost units; compaction is separate. The pilot does *not* implement Eq. (14): it uses a smoothed 24-bin query-location histogram and one day-50 rebuild after a scripted shift if $JSD > 0.05$. This checks mechanism, not online optimality or partial rebuilds.

More generally, a configuration π could be selected with normalized costs and hard guarantees:

$$\begin{aligned} \min_{\pi} \quad & w_L \widehat{L}_{p99} + w_U \widehat{C}_{\text{update}} + w_M \widehat{M} \\ & + w_S(\widehat{t_q - W}) + w_B \widehat{C}_{\text{rebuild}} \\ \text{s.t.} \quad & \underline{R}_{\pi} \geq \rho_{\text{req}}, \quad t_q - W \leq \Delta_{\text{max}}, \quad M_{\pi} \leq M_{\text{max}}. \end{aligned} \quad (15)$$

Hats denote division by preregistered reference scales. Equation (15) uses M_{π} for memory footprint (not M_{rem}) and is a design objective, not an optimizer implemented or tuned in this pilot.

E. Complexity, Limits, and Edge Cases

For m surviving blocks, n_{scan} examined base slots, delta size $|D_W|$, and graph degree g , query work is $O(md + n_{\text{scan}}d + |D_W|d)$ plus at most e_f vertices' graph work; worst-case exact search is linear. Storage is $O((N + |D_W|)d + Ng)$, including duplicated graph vectors. Brute-force neighbor construction can be $O(N^2d)$: no HNSW or sublinear guarantee is claimed. Full-prefix replay and embedding makes a pilot publication $O(|E_{\leq W}|d)$, not a production append-only $O(d)$; reported update throughput is implementation-specific.

Boundary equality is resolved only by the exact inclusive predicates. A zero radius returns co-located matches; $\tau = -1$ disables semantic exclusion, while $\tau = 1$ is checked in float64. Great-circle distance and wrapped longitude handle date-line and polar queries. Multiple events for an identifier resolve by final sequence order; a tombstone hides all base state. An exact empty answer has certificate one, whereas an early empty answer with residual slots has certificate zero. Unknown-region delta objects remain safe because the delta is not spatially pruned. Missing trustworthy observation time should disable the object-age predicate rather than substitute last-edit time. The method assumes finite embeddings in one fixed model generation; the current prototype normalizes nonzero outputs, treats a zero vector as having zero cosine with every query, and does not implement dual-generation migration. These restrictions delimit the claims evaluated in Section IV.

IV. EXPERIMENTAL EVALUATION

This section reports a functional CPU pilot, not a publication-scale comparison. The narrow purpose is to check the query semantics, exact mode, certificate accounting, and update-overlay plumbing on an executable workload. It does not substitute for the proposed OSM-diff evaluation, and the IVF, graph, lexical, and LIST-router implementations are local reference proxies rather than the authors' systems. Results are therefore interpreted within this pilot only.

TABLE III
PILOT DATA SNAPSHOTS. OBJECT ATTRIBUTES ARE REAL; REPLAY OPERATIONS AND TIMES ARE SIMULATED.

Snapshot	Objects	Countries	Events
Pinned raw sample	2500	155	0
Base	2000	132	0
Live after replay	2420	154	740

TABLE IV
RESOLVED PILOT CONFIGURATION. VALUES ARE EXPORTED FROM THE YAML FILES AND EXECUTION MANIFEST.

Setting	Resolved value
Raw/base/live objects	2,500 / 2,000 / 2,420
Simulated replay events	740
Primary seeds	101, 211, 307, 401, 503
Queries per seed / total	36 / 180
Radii (km)	150, 500, 1500, 4000
Cosine thresholds	0.15, 0.25, 0.35, 0.45
Maximum simulated ages (days)	14, 30, 60, 60
Encoder / dimensions	TF-IDF+LSA / 48
Cell / block size	10.0 deg / 64 records
Graph degree / proposal budget	12 / 64
Certificate target	0.95
CPU / logical CPUs	AMD EPYC 9V74 80-Core Processor / 9
Python / scikit-learn	3.12.14 / 1.8.0

A. Experimental Setup

a) *Data and replay.*: We pinned a 2,500-row airport sample from the OpenFlights-format airport file [32]; every retained row declares OurAirports, whose official page releases the data to the Public Domain [33], as its source (SHA-256 0ad06d...44cd). Its names, cities, countries, facility fields, and coordinates are source attributes. We used the first 2,000 rows as the day-30 base and held out 500 real rows for insertions on days 31–60. A seeded generator added 160 metadata updates and 80 deletions, yielding 740 replay events and 2,420 live objects on day 60. Event types, event days, and `observed_day` are simulated, so this experiment tests a simulated age predicate and source-watermark behavior; it provides no evidence about historical airport changes or real record-freshness distributions. Table III records the snapshots.

