

RAGOps Proof PDF: Nonsensical Token Corpus

Purpose: deterministic retrieval tests using unique tokens embedded in text.

Generated: 2026-01-30

If your system answers without citing the exact token-containing chunk, treat it as hallucination.

Token Index (for evaluation prompts)

ID	Token	Notes
1	XQJ-77-NEBULA-SHARD	Unique string; must be retrieved verbatim
2	KAPPA-13-GLASSFROG	Unique string; must be retrieved verbatim
3	OMEGA-9-DRIFT-LATTICE	Unique string; must be retrieved verbatim
4	ZETA-42-VOID-ANCHOR	Unique string; must be retrieved verbatim
5	N4N0-QUANT-PLUME	Unique string; must be retrieved verbatim
6	RUNEKEY-ALPHA-999	Unique string; must be retrieved verbatim
7	SIGIL-BETA-0XDEAD	Unique string; must be retrieved verbatim
8	CRYPT-ORCHID-17	Unique string; must be retrieved verbatim
9	FROSTBYTE-TOTEM-808	Unique string; must be retrieved verbatim
10	ECHO-TORUS-31415	Unique string; must be retrieved verbatim

Section A: Narrative Corpus

This document contains deliberately nonsensical identifiers used to verify retrieval-augmented generation. Each token is unique and should appear in only one local context. Example: XQJ-77-NEBULA-SHARD appears in a paragraph about imaginary mineral logistics.

Imaginary logistics note: the shipment label reads 'KAPPA-13-GLASSFROG' and is attached to crate 6B. Any answer referencing this label must cite the relevant chunk.

Formatting stress test: [brackets] {braces} <angles> and odd spacing. Token: OMEGA-9-DRIFT-LATTICE.

Another line with line breaks and an inline code-ish marker: `ZETA-42-VOID-ANCHOR`.

Mini QA Prompts (ask your system)

Q1: What is the shipment label attached to crate 6B?

Q2: Which token is presented as a code-ish marker?

Q3: Identify the token mentioned in the formatting stress test line.

Section B: Sparse Records

Record 01 | asset_id=MORPHIC-TEA-KETTLE-7 | status=ACTIVE | region=NW
Record 02 | asset_id=PLASMA-PENGUIN-204 | status=QUARANTINED | region=SE
Record 03 | asset_id=DUSTY-COMET-61 | status=ARCHIVED | region=NE
Record 04 | asset_id=ARC-LAMP-SQUID-12 | status=ACTIVE | region=SW
Record 05 | asset_id=PAPERCLIP-PARADOX-3 | status=UNKNOWN | region=CENTRAL

Evaluation Notes

Expected behavior:

- Retrieval should surface the record line containing the asked asset_id.
- Answers should include the exact token string and cite the supporting record line.
- If the token is not in any retrieved chunk, the system should respond 'not found' rather than guessing.

Additional token for edge case: SIGIL-BETA-0XDEAD (do not confuse with similar hex strings).