2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

SHARED EMBEDDING SPACES FOR AEC ONTOLOGIES: AN EMPIRICAL STUDY OF DISCOVERY AND ALIGNMENT

Tobias Heimig-Elschner[1,2], Jyrki Oraskari[2,3], and Jakob Beetz[2]
[1]Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), Germany
[2]Design Computation, RWTH Aachen University, Germany
[3]Cluster of Excellence CARE, TU Dresden and RWTH Aachen University, Germany

Abstract

The adoption of ontologies in the AEC domain is increasing, yet ontology reuse remains constrained because identifying suitable ontologies and concepts is still largely manual. This paper investigates whether shared embedding spaces can capture semantic relatedness across AEC ontologies, thereby supporting ontology discovery and alignment. We test three hypotheses regarding expert-aligned similarity, the spatial proximity of related classes and the superiority of structure-aware over purely lexical models. We embedded several widely used AEC ontologies and evaluated ontology and class-level similarities. Results suggest that embeddings operationalize expert intuition and lower barriers to ontology reuse.

Introduction

A significant obstacle to ontology reuse is how many ontologies are designed: they often lack adequate documentation and provide no human-readable specifications, thereby imposing unnecessary friction on anyone attempting to adopt or extend them (Fernández-López et al., 2019). Many built-environment ontologies have been created recently, but users find it difficult to locate suitable ones, which limits their practical application (Akbarieh et al., 2025). Identifying suitable ontologies and concepts is still largely manual.

We assume that ontologies and their constituent entities can be represented in a vector space such that semantic relatedness systematically correlates with geometric proximity. We explore hypotheses about how the embeddings of lexical ontology notations relate to expert assessments and known alignments. Specifically, we suggest that:

- **H1 (Ontology-level similarity correlates with expert recommendations).** If ontology embeddings capture domain structure, then similarity in the embedding space should align with expert judgments of task or domain suitability.
- **H2 (Semantically corresponding classes are closer than unrelated classes).** Useful ontology embeddings must preserve class-relatedness.
- **H3 (Context-enriched lexical representations improve embedding quality).** Ontologies provide structural context, such as class connectivity and local relations.

We evaluate our hypotheses by embedding a curated set of widely used AEC ontologies and their classes using both text-based (labels + descriptions) and graph-based (RDF/OWL) strategies. These embeddings are employed for similarity-based ontology search and class-level alignment suggestions. Expert-guided evaluations and quantitative analyses of the embedding space reveal how closely the embeddings reflect expert judgment and semantic structure, and demonstrate that they can streamline ontology engineering workflows by lowering barriers to ontology reuse and integration in construction informatics.

Our main contributions include: (i) an LLM embedding–based, ontology-level semantic search framework; (ii) a class-level alignment method using similarity neighborhoods for an ontology base; (iii) an expert-aligned evaluation approach for comparing and assessing embedding spaces and (iv) the validation and development of a foundational embedding hypothesis for AEC ontologies.

Related Work

Linked Open Vocabularies (LOV) is a curated registry of RDF vocabularies that shows how they link, tracks usage in the Linked Data cloud, and offers search, SPARQL, and API access to find and reuse vocabularies. Despite strong reuse of LOV's core vocabularies, domain-specific vocabularies remain underlinked within and across domains (Maratsi et al., 2023). More recently, Akbarieh et al. (2025) presented BE-OLS, a built-environment-specific ontology lookup service that catalogs and organizes over 100 ontologies in the building environment domain with structured metadata and search/filter functions. It provides a custom, domain-aware evaluation framework (connectivity, accessibility, and documentation/reuse). Current limitations include an incomplete ontology collection, manually applied evaluation, and no LLM-based assistance for ontology search and analysis.

LLM embeddings (Tao et al., 2024) used in semantic searches encode text as high-dimensional vectors that capture semantic meaning. Models like BERT or all-MiniLM-L6-v2 (Reimers and Gurevych, 2024) convert text into vector representations, ensuring that semantically similar content is aligned closely and facilitating comparison based on meaning rather than exact words.

Wang et al. (2015) presented a method for short text categorization that combines semantic clustering with convolutional neural networks (CNN). Although newer techniques such as transformers and large language models (LLMs) now outperform this approach, it remains effective for lightweight applications. This method employs a small, supervised CNN operating on semantic matrices, in contrast to LLMs that produce embeddings via unsupervised transformer pretraining with attention mechanisms, which offer richer context but require more computational resources.