Table IV records the resolved, executable settings rather than an intended configuration. All later plot data and prose macros are exported from the matching run.

b) *Representation and index settings.*: A deterministic word-and-character TF-IDF pipeline followed by 48-dimensional truncated SVD produced unit-normalized vectors. It was fit on the day-30 base only; insertion and update text was unseen during fitting. FRESH-GEORANGE used 10° spatial cells, semantic blocks of at most 64 records, a degree-12 exact- k NN reference graph used only for proposal order, and search budget 64. Certified queries requested a 0.95 lower bound; exact queries continued verification to completion. The query driver applied the complete event history through a watermark before querying. This replacement-batch API assumes one writer and a complete history; it is not a retained, concurrent micro-batch implementation.

c) *Queries, comparators, and ground truth.*: For each of five seeds (101, 211, 307, 401, and 503), we sampled 36 day-60 queries from the live corpus, for 180 distinct queries.

A query center is a small perturbation of a sampled airport and its phrase is formed from that row’s derived facility kind, city, and country. Radii rotate over 150, 500, 1,500, and 4,000 km; cosine thresholds over 0.15, 0.25, 0.35, and 0.45; and maximum simulated ages over 14, 30, 60, and 60 days. These broad radii and source-derived phrases make this a predicate smoke test, not an independently judged emergency-resource workload.

Brute-force, spatial-first, and semantic-first exact scans establish set-valued ground truth. We additionally tested a lexical spatial grid, IVF post-filter, an exact- k NN graph post-filter, a LIST-style cluster router, and a spatial index rebuilt every 15 days, plus FGR-Cert95 and FGR-Exact. The local IVF, graph, lexical, and LIST-style implementations neither reproduce HNSW, LIST, nor an IR-tree nor justify external superiority claims. Ten methods over 180 queries produced 1800 method–query executions.

d) Measurement and statistics. For truth T and output A , recall, precision, and Jaccard are $|A \cap T|/|T|$, $|A \cap T|/|A|$, and $|A \cap T|/|A \cup T|$ (empty–empty is one). Higher set scores and throughput are better; lower latency, staleness, footprint, and rebuild cost are better. Method order was randomized within each measured query, but every query–method pair was timed only once. Latency is therefore exploratory single-query wall-clock time. The driver issued one query at a time on a CPU environment exposing 9 logical CPUs: AMD EPYC 9V74 80-Core Processor, Linux 6.18.35, Python 3.12.14, NumPy 2.3.5, pandas 2.2.3, scikit-learn 1.8.0, and SciPy 1.17.0. Peak process RSS was 283,452 KiB. Footprint is protocol-5 serialized object size, not resident memory. “Examined” is meaningful within a method but is not comparable across methods because the implementations count different operations.

For inferential summaries, the five independently constructed seeds—not the 180 dependent query rows—are the paired units. We applied two-sided Wilcoxon signed-rank tests and Holm adjustment within each metric; intervals in Table VI bootstrap the five paired seed differences. The pooled query-row recall intervals in the main summary are not used for inference. With $n = 5$, the tests are coarse and underpowered and are included as an uncertainty check rather than evidence of equivalence or superiority.

B. Primary Result: Correctness at a Latency Cost

Table V gives the principal result. FGR-Exact attained 100.00% recall, precision, and Jaccard on all 180 primary queries. FGR-Cert95 attained 99.91% mean empirical recall and 100.00% precision while reporting a 99.41% mean lower bound. The mean empirical-minus-certified gap was only 0.50 pp; the certificate-violation count was 0 of 180, the minimum certificate was 95.26%, and the below-target count was 0. Across five seeds, a separate audit made 6,080 sampled semantic-bound comparisons; the smallest slack was positive (3.56×10^{-12}). Ten unit tests also covered exact/certificate behavior, normalization, terminal cells, non-divisor grids, poles, the dateline, singleton graphs, and watermark rules. These checks support the certificate on the declared path—normalized vectors, a committed snapshot, complete-history

replay, and one ordered writer—but are not a general production proof. Repeated versions, moves, late or gapped events, concurrent writers, malformed embeddings, and semantic outlier streams remain outside the property suite.

The efficiency result is negative. FGR-Cert95 required a 7.24-ms median and 12.39-ms p95, versus 1.24 ms and 1.99 ms for spatial-first exact—a $5.85\times$ median slowdown ($6.22\times$ at p95). It was also essentially tied with FGR-Exact (7.27 ms), saving only 1.1 reported examinations on average. On this small corpus, certification and overlay bookkeeping dominate, and early termination provides no meaningful latency win. Serialized footprint was 5.39 MB versus 5.18 MB for spatial-first. The correct conclusion is feasibility of the narrow correctness contract, not a performance advantage.

Figure 2 reaches the same conclusion across budgets. Increasing the FGR request from 0 to 0.95 raised empirical recall from 77.88% to 100.00% and its reported bound from 55.63% to 99.71%; median latency remained between 5.56 and 6.85 ms because the corpus is tiny and timing noise masks a monotone work trend. The IVF proxy reaches 100.00% only when all 32 clusters are probed, whereas the graph proxy reaches just 37.46% at its largest budget. The plotted proxy curves measure these local implementations only and provide no comparison to official ANN systems.

