NASA Vitalis Project

Methodological Framework and Analysis Rationale

1 Research Objective and Analytical Framing

The NASA Vitalis Project examines how expert domain knowledge is encoded, structured, and organized in scientific writing. Rather than summarizing content or evaluating individual claims, the project models the *structural properties* of expert knowledge representations as they emerge across large collections of technical texts.

Expert-authored scientific literature is treated as a high-density reference point for how understanding appears when it is well-formed rather than misperceived. This framing enables analysis of organizational features such as conceptual hierarchy, connectivity, and redundancy, which are often absent or distorted in learner-generated texts despite surface-level confidence. The project is explicitly comparative, using expert structures as a reference distribution against which learner misconception patterns identified in the parallel *Illusion of Mastery* project are examined.

The objective is not to determine what experts know, but how expert understanding is *organized* at scale.

2 Corpus Selection and Scope Definition

The corpus comprises over 600 expert-authored scientific papers drawn from the publicly available NASA research database. These texts represent mature, peer-reviewed scientific communication within highly technical domains, providing a consistent reference for expert knowledge organization.

Expert-authored material was defined using corpus-level criteria rather than subjective assessment, including document type, length, and source consistency. This approach emphasizes structural regularities that emerge across many texts rather than idiosyncratic features of individual papers.

Corpus scale is central to the project's design. Structural properties such as conceptual density and hierarchy become visible only when examined across hundreds of documents, making large-scale analysis essential to the research question.

The project does not assess factual correctness, scientific novelty, or individual claim validity. Its scope is intentionally limited to modeling large-scale organizational patterns characteristic of expert knowledge representations.

3 Preprocessing Methodology

Preprocessing decisions were guided by the goal of preserving conceptual structure while minimizing noise unrelated to knowledge organization.

Text cleaning involved removing or normalizing elements such as citations, equations, and metadata that contribute minimally to conceptual relationships when treated as raw text. Care was taken to preserve domain-relevant terminology embedded in explanatory language.

Segmentation followed scientific rhetorical structure rather than fixed-length token windows. Texts were divided into sections and further segmented into semantically coherent units corresponding to claims, mechanisms, or explanatory blocks. This approach reflects how expert arguments are constructed and supports modeling at an appropriate level of abstraction.

Normalization procedures were applied consistently across the corpus to ensure comparability while avoiding transformations that would collapse meaningful conceptual distinctions.

4 Analytical Framework and Structural Modeling

Within this framework, a concept is operationally defined as a semantically coherent unit of meaning derived from expert-authored text segments. Concepts are not mapped to predefined ontologies but emerge from patterns of semantic similarity across the corpus.

Relationships between concepts are inferred through similarity-based modeling and represented as explicit structural connections. These relationships are aggregated into graph-based representations, with nodes corresponding to concepts and edges encoding inferred associations.

Analysis focuses on emergent structural properties rather than individual semantic interpretations. Examined features include conceptual clustering, hierarchical depth, connectivity patterns, and the distribution of central versus peripheral concepts. Together, these properties characterize how expert knowledge balances compression, redundancy, and specialization at scale.

5 Comparative Analysis with Learner Knowledge Models

To contextualize expert structures, selected features were compared with learner misconception patterns identified in the *Illusion of Mastery* project. Learner models capture cases where perceived understanding diverges from underlying conceptual organization.

Comparisons were conducted at the level of structural features such as conceptual density, hierarchy, and stability rather than correctness or content overlap. This approach enables comparison without relying on normative judgments about accuracy.

The analysis shows that expert texts tend to exhibit tightly integrated conceptual cores with peripheral specialization, while learner texts more frequently display shallow hierarchies, fragmented connections, and elevated redundancy.

6 Scope, Constraints, and Future Directions

This framework is constrained by its reliance on textual representations of knowledge and does not capture non-textual forms of expertise. Additionally, structural similarity does not imply pedagogical effectiveness or conceptual accessibility.

Despite these constraints, the framework offers a scalable approach for modeling expert knowledge organization across domains. Future applications could include evaluating instructional materials by comparing their structural properties against expert reference models or extending the analysis to other technical corpora to examine cross-domain organizational patterns.

This document accompanies a separate technical evidence file containing representative code excerpts and pipeline diagrams illustrating the implementation of the methods described above.