A more recent method, introduced by Khan et al. (2020), uses sentence-embedding-based clustering to summarize discussion threads. It leverages Word2Vec[1] to create sentence embeddings and applies K-medoids clustering—a partitioning technique that selects actual data points as cluster centers—to group replies that are semantically similar. A ranking mechanism based on quality-related textual features then determines the most relevant responses for the summary. This fully automatic, domain-independent approach performs well on precision, recall, and F-measure. However, its reliance on Word2Vec and traditional clustering may limit its ability to capture complex semantic nuances compared with more advanced models, such as large language models (LLMs).

Teymurova et al. (2024) introduced OWL2Vec4OA, a version of OWL2Vec* tailored for ontology alignment by integrating confidence-weighted seed mappings into the random-walk process. This makes the learned embeddings more alignment-aware and improves matching performance in experiments. Its main limitations are dependence on seed-mapping quality, added computational complexity, and sensitivity to hyperparameters.

Methodology

Problem Setting and Representation Framework

To make the decisions made in representation and evaluation explicit, we introduce the following formal setting:

Let an ontology artifact be denoted by O. Let $\mathcal{D} = \{O_1, \dots, O_N\}$ denote the ontology dataset, i.e. a finite collection of ontology artifacts processed jointly.

For each ontology O we assume a materialized RDF graph:

$$\mathcal{G}_O = (V, E) \tag{1}$$

where V denotes the set of RDF nodes and E the set of triples with predicates P. A triple is written as $(s, p, o) \in E$ with $s, o \in V$ and $p \in P$. Based on this, let C_O be the set of classes declared in O (e.g., resources typed as `owl:Class` or `rdfs:Class`) and P_O be the set of properties declared in O (e.g., `owl:ObjectProperty`, `owl:DatatypeProperty`, `rdf:Property`).

We define the set of representable entities [2] at both dataset and ontology level:

$$\mathcal{X}_O := \{O\} \cup C_O \cup P_O \tag{2}$$

$$C_{\mathcal{D}} := \bigcup_{O \in \mathcal{D}} C_O, \quad P_{\mathcal{D}} := \bigcup_{O \in \mathcal{D}} P_O, \quad \mathcal{X}_{\mathcal{D}} := \bigcup_{O \in \mathcal{D}} \mathcal{X}_O. \tag{3}$$

The formalism for representation and embeddings is defined as follows: Let $\mathcal{T}$ be the space of textual representations (strings), $f : \mathcal{T} \to \mathbb{R}^d$ be an embedding function and $\mathrm{t}_m : \mathcal{X} \to \mathcal{T}$ be a representation variant m. The embedding of an entity $x \in \mathcal{X}$ under variant m is:

$$\mathrm{v}_x^{(m)} := f(\mathrm{t}_m(x)) \tag{4}$$

All embeddings of interest for a fixed representation variant m live in a shared dataset-level embedding space

$$\mathcal{E} := \mathcal{E}_m = \{\mathrm{v}_x^{(m)} \mid x \in \mathcal{X}_{\mathcal{D}}\}. \tag{5}$$

Let $\delta(\cdot, \cdot)$ denote cosine distance on $\mathcal{E}$.

For any entity $x \in \mathcal{X}$, define three optional lexical fields: $\ell(x)$ – label or title string (identifier-like), $d(x)$– description or definition string (content-like) and $c(x)$– comment or remark string (additional notes). Extraction is performed via documented, stable predicate lists (implementation detail), but the formalism assumes only that $\ell(x), d(x), c(x)$ are string-valued or empty.

Use-case descriptions U are treated as external textual queries embedded via the same function f, but are not elements of $\mathcal{X}_O$.

Throughout the remainder of this section, $\mathbb{E}[\cdot]$ denotes the empirical expectation with respect to the sampling procedure specified in each context. Unless stated otherwise, sampling is performed uniformly over dataset-level entities $x \in \mathcal{X}_{\mathcal{D}}$ (or over constrained subsets such as relation-defined pairs).

Definition of a Well-Formed Embedding Space

Based on the foundational hypothesis that: *There exists an embedding space in which semantic relatedness between ontologies and ontology entities is reflected by geometric proximity*. We define a set of geometric (*G1–G3* and semantic properties *S1 & S2* that suggest that such an embedding space is well-formed. The following criteria define necessary geometric and semantic properties that an embedding space must satisfy to be considered suitable for ontology discovery and alignment tasks. G1–G3 are treated as necessary but not sufficient conditions for semantic validity.