Table VI reinforces the limits on comparative interpretation. The estimated FGR-Cert95 recall differences were $+0.143$ versus the LIST-router proxy, $+0.145$ versus IVF, and $+0.673$ versus the graph proxy, but every Holm-adjusted p -value was at least 0.5625. Its difference from FGR-Exact was -0.00092 . Five seed pairs cannot sustain a strong performance claim.

C. Sensitivity, Freshness, and Replay Cost

Radius and threshold sweeps in Fig. 3 reuse a fixed cohort within each panel. At radii through 1,000 km, both recall and certificate were 100.00% and 100.00%, respectively; at 5,000 km, latency rose to 8.45 ms, recall was 99.90%, and the bound was 99.00%. Tightening the cosine threshold from 0.05 to 0.60 reduced mean answer count from 148.3 to 29.6 and latency from 10.29 to 5.44 ms; recall rose from 99.83% to 100.00%, while the bound rose from 98.97% to 99.81%. This is consistent with fewer qualifying records, but only for this cohort and encoder.

Figure 4(a) times full-history replacement, not sustained incremental ingestion. Replaying 50, 100, 250, 500, and 740 events took 55.4, 108.2, 285.0, 576.5, and 721.6 ms, respectively, for a median 903 and maximum 1025 replayed events/s. A full re-embedding and rebuild took 384.6–509.8 ms. The last evaluated batch where replay was cheaper contained 250 events, and the first where it was slower contained 500; the present delta representation therefore does not demonstrate update scalability. A credible dynamic benchmark still needs retained micro-batches, operation-stratified latency, compaction, and concurrent query interference.

Panel (b) isolates watermark staleness on one 24-query cohort. FGR-Exact, given the complete history through each query day, retained at least 100.00% recall. Immediately before each scheduled rebuild, the periodic view’s recall fell from

TABLE V
EXECUTED DAY-60 PILOT RESULTS ACROSS FIVE SEEDS AND 180 QUERIES PER METHOD. RECALL AND PRECISION ARE PERCENTAGES. FOOTPRINT IS STANDALONE PICKLE SIZE, NOT RESIDENT MEMORY.

Method	Recall $\uparrow$	Precision $\uparrow$	Median ms $\downarrow$	p95 ms $\downarrow$	Examined	Serialized MB
Brute-force exact	100.00	100.00	1.39	2.03	2420.0	5.07
Spatial-first exact	100.00	100.00	1.24	1.99	265.4	5.18
Semantic-first exact	100.00	100.00	1.23	1.86	448.5	5.07
Lexical spatial	99.96	100.00	1.44	2.76	265.4	5.40
IVF-post	85.45	97.78	1.29	1.84	592.9	5.12
kNN-graph post-filter	32.56	55.56	2.31	3.19	96.0	5.18
LIST-router proxy	85.59	98.33	1.27	1.79	504.8	5.11
Periodic rebuild (15 d)	68.01	78.20	1.28	2.14	257.9	5.04
FGR-Cert95 (this work)	99.91	100.00	7.24	12.39	1047.4	5.39
FGR-Exact (this work)	100.00	100.00	7.27	12.53	1048.5	5.39

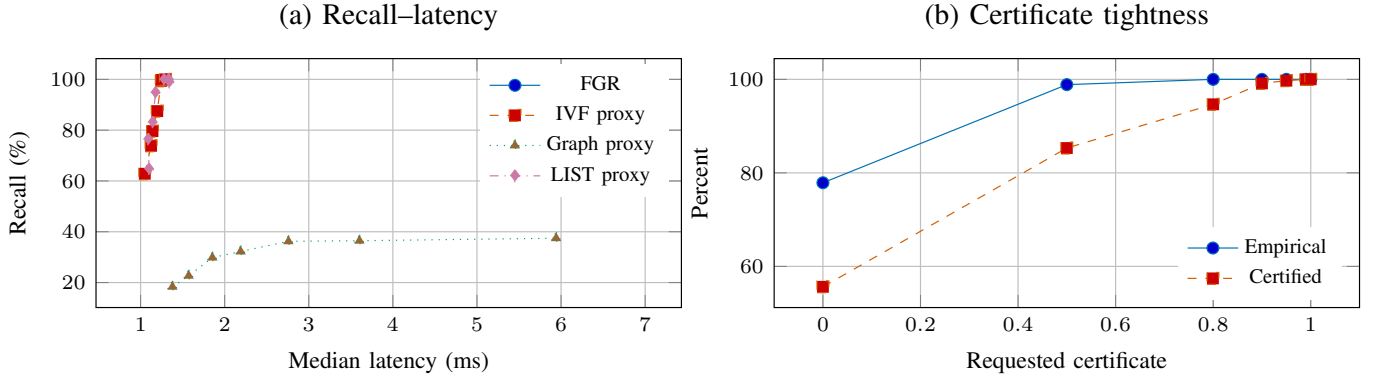


Fig. 2. Recall-latency and certificate frontier. This is a single-query CPU pilot over 2,500 airport rows with a simulated 740-event replay; IVF and graph lines are local proxies, not official implementations.

