CompreSeed-ZDS

Zero-Decompression Semantic Retrieval via Irreversible Core Compression

Author: Yoshikazu Nakamura
Affiliation: Independent Researcher, Aichi, Japan
Location: Aichi, Japan
Email: info@xinse.jp

Table of Contents

1. Abstract

Conventional information retrieval systems rely on a multi-step process: decompress data, tokenize, vectorize, embed, and then search. Each stage introduces latency, computational load, and memory overhead. Even modern RAG and vector-search systems require either GPU resources or reconstructable embedding spaces, both of which carry operational and security costs.

This paper introduces **CompreSeed-ZDS**, a retrieval paradigm in which **all search operations occur directly within the compressed semantic space**, without any decompression, tokenization, or embedding. CompreSeed's irreversible semantic compression produces compact meaning-cores that can be searched with high semantic accuracy, enabling a **zero-decompression, CPU-only retrieval pipeline**.

CompreSeed-ZDS represents a new class of efficient, scalable, and secure information retrieval, suitable for local devices, enterprise systems, and LLM-hybrid architectures.

2. Introduction

Traditional IR systems require full or partial decompression before retrieval:

- Full-text search requires storing and scanning raw text.
- Embedding-based retrieval requires vectorizing documents and storing large vectors.
- RAG pipelines depend on GPU resources for embedding generation.

These approaches are costly, heavy, and vulnerable to data leakage.

CompreSeed-ZDS proposes a simpler formula:

compressed documents → search directly on compressed cores → return meaningfully relevant results

No decompression. No vectors. No tokenization.

This creates a new IR category: **zero-decompression semantic retrieval**.

3. Background

3.1 Full-Text Search

Efficient for lexical matching but poor for semantic reasoning.

3.2 Embedding-Based Vector Search

Provides high semantic accuracy but requires:

- GPU resources
- Large embedding stores
- Vulnerable vector spaces
- Expensive indexing

3.3 BM25 and Keyword-Based IR

Lightweight but often fails on synonymy and long-text semantics.

3.4 Need for a New Retrieval Paradigm

The world needs:

- GPU-free semantic retrieval
- Zero decompression overhead
- Non-reconstructable storage

- High accuracy on meaning-level queries

CompreSeed-ZDS emerges as the first system meeting all of these criteria simultaneously.

4. The CompreSeed Compression Pipeline

CompreSeed converts each document into a compact **semantic core** through multiple stages:

1. **Normalization** – cleaning, structural reduction
2. **Topic Extraction** – identifying dominant meaning axes
3. **Semantic Distillation** – collapsing lexical redundancy
4. **Surface Information Removal** – discarding reconstructable elements
5. **Core Generation** – outputting compressed meaning units

The result is a small, non-reconstructable representation (400–1200 chars) that captures meaning but **cannot produce original sentences**.

This irreversible compression is the foundation of zero-decompression retrieval.

5. Zero-Decompression Retrieval

Traditional Paradigm

compressed storage → decompression → tokenization → embedding → search

CompreSeed-ZDS Paradigm

compressed storage → direct semantic search (no decompression)

Advantages

- **Zero decompression overhead**
- **Significantly reduced I/O**
- **No tokenization or embedding steps**
- **Faster retrieval pipeline**
- **No GPU dependency**
- **Non-reconstructable semantic space**

This enables large-scale semantic retrieval on devices as small as consumer laptops.

6. Similarity Model and Scoring

CompreSeed-ZDS compares meaning cores using:

- Topic-axis overlap
- Semantic-feature distance
- Compressed-pattern similarity
- Key-concept density
- Core-structure alignment

These methods allow meaningful semantic matching **without vectors**, **without neural embeddings**, and **without decompressing documents**.

The scoring function is optimized for:

- high recall of meaning-relevant documents
- low noise from lexical variations
- robustness to paraphrasing
- fast execution on CPU-only environments

7. Performance Characteristics

Dataset Scale

- Over **3 million documents** tested
- Disk size reduced to **1.8 GB**

Retrieval Speed

- **0.2–0.8 seconds** typical latency
- Fully CPU-based
- 2–3 GB RAM usage

Resource Footprint

- Zero decompression
- No GPU requirement
- No embedding generation
- No large vector DB

CompreSeed-ZDS achieves practical, fast semantic retrieval under minimal hardware conditions.

8. Comparison with Traditional Retrieval Methods

Method	Decompression	GPU	Semantic Ability	Security	Cost
Full-Text Search	Required	No	Low-Medium	Low	Medium
BM25	No	No	Low	Low	Low
Embedding Search (RAG)	No	Yes	High	Medium	High
Encrypted DB	Yes	No	N/A	Medium	Medium
CompreSeed-ZDS	**No**	**No**	**High**	**Very High**	**Very Low**

CompreSeed-ZDS is the **only method** that simultaneously offers:

- semantic retrieval
- zero decompression
- zero GPU
- irreversibility
- low resource cost

9. Applications

9.1 Local and Offline Devices

Laptops, field devices, kiosks, offline terminals.

9.2 Enterprise Search

High-volume document repositories without GPU budgets.

9.3 Government and Municipalities

Regulatory documents, guidelines, policy databases.

9.4 Medical and Legal Fields

Meaning-level search without privacy risk.

9.5 LLM Memory Systems

Acts as a **semantic external memory** for small LLMs.

9.6 Embedded or Edge Environments

IoT systems, industrial machines, secure devices.

10. Conclusion

CompreSeed-ZDS introduces a fundamentally new retrieval paradigm:

semantic search directly in compressed space,
with no decompression, no tokens, and no vectors.
By combining irreversible semantic compression with lightweight CPU execution,
CompreSeed-ZDS achieves:

- fast semantic search,
- low hardware cost,
- high security,
- zero reconstructability,
- excellent integration with LLM systems.

This architecture opens the door to next-generation retrieval systems
that are efficient, secure, and scalable.

11. Final Remarks

CompreSeed-ZDS is not merely an optimization—it is a new category of information retrieval.
Future developments may include:

- advanced scoring models,
- compressed multi-modal retrieval,
- hybrid LLM-ZDS reasoning systems,
- federated semantic search across secure nodes.

The methodology sets a new foundation for
lightweight, secure, and intelligent retrieval systems.