G1 – Global Discriminability

In high-dimensional spaces, distances tend to concentrate, making nearest-neighbor relations unreliable; measuring distance contrast provides a global diagnostic of whether the embedding space preserves discriminative power (Beyer et al., 1999; Aggarwal et al., 2001). A well-formed embedding space must exhibit sufficient distance contrast, such that semantic distinctions are not lost due to high-dimensional distance concentration. This is operationalized via a distance contrast measure introduced by Beyer et al. (1999):

[1] `https://code.google.com/archive/p/word2vec/`

[2] When the context is clear, we write $\mathcal{X}$ for $\mathcal{X}_{\mathcal{D}}$ (dataset-level entities) and use $\mathcal{X}_O$ when referring to a specific ontology.

$$D = \{\delta(\mathrm{v}_i, \mathrm{v}_j) \mid i \neq j\}. \tag{6}$$

$$RC_q = \frac{\mathbb{E}[D] - Q_5(D)}{Q_5(D)} \tag{7}$$

where δ ranges over all pairwise distances in $\mathcal{E}$, with i, j indexing embeddings corresponding to entities $x_i, x_j \in \mathcal{X}_{\mathcal{D}}$.

G2 – Local Semantic Coherence

Relative neighborhood statistics remain meaningful in high-dimensional spaces because they rely on comparative distances rather than absolute magnitudes (Beyer et al., 1999; Radovanović et al., 2009). They serve as a robust replacement for classical clusterability tests (e.g. Hopkins (Hopkins and Skellam, 1954)), which degrade in high dimensions (Adolfsson et al., 2019). A well-formed embedding space must exhibit locally coherent neighborhoods, such that nearby vectors are significantly closer than randomly chosen vectors. To capture local coherence robustly, we adopt *relative k-nearest-neighbor distance ratio*, which compares expected distances to k nearest neighbors (k-NN) against those drawn from random pairs, in the spirit of neighborhood preserving metrics used in embedding evaluation (Boggust et al., 2022).

$$\bar{\delta}_k(x) := \frac{1}{k} \sum_{x' \in k\text{-NN}(x)} \delta(\mathrm{v}_x, \mathrm{v}_{x'}) \tag{8}$$

$$R_k = \frac{\mathbb{E}[\bar{\delta}_k(x)]}{\mathbb{E}[\delta(\mathrm{v}_x, \mathrm{v}_{x'})]} \quad \text{with} \quad R_k \ll 1 \tag{9}$$

where k-NN(x) denotes the set of the k nearest neighbors of x in $\mathcal{E}$, (x, x') a uniformly sampled random pair of entities, and expectations are taken over uniformly sampled entities $x \in \mathcal{X}_{\mathcal{D}}$. The parameter k controls the neighborhood size.

G3 – Non-Degenerated Dimensionality

Effective dimensionality captures whether embeddings meaningfully occupy a structured manifold rather than degenerating into noise or redundancy (Recanatesi et al., 2022; Giaffar et al., 2024) which can be assessed via the *participation ratio (effective dimensionality)* as proposed by Recanatesi et al. (2022):

$$D_{\text{eff}} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2} \tag{10}$$

where λ_i are the eigenvalues of the covariance matrix of the embedding vectors in $\mathcal{E}$.
Following this, a well-formed embedding space satisfies:

$$1 \ll D_{\text{eff}} \ll d \tag{11}$$

While G1–G3 assess geometric well-formedness independently of domain semantics, S1 and S2 evaluate whether the embedding space aligns with ontological meaning and expert judgment.