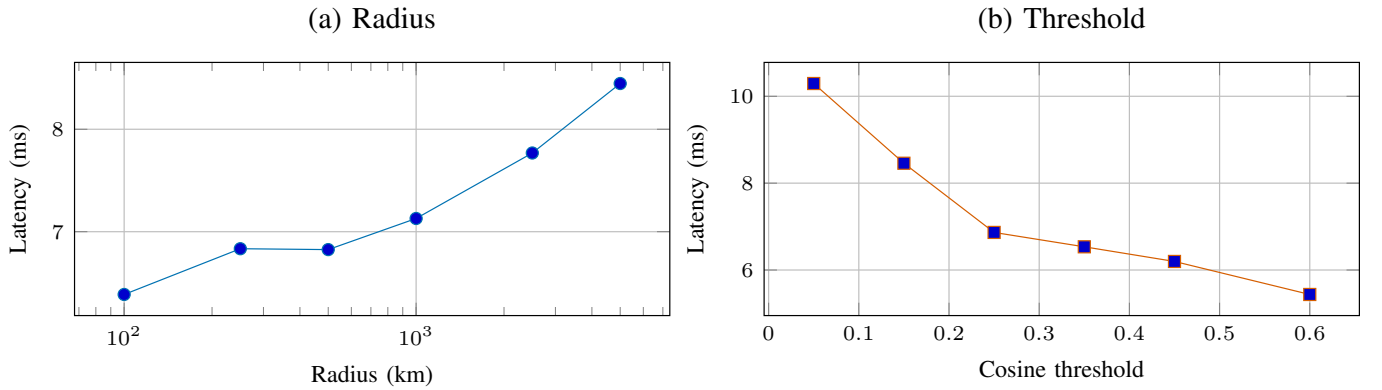


Fig. 3. Radius and threshold sensitivity in the single-query, 2,500-row airport pilot with simulated updates. Points are aggregates over fixed 20-query cohorts; they are not population-level estimates.

98.95% at five days of lag to 95.52% at ten days and 81.40% at 15 days. After the next rebuild it recovered to 95.81%, then declined to 77.89% recall and 82.44% precision at day 60. This supports the elementary point that an old watermark changes the answer set; because event time and record age are simulated, it does not quantify real-world source freshness.

D. Diagnostic Stress Tests and Ablations

Figure 5(a) is deliberately a scripted trigger test. A forced query-region shift raises Jensen-Shannon divergence from

0.09 on day 40 to 0.55 on day 45; a hard-coded day-50 rebuild costs 298.3 ms. The measured p95 then changes from 7.43 ms (day 45) to 5.85 ms (day 50), while the mean certificate changes from 99.43% to 98.09%. Because the workload changes and there is no simultaneous no-rebuild counterfactual, this cannot be attributed causally to recovery or used to validate an adaptive policy. Panel (b) shows serialized FGR size increasing from 3.88 MB at 500 objects to 5.01 MB at 2,000, compared with 3.82 to 4.88 MB for spatial-first. Four small points describe the pilot footprint; they do not establish asymptotic scaling.

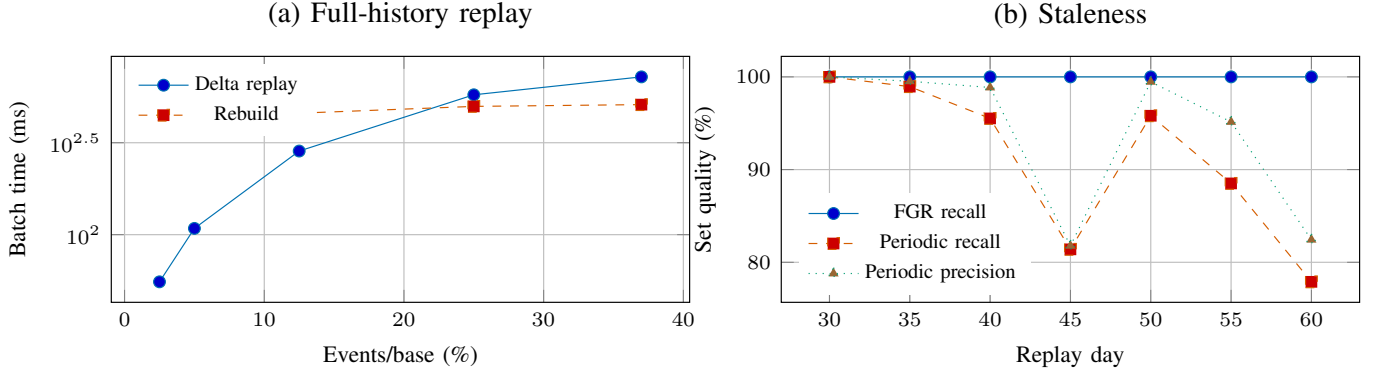


Fig. 4. Full-history replay cost and 15-day rebuild staleness. Both panels use the 2,500-row CPU pilot and simulated event stream; replay is replacement batch application, not concurrent incremental throughput.

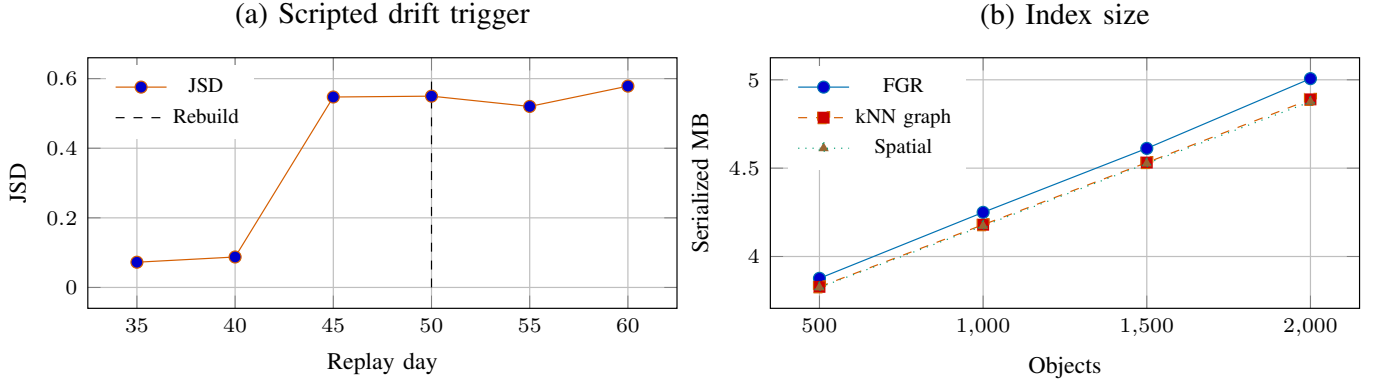


Fig. 5. Scripted drift instrumentation and serialized footprint in the 2,500-row CPU pilot with simulated replay. The trigger is hard-coded and the graph line is a local exact- k NN proxy.

TABLE VI
PAIRED RECALL DIFFERENCES: FGR-CERT95 MINUS COMPARATOR. CIS ARE BOOTSTRAP 95%; p_H IS HOLM-ADJUSTED WILCOXON.

Comparator	Mean diff.	95% CI	p_H
LIST-router proxy	0.143	[0.105, 0.194]	0.56
FGR-Exact	-0.001	[-0.001, -0.000]	0.62
Periodic rebuild (15 d)	0.319	[0.280, 0.352]	0.56
Spatial-first exact	-0.001	[-0.001, -0.000]	0.62
IVF-post	0.145	[0.117, 0.174]	0.56
kNN-graph post-filter	0.673	[0.631, 0.709]	0.56

Selectivity in Fig. 6(a) exposes the same overhead: FGR-Cert95 median latency rises from 5.93 ms for empty answers to 9.23 ms above 2% selectivity, while spatial-first remains between 1.15 and 1.39 ms. FGR recall is at least 100.00% through the 0.1–0.5% bin and 99.67% above 2%. We do not compare “examined” counts because their definitions differ by method. Panel (b) is a dictionary-normalization sanity check, not a multilingual retrieval evaluation. The fixed lexicon raises category recall from 56.82% to 61.36% for French, 56.53% to 61.36% for Spanish, 57.10% to 60.51% for German, and 55.97% to 60.51% for an English synonym; English remains 60.51%. Labels are circularly inferred from the word “international,” and tested translations appear in the lexicon, so no language-generalization claim follows.

Finally, Table VII and Fig. 7 show that removing the

TABLE VII
COMPONENT ABLATION. THE NO-DELTA ROW IS DELIBERATELY FRESHNESS-NONCOMPLIANT, SO NO CERTIFICATE IS REPORTED.

Variant	Recall (%)	Cert. (%)	Latency (ms)
Full	99.68	99.20	8.17
No graph proposer	99.68	99.20	6.93
No semantic bounds	99.88	99.20	7.93
One semantic block per cell	99.68	99.20	8.18
Stale base, no delta (uncertified)	59.38	—	4.73

graph proposer preserves recall (99.68% versus 99.68%) and the mean certificate (99.20% versus 99.20%) while changing latency from 8.17 to 6.93 ms; the no-graph variant is faster in this run. Removing semantic bounds or merging to one semantic block per spatial cell changes latency to 7.93 and 8.18 ms. These small, unreplicated ablations provide no positive component-level speed evidence and cannot establish behavior at scale. The decisive row is the deliberately unsafe stale-base variant: omitting the delta lowers recall from 99.68% to 59.38%. Its certificate is suppressed because advancing the watermark without applying the events violates the certificate premise.