S1 – Ontology-Aware Semantic Alignment

An intrinsic characteristic of ontologies is the explicit declaration of semantic relations between classes (intra- and inter-ontology), such as equivalence mappings, subsumption, and domain-specific predicates. A meaningful embedding space should preserve these relations, embedding semantically related classes closer together than unrelated ones. Intrinsic evaluation frameworks for ontology embeddings therefore emphasize explicit measurement of relational preservation (Alshargi et al., 2019).
Formally, let $\mathcal{R}$ denote the set of declared semantic relations between classes in the ontology dataset $\mathcal{D}$:

$$\mathcal{R} := \bigcup_{O \in \mathcal{D}} \mathcal{R}_O \;\cup\; \mathcal{R}_{\text{cross}}, \quad \mathcal{R} \subseteq C_{\mathcal{D}} \times C_{\mathcal{D}} \tag{12}$$

where $\mathcal{R}_O$ contains intra-ontology class pairs and $\mathcal{R}_{\text{cross}}$ inter-ontology class pairs. Class pairs $(i, j) \in C_{\mathcal{D}} \times C_{\mathcal{D}}$ are induced by predicate assertions (i, p, j) or (j, p, i). No logical inference or reasoning closure is applied. Let the binary semantic relatedness indicator $s_{ij} \in \{0, 1\}$ be

$$s_{ij} := \begin{cases} 1 & \text{if } (i, j) \in \mathcal{R}, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Define the sets of *related* and *unrelated* class pairs:

$$P_{\text{rel}} = \{(i, j) \mid s_{ij} = 1\}, \quad P_{\text{unrel}} = \{(i, j) \mid s_{ij} = 0\} \tag{14}$$

Pairs in P_{unrel} are uniformly sampled such that $|P_{\text{rel}}| = |P_{\text{unrel}}|$. Let $d_{ij} := \delta(\mathrm{v}_i, \mathrm{v}_j)$.
A well-formed embedding space satisfies

$$\mathbb{E}_{(i,j) \in P_{\text{rel}}}[d_{ij}] \;<\; \mathbb{E}_{(i,j) \in P_{\text{unrel}}}[d_{ij}]. \tag{15}$$

We further define the alignment ratio

$$R_{\text{align}} := \frac{\mathbb{E}_{(i,j) \in P_{\text{unrel}}}[d_{ij}]}{\mathbb{E}_{(i,j) \in P_{\text{rel}}}[d_{ij}]} \tag{16}$$

$R_{\text{align}} > 1$ indicates ontology-aware semantic alignment.
Statistical significance is assessed via a stratified class-identity permutation test ($K = 1000$), where class identities are shuffled within ontology strata while distances are held fixed. The empirical one-sided p-value $P(\rho_{\text{null}} \geq \rho_{\text{obs}} \mid H_0)$ is computed from the resulting null distribution of Spearman $\rho(-d_{ij}, s_{ij})$.

S2 – External Semantic Consistency

While S1 evaluates semantic alignment intrinsic to the ontology structure, external semantic consistency assesses whether embedding-based similarity aligns with expert intuition in practical ontology selection scenarios. If an embedding space is well-formed, ontologies deemed relevant by domain experts for a given use case should appear among the most similar ontologies according to embedding-based rankings. Ontology-level similarity is computed using lexical and context-enriched lexical representations. Ontology embeddings are obtained directly from the corresponding textual representations.

Let U denote a use-case description and O the set of candidate ontologies. Let $E_U \subseteq O$ denote the set of ontologies identified as relevant by domain experts for use case U, allowing multiple ontologies to be considered equally relevant.

Let $r_{\text{embed}}(U) = (o_1, o_2, \dots)$ be the ontology ranking induced by embedding similarity between U and ontology representations, where $o_i \in O$. Define the set of top-k retrieved ontologies using precision-at-k and recall-at-k using the top-k set $R_k(U) := \mathrm{TopK}(r_{\text{embed}}(U), k)$:

$$P_k(U) = \frac{1}{k} \sum_{o \in R_k(U)} 1[\, o \in E_U \,], \tag{17}$$

$$R_k(U) = \frac{1}{|E_U|} \sum_{o \in R_k(U)} 1[\, o \in E_U \,]. \tag{18}$$

Here, Precision@k measures the proportion of relevant ontologies among the top-k retrieved results, while Recall@k measures the proportion of all expert-relevant ontologies retrieved within the top-k. Since expert relevance sets may contain fewer than k ontologies, the maximum achievable precision is bounded by $|E_U|/k$, while recall remains unaffected.

To assess whether retrieval performance exceeds chance, observed $P_k(U)$ values are compared against a random baseline. Under random selection, the number of relevant ontologies among the top-k follows a hypergeometric distribution, with expected precision $\mathbb{E}[P_k(U)] = |E_U|/|O|$. A well-formed embedding space exhibits consistently high precision and recall values across use cases, indicating that expert-relevant ontologies are ranked among the most similar results, and achieves precision values that systematically exceed the corresponding random baseline, reflecting performance beyond chance.