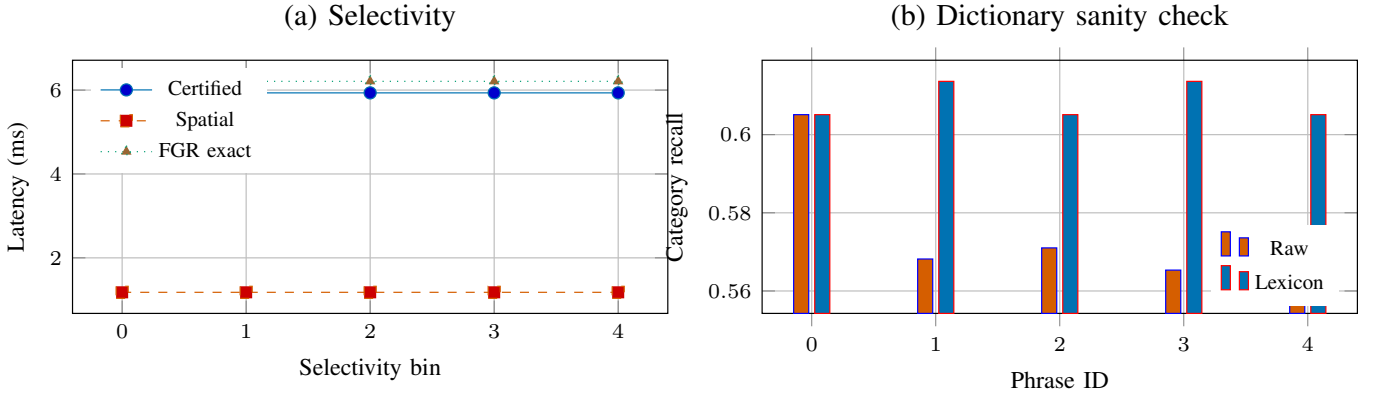


Fig. 6. Selectivity and dictionary-normalization diagnostics on the single-query airport pilot with simulated replay. The lexical intervention is constructed and does not measure multilingual robustness.

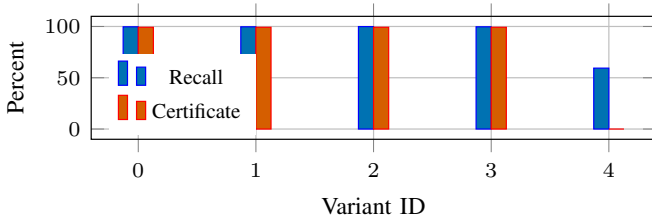


Fig. 7. Guarantee ablation; variant IDs and latency are in Table VII. The stale-base variant is uncertified.

E. Threats to Validity and Deployment Limits

Local proxies, single timings, and Python code limit internal validity; source-derived queries, simulated freshness, and serialized size limit construct validity. One uneven 2,500-airport sample, five seeds, and one CPU limit external and statistical validity. Single-writer replay omits concurrency. No personal data are used, but the pilot is not safety-ready. Pinned artifacts aid reproduction. Archived OSM diffs [34], official baselines, independent queries, concurrent updates, and broader property tests remain required. Correctness passes, but efficiency fails: FGR-Cert95 is $5.85\times$ slower than spatial-first exact. Publication claims are no-go.

V. CONCLUSION

FRESH-GEORANGE addresses semantic-spatial range enumeration when applications need every qualifying object from a current snapshot rather than an approximate ranking. The design combines geographic cells, semantic microblocks, admissible spatial and cosine bounds, a graph used for proposal ordering, and a latest-write delta overlay. Exact mode accounts for every nonprunable record; certified mode converts verified answers and an unresolved-record count into a deterministic recall lower bound. In the CPU pilot, exact mode returned the complete reference set on all 180 queries, while the 95-percent mode reached 99.91% empirical mean recall with a 99.41% reported mean certificate and no certificate violation. These correctness results are encouraging, but the efficiency evidence is negative: certified retrieval required 7.24 ms median latency, 5.85 times the spatial-first exact baseline, and complete-history

replay exceeded rebuild time for large batches. Consequently, this package demonstrates a functional proof of concept, not a submission-ready dynamic index or a deployment claim. Its limitation is scope: 2,500 real airport records were paired with simulated update histories and local proxy implementations rather than real OpenStreetMap diffs and official systems. Future work should implement incremental multi-version publication with sequence validation, then test million-scale diffs, concurrent workloads, independently judged semantic queries, and recall-matched official baselines.

REFERENCES

- [1] Z. Yin, S. Feng, S. Liu, G. Cong, Y. S. Ong, and B. Cui, "LIST: Learning to index spatio-textual data for embedding based spatial keyword queries," *The VLDB Journal*, vol. 34, 2025. [Online]. Available: <https://doi.org/10.1007/s00778-024-00886-5>
- [2] Y. Song, B. Yao, Z. Chen, X. Yang, J. Xie, F. Li, and M. Chen, "Efficient top-k spatial-range-constrained approximate nearest neighbor search on geo-tagged high-dimensional vectors," *The VLDB Journal*, vol. 34, 2025. [Online]. Available: <https://doi.org/10.1007/s00778-024-00894-5>
- [3] T. Lee, J. woo Park, S. Lee, S. won Hwang, S. Elnikety, and Y. He, "Processing and optimizing main memory spatial-keyword queries," *Proceedings of the VLDB Endowment*, vol. 9, no. 3, pp. 132–143, 2015. [Online]. Available: <https://www.vldb.org/pvldb/vol9/p132-lee.pdf>
- [4] Y. Sheng, X. Cao, Y. Fang, K. Zhao, J. Qi, G. Cong, and W. Zhang, "WISK: A workload-aware learned index for spatial keyword queries," *Proceedings of the ACM on Management of Data*, vol. 1, no. 2, 2023. [Online]. Available: <https://doi.org/10.1145/3589332>
- [5] P. Tampakakis, D. Spyrellis, C. Doulkeridis, N. Pelekis, C. Kalyvas, and A. Vlachou, "A novel indexing method for spatial-keyword range queries," in *Proceedings of the 17th International Symposium on Spatial and Temporal Databases (SSTD '21)*. Association for Computing Machinery, 2021, pp. 54–63. [Online]. Available: <https://doi.org/10.1145/3469830.3470897>
- [6] S. Oh, H. Jung, J. Koo, and U.-M. Kim, "Efficient method for processing range spatial keyword queries over moving objects based on Word2Vec," in *Human Interface and the Management of Information (HIMI 2018)*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2018, vol. 10905, pp. 620–639. [Online]. Available: https://doi.org/10.1007/978-3-319-92046-7_51
- [7] S. Alkobaisi, "Optimizing map labeling of point features based on an onion peeling approach," *Journal of Spatial Information Science*, 2011.
- [8] W. D. Bae, S. Alkobaisi, M. Horak, C.-S. Park, S. Kim, and J. Davidson, "Predicting health risks of adult asthmatics susceptible to indoor air quality using improved logistic and quantile regression models," *Life*, vol. 12, no. 10, p. 1631, 2022.
- [9] A. R. Mahmood, A. M. Aly, and W. G. Aref, "FAST: Frequency-aware indexing for spatio-textual data streams," in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE,