Ontology Representations and Embeddings

In this paper, we define and evaluate two complementary lexical representations for ontologies O and their associated classes C_O. These representations rely solely on human-readable lexical information such as labels, descriptions, and comments, or enrich this information with bounded contextual signals derived from related entities. In addition, we include a purely structural RDF2Vec-based representation - as introduced by Ristoski and Paulheim (2016) - as an established baseline focusing on graph structure, used for comparative evaluation under criteria *G1–G3* and *S1*.

Lexical embedding v_{lex}

The *lexical representation* variant t_{lex} captures only the immediate human-readable lexical documentation attached to an entity, without incorporating any explicit structural context. For any $x \in \mathcal{X}$:

$$t_{\text{lex}}(x) := \mathrm{concatFields}\big(\ell(x),\ d(x),\ c(x)\big) \tag{19}$$

Here, $\mathrm{concatFields}(\cdot)$ denotes deterministic concatenation with a fixed ordering and separators.

The representation t_{lex} is deterministic, structure-independent, and fully interpretable. It serves as the lexical baseline for evaluating embedding quality and semantic alignment.

$$v_{\text{lex}}(x) = f(t_{\text{lex}}(x)) \text{ for all } x \in \mathcal{X} \tag{20}$$

Context-Enriched Lexical embedding v_{conlex}

The *context-enriched lexical representation* variant t_{conlex} extends t_{lex} by incorporating bounded and deterministic lexical context derived from related entities, while remaining text-based. For an ontology O, let $\bigoplus$ denote deterministic aggregation (concatenation) with a fixed ordering:

$$t_{\text{conlex}}(O) := \mathrm{concatFields}\Big(t_{\text{lex}}(O),\ \bigoplus_{c \in C'_O} t_{\text{lex}}(c)\Big) \tag{21}$$

where $C'_O \subseteq C_O$ is a fixed, deterministic subset of representative classes. In this work, classes are ranked by first-order connectivity (number of RDF edges), with ties resolved by alphabetical ordering of IRIs. The top $\min(50, |C_O|)$ classes are selected.

For a class $c \in C_O$:

$$t_{\text{conlex}}(c) := \mathrm{concatFields}\Big(t_{\text{lex}}(c),\ \bigoplus_{p \in P_c^{(n)}} t_{\text{lex}}(p)\Big) \tag{22}$$

where $P_c^{(n)} \subseteq P_O$ denotes properties connected to c within RDF graph depth n under a fixed neighborhood policy; depth $n = 1$ is used by default.

The representation t_{conlex} remains deterministic and human-interpretable, while introducing controlled local contextual signal in a reproducible manner.

$$v_{\text{conlex}}(x) = f(t_{\text{conlex}}(x)) \text{ for all } x \in \mathcal{X} \tag{23}$$

Structural Embedding v_{struct}

In addition to the lexical representation variants, we include a purely structural embedding baseline based on RDF2Vec. In contrast to v_{lex} and v_{conlex}, which follow the mapping $v_x^{(\bullet)} = f(t_\bullet(x))$, the structural variant directly induces embeddings from RDF graph structure. Formally, v_{struct} is defined only for class-level entities $c \in C_{\mathcal{D}}$, yielding embeddings $\mathrm{v}_c^{(\text{struct})} \in \mathbb{R}^d$. Ontology-level embeddings are derived via aggregation over class-level embeddings:

$$\mathrm{v}_O^{(\text{struct})} := \frac{1}{|C_O|} \sum_{c \in C_O} \mathrm{v}_c^{(\text{struct})}. \tag{24}$$

The resulting embeddings capture global structural patterns of the RDF graph, are stochastic but reproducible under fixed seeds, and are not directly human-interpretable.