- 2018, pp. 305–316, preprint title: FAST: Frequency-Aware Spatio-Textual Indexing for In-Memory Continuous Filter Query Processing; arXiv:1709.02529. [Online]. Available: <https://doi.org/10.1109/ICDE.2018.00036>
- [10] Y. Dong, C. Xiao, H. Chen, J. X. Yu, K. Takeoka, M. Oyamada, and H. Kitagawa, “Continuous top-k spatial-keyword search on dynamic objects,” *The VLDB Journal*, vol. 30, no. 2, pp. 141–161, 2021. [Online]. Available: <https://doi.org/10.1007/s00778-020-00627-4>
- [11] C. Salgado, M. A. Cheema, and M. E. Ali, “Continuous monitoring of range spatial keyword query over moving objects,” *World Wide Web*, vol. 21, pp. 687–712, 2018. [Online]. Available: <https://doi.org/10.1007/s11280-017-0488-3>
- [12] Y. Xu, H. Liang, J. Li, S. Xu, Q. Chen, Q. Zhang, C. Li, Z. Yang, F. Yang, Y. Yang, P. Cheng, and M. Yang, “SPFresh: Incremental in-place update for billion-scale vector search,” in *Proceedings of the 29th ACM Symposium on Operating Systems Principles (SOSP ’23)*. Association for Computing Machinery, 2023. [Online]. Available: <https://doi.org/10.1145/3600006.3613166>
- [13] S. Gong, H. Sun, Z. Fang, L. Liu, L. Chen, and Y. Gao, “VStream: A distributed streaming vector search system,” *Proceedings of the VLDB Endowment*, vol. 18, no. 6, pp. 1593–1606, 2025. [Online]. Available: <https://www.vldb.org/pvldb/vol18/p1593-gao.pdf>
- [14] J. Mohoney, D. Sarda, M. Tang, S. R. Chowdhury, A. Pacaci, I. F. Ilyas, T. Rekatsinas, and S. Venkataraman, “Quake: Adaptive indexing for vector search,” in *19th USENIX Symposium on Operating Systems Design and Implementation (OSDI 25)*. Boston, MA: USENIX Association, Jul. 2025, pp. 153–169. [Online]. Available: <https://www.usenix.org/conference/osdi25/presentation/mohoney>
- [15] W. D. Bae, S. Alkobaisi, S. H. Kim, S. Narayanappa, and C. Shahabi, “Supporting range queries on web data using k-nearest neighbor search,” in *International Symposium on Web and Wireless Geographical Information Systems*. Springer, 2007, pp. 61–75.
- [16] C. Zuo, M. Qiao, W. Zhou, F. Li, and D. Deng, “SeRF: Segment graph for range-filtering approximate nearest neighbor search,” *Proceedings of the ACM on Management of Data*, vol. 2, no. 1, 2024. [Online]. Available: <https://doi.org/10.1145/3639324>
- [17] Y. Xu, J. Gao, Y. Gou, C. Long, and C. S. Jensen, “iRangeGraph: Improvising range-dedicated graphs for range-filtering nearest neighbor search,” *Proceedings of the ACM on Management of Data*, vol. 2, no. 6, 2024. [Online]. Available: <https://doi.org/10.1145/3698814>
- [18] F. Ahmed, M. Zia, H. Mahmood, and S. Al Kobaisi, “Open source computer game application: An empirical analysis of quality concerns,” *Entertainment Computing*, vol. 21, pp. 1–10, 2017.
- [19] Z. Peng, M. Qiao, W. Zhou, F. Li, and D. Deng, “Dynamic range-filtering approximate nearest neighbor search,” *Proceedings of the VLDB Endowment*, vol. 18, no. 10, pp. 3256–3268, 2025. [Online]. Available: <https://www.vldb.org/pvldb/vol18/p3256-deng.pdf>
- [20] M. Jiang, Z. Yang, F. Zhang, G. Hou, J. Shi, W. Zhou, F. Li, and S. Wang, “DIGRA: A dynamic graph indexing for approximate nearest neighbor search with range filter,” *Proceedings of the ACM on Management of Data*, vol. 3, no. 3, 2025. [Online]. Available: <https://doi.org/10.1145/3725399>
- [21] F. Zhang, M. Jiang, G. Hou, J. Shi, H. Fan, W. Zhou, F. Li, and S. Wang, “Efficient dynamic indexing for range filtered approximate nearest neighbor search,” *Proceedings of the ACM on Management of Data*, vol. 3, no. 3, 2025, introduces RangePQ and RangePQ+. [Online]. Available: <https://doi.org/10.1145/3725401>
- [22] Z. Wang, J. Zhang, and W. Hu, “WoW: A window-to-window incremental index for range-filtering approximate nearest neighbor search,” *Proceedings of the ACM on Management of Data*, vol. 3, no. 6, 2025, published online December 2025; presented at SIGMOD 2026. [Online]. Available: <https://doi.org/10.1145/3769843>
- [23] M. D. Manohar, T. Kim, and G. E. Bluelloch, “Range retrieval with graph-based indices,” 2025, active preprint, version 3. [Online]. Available: <https://arxiv.org/abs/2502.13245v3>
- [24] S. Horchidan, F. Zeiher, H. Boström, and P. Carbone, “ConANN: Conformal approximate nearest neighbor search,” *Proceedings of the VLDB Endowment*, vol. 19, no. 1, pp. 29–42, 2025. [Online]. Available: <https://www.vldb.org/pvldb/vol19/p29-horchidan.pdf>
- [25] Z. Li, K. C. K. Lee, B. Zheng, W.-C. Lee, D. L. Lee, and X. Wang, “IR-Tree: An efficient index for geographic document search,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 4, pp. 585–599, 2011. [Online]. Available: <https://doi.org/10.1109/TKDE.2010.149>
- [26] X. Chen, J. Xu, R. Zhou, P. Zhao, C. Liu, J. Fang, and L. Zhao, “S²R-tree: A pivot-based indexing structure for semantic-aware spatial keyword search,” *GeoInformatica*, vol. 24, pp. 3–25, 2020. [Online]. Available: <https://doi.org/10.1007/s10707-019-00372-z>
- [27] S. Liu, G. Cong, K. Feng, W. Gu, and F. Zhang, “Effectiveness perspectives and a deep relevance model for spatial keyword queries,” *Proceedings of the ACM on Management of Data*, vol. 1, no. 1, pp. 1–25, 2023. [Online]. Available: <https://doi.org/10.1145/3588691>
- [28] S. Gollapudi, N. Karia, V. Sivashankar, R. Krishnaswamy, N. Begwani, S. Raz, Y. Lin, Y. Zhang, N. Mahapatro, P. Srinivasan, A. Singh, and H. V. Simhadri, “Filtered-DiskANN: Graph algorithms for approximate nearest neighbor search with filters,” in *Proceedings of the ACM Web Conference 2023*. Association for Computing Machinery, 2023, pp. 3406–3416. [Online]. Available: <https://doi.org/10.1145/3543507.3583552>
- [29] A. Liang, P. Zhang, B. Yao, Z. Chen, Y. Song, and G. Cheng, “UNIFY: Unified index for range filtered approximate nearest neighbors search,” *Proceedings of the VLDB Endowment*, vol. 18, no. 4, pp. 1118–1130, 2024. [Online]. Available: <https://www.vldb.org/pvldb/vol18/p1118-yao.pdf>
- [30] J. Engels, B. Landrum, S. Yu, L. Dhulipala, and J. Shun, “Approximate nearest neighbor search with window filters,” in *Proceedings of the 41st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 235. PMLR, 2024, pp. 12 469–12 490. [Online]. Available: <https://proceedings.mlr.press/v235/engels24a.html>
- [31] Y. A. Malkov and D. A. Yashunin, “Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 824–836, 2020. [Online]. Available: <https://doi.org/10.1109/TPAMI.2018.2889473>
- [32] OpenFlights, “Airport and airline data,” Official data page, 2026, accessed 9 September 2026. [Online]. Available: <https://openflights.org/data/>
- [33] OurAirports, “Open data downloads,” Official data page, 2026, accessed 9 September 2026. [Online]. Available: <https://ourairports.com/data/>
- [34] OpenStreetMap Foundation, “Planet.osm replication diffs and licensing,” Official documentation, 2026, accessed 9 September 2026. [Online]. Available: <https://wiki.openstreetmap.org/wiki/Planet.osm/diffs>