Results

Implementation and Dataset

Implementation

The methodology was instantiated as a modular processing pipeline with separated responsibilities for identity management, semantic storage, and representation. Ontology identity and provenance are maintained via a JSON-based registry. Resolved ontology serializations are loaded into an Apache Jena Fuseki triple store[3], which serves as the authoritative semantic state and supports SPARQL-based querying. Normalized ontology artifacts and intermediate results are persisted as file-based materializations to ensure reproducibility. Vector representations derived from textual and graph-based encodings are stored separately in a Weaviate vector database, allowing symbolic semantic data and learned embeddings to be managed independently while remaining linked through stable ontology identifiers. Vector representations are derived from both textual and graph-based encodings. Textual embeddings are computed using the transformer-based model `snowflake-arctic-embed2` (1024 dimensions), selected as a representative open-weight model for semantic similarity tasks Yu et al. (2024). Structural embeddings are generated using RDF2Vec (100 dimensions) with random walks (depth $d = 4$, 100 walks per entity) and a Word2Vec backend, consistent with commonly used configurations in prior RDF2Vec studies. Traversal excludes high-degree hub predicates (e.g., `rdf:type`) to reduce structural noise Ristoski and Paulheim (2016). Full embedding configurations are provided in the appendix [4] .

Dataset

From the 144 ontologies listed in the Built Environment Ontology Lookup Service (BE-OLS) (Akbarieh et al., 2025), 101 ontologies provided a reachable reference or download URL. Of these, 100 could be successfully resolved and parsed. Two ontologies were fully contained as subgraphs of others already included and were therefore not treated as independent artifacts. The resulting dataset comprises 98 distinct ontologies.

To characterize the availability of lexical information, ontologies were classified as **meta** (no class definitions), **formal** (classes without lexical signals, i.e., only blank nodes or unlabeled IRIs), **structural** (fewer than 25% of classes provide at least one lexical signal (e.g., `rdfs:label`, `skos:prefLabel`, `rdfs:comment`, `dc:description`)), or **lexical** (at least 25% of classes provide lexical information). In addition, we assessed whether available lexical information follows common, human-readable documentation patterns suitable for downstream processing by large language models.

For inclusion in the vectorization experiments, ontologies were required to provide ontology-level lexical information, a resolvable serialization (i.e., not metadata-only), and a lexical profile classified as either structural or lexical. Applying these criteria yields a final subset of 84 ontologies, comprising 7,086 classes.

Evaluation of the Embedding Space

This section reports the results for three global well-formedness criteria (G1–G3) introduced above. All evaluations are reported for ontology-level and class-level embeddings and for all three representation variants (Lex, ConLex, Struct).

G1 – Global Discriminability

G1 evaluates global discriminability by measuring distance contrast across the embedding space. All representations satisfy this criterion at both the ontology and class levels, indicating that none of the embedding spaces suffers from distance concentration effects. Structural embeddings exhibit the highest contrast, reflecting the sparse and the role-oriented nature of graph-based representations, while lexical and contextualized lexical embeddings show moderate but stable contrast characteristic of dense language-model spaces. The consistency of G1 across granularities confirms that subsequent local semantic analyses are not confounded by degenerate global geometry. (see Table 1)

Table 1: G1 – Global discriminability (distance contrast).

Level	v_{lex}	v_{conlex}	v_{struct}
O	0.71 ✓	0.55 ✓	2.18 ✓
C_O	0.64 ✓	0.66 ✓	2.22 ✓

G2 – Local Semantic Correspondence ($\tau = 0.5$)

Table 2: G2 – Local semantic correspondence for $k = 10$.

Level	v_{lex}	v_{conlex}	v_{struct}
O	0.68 ×	0.73 ×	0.55 ×
C_O	0.42 ✓	0.43 ✓	0.06 ✓

G2 reveals a clear granularity effect (see Table 2). While none of the representations satisfy the criterion at the ontology level, all representations consistently meet G2 at the class level. This indicates that semantic correspondence is preserved locally between individual ontology entities, but not necessarily when embeddings aggregate heterogeneous concepts at the ontology scale. Rather than indicating a deficiency, this result confirms that class-level embeddings are the appropriate resolution for evaluating fine-grained semantic relatedness, whereas ontology-level embeddings primarily capture broader thematic or structural similarity.

G3 – Non-Degenerated Dimensionality

G3 evaluates whether embedding spaces exhibit non-degenerate effective dimensionality (see Table 3). All representations pass this criterion at both the ontology and class levels, with effective dimensionalities substantially

[3] `https://jena.apache.org/documentation/fuseki2/`

[4] The appendix is provided online at `https://git-ce.rwth-aachen.de/Tobias1/ec3-2026-shared-embedding-spaces-for-aec-ontologies/-/tree/v1.0-ec3-paper-2026`.

greater than one and far below the nominal embedding dimension. Lexical and contextualized lexical embeddings occupy a relatively low- dimensional subspace of the original 1024-dimensional space, reflecting semantic compression through aggregation, while structural embeddings exhibit even more compact but information-dense representations. The increase in effective dimensionality from ontology to the class level aligns with higher semantic granularity and complements the G2 findings.

Table 3: G3 – Effective dimensionality of embedding spaces.

Level	v_{lex}	v_{conlex}	v_{struct}
O	25.58 ✓	36.30 ✓	16.82 ✓
C_O	59.47 ✓	65.26 ✓	13.20 ✓

S1 – Ontology-Aware Semantic Alignment

All representations achieve statistically significant alignment ($R_{\text{align}} > 1$, $p < 10^{-3}$), with semantically related pairs consistently closer than unrelated ones. Lexical and context-enriched representations show moderate alignment ($R_{\text{align}} \approx 1.55$), while the structural embedding yields substantially stronger separation ($R_{\text{align}} = 2.94$), indicating that graph-based context reinforces semantic correspondence. The observed rank correlations ($\rho \approx 0.56$–0.60) confirm robust but non-trivial alignment between embedding distances and relatedness (see Table 4).

Table 4: S1 - Ontology-Aware Semantic Alignment Evaluation.

Metric	v_{lex}	v_{conlex}	v_{struct}
R_{align}	1.56	1.55	2.94
Spearman ρ	0.58	0.56	0.60
p_{perm}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
$\mu_{\rho,\text{null}}$	0.08	0.06	0.05
Significance	✓	✓	✓

S2 - External Semantic Consistency

Across the five evaluated use cases, both lexical and contextualized lexical representations achieve precision@5 values above the random baseline, indicating that embedding-based retrieval captures some degree of expert-relevant signal. However, absolute performance remains low (P@5 ≈ 0.24–0.28), and no statistically significant results are observed across use cases. Lexical representations exhibit higher recall, while contextualized lexical representations achieve slightly higher precision (see Table 5), suggesting a trade-off between coverage and selectivity. Nevertheless, the overall performance indicates that ontology-level embeddings alone are insufficient to reliably reproduce expert ontology selection behavior under the tested conditions.

In contrast to S1, which evaluates intrinsic semantic alignment at the class level, S2 operates on highly aggregated ontology representations and sparse, non-ranked expert feedback, resulting in reduced discriminative power. Consequently, S2 does not provide confirmatory evidence for external semantic consistency in this setting.

Table 5: S2 – External semantic consistency (P@5 and R@5).

Metric	v_{lex}	v_{conlex}
Mean P@5	0.24	0.28
Mean R@5	0.30	0.15
Baseline $E[P@5]$	0.11	0.11
Baseline $E[R@5]$	0.07	0.07
$\Delta P@5$	0.13	0.17
$\Delta R@5$	0.23	0.08
Significant	0/5	0/5

Discussion

We investigated whether shared embedding spaces facilitate ontology discovery and alignment in the AEC domain (hypotheses H1–H3). All embedding variants met the geometric well-formedness criteria G1–G3, confirming stable, non-degenerate spaces. Semantic alignment at the class level (S1) was consistently observed across lexical, context-enriched lexical, and structural representations. In contrast, ontology-level semantic consistency (S2) yielded weak but above-random results, insufficient for reliable ontology selection. Thus, embedding-based similarity is valid primarily at the class-level abstraction, supporting the core hypothesis that semantic relatedness in ontologies is reflected by geometric proximity, while highlighting the abstraction levels where this relationship holds reliably.

With respect to H1, which hypothesizes that ontology-level embedding similarity correlates with expert recommendations, the results do not provide confirmatory evidence under the current evaluation setup. A primary limiting factor is the statistical underpowering of S2: unlike S1, which operates over thousands of class-level relation pairs, S2 relies on a small number of use cases and sparse, largely unranked expert relevance judgments. Combined with the strong aggregation inherent in ontology-level embeddings, this substantially reduces the sensitivity of retrieval-based evaluation measures such as precision@k and recall@k. Consequently, H1 cannot be rejected outright, but remains unconfirmed within the scope of this study and should be revisited using larger and more structured expert evaluations.

Beyond methodological considerations, the inconclusive results for H1 highlight a fundamental issue affecting ontology reuse. Ontology-level embeddings, derived solely from explicit information like labels, descriptions, and structure, fail to replicate expert selection behavior. This suggests that many ontologies lack sufficient information for reuse decisions, as experts also consider contextual factors like maturity, adoption, tooling, conventions, and experience—not encoded in standard documentation. This deficiency indicates broader communication issues within ontology ecosystems, echoing barriers caused by inadequate documentation. Additionally, S2 results imply that expert selection depends more on high-level lexical descriptions of scope and application than on detailed internal content, which is mainly reflected in class structure and relations.

The results strongly support H2, showing that semantic classes are embedded closer together when related, across all variants, and confirming that embedding spaces preserve fine-grained semantic relations within and across ontologies. This demonstrates their suitability for class alignment, correspondence discovery, and semantic exploration. The evaluation focuses on explicitly declared relations, isolating the effect of representation learning from reasoning variability. Including inferred relations could further improve semantic alignment by revealing implicit structures, highlighting the potential of knowledge graphs. The consistent results across different representations suggest that class-level semantic relatedness is a stable property of the embedding space, largely independent of the representation used.

The results support H3, which suggests that adding structural context improves embedding quality. Context-enriched lexical representations achieve results comparable to purely lexical ones in semantic alignment while remaining deterministic and interpretable, indicating that additional contextual aggregation does not consistently translate into improved alignment. Structural RDF2Vec embeddings have the highest alignment ratios, showing the importance of graph structure in maintaining ontological relations. Interestingly, large language model–based lexical embeddings already meet all geometric criteria (G1–G3) and achieve strong semantic alignment (S1), even though they weren't trained specifically for ontology embedding. This suggests that modern language models encode substantial latent domain knowledge, allowing them to recover meaningful semantic structures from sparse, diverse ontology data.

Taken together, these findings clarify the appropriate role of embedding-based methods in ontology engineering workflows. Embeddings should not be viewed as replacements for expert judgment at the ontology selection level (H1), but as powerful instruments for semantic exploration and alignment at the class level (H2), particularly when enriched with bounded structural context (H3). The context depth is intentionally limited to one hop ($n = 1$), focusing on directly related classes and properties while avoiding noise propagation and cyclic dependencies in graph traversal. Exploring deeper contextual representations may further enrich semantic signals, but introduces increased structural variance and is left for future work. By making explicit where semantic information is preserved and where it is absent, embedding-based representations both support ontology reuse and expose documentation gaps that warrant further attention.

Conclusion

This paper explored whether embedding-based representations can support ontology discovery and alignment in the AEC domain. Across all evaluated representations, the geometric well-formedness criteria (G1–G3) are satisfied, indicating that the resulting embedding spaces are stable and suitable for semantic analysis. At the class level, ontology-aware semantic alignment (S1) is consistently observed: semantically related classes are embedded closer together than unrelated ones, independent of representation choice. This confirms that embeddings preserve fine-grained semantic relationswithin and across ontologies and are well suited for class-level exploration and alignment tasks. These results further provide empirical support for the foundational hypothesis that semantic relatedness between ontology entities can, to a meaningful extent, be reflected by geometric proximity in a shared embedding space.

A key empirical finding is that LLM-based lexical embeddings perform surprisingly well in this setting, despite not being trained on ontologies or formal knowledge graphs. Even without explicit access to ontological structure, these embeddings satisfy all geometric criteria and exhibit statistically significant semantic alignment. While context-enriched lexical representations do not yield a significant improvement, incorporating structural information enhances alignment performance. Purely structural embeddings achieve the strongest alignment effects, albeit at the cost of interpretability.

At the ontology level, external semantic consistency with expert recommendations (S2) remains inconclusive. This result reflects both methodological limitations—strong aggregation and sparse expert feedback—and fundamental constraints on ontology reuse, where expert selection relies on contextual factors not explicitly encoded in ontology artifacts. Overall, the results support embeddings as a reliable semantic substrate for exploration and alignment at appropriate levels of abstraction, while clarifying their limitations for ontology-level decision-making.

Future Work

Future work should extend this study along three complementary directions. First, external semantic consistency (S2) should be evaluated using more mature and statistically robust designs, including a larger set of use cases, ranked or graded expert judgments, and repeated evaluations across expert groups. Such designs would enable stronger conclusions about ontology-level similarity and better isolate the factors influencing expert ontology selection. Second, the role of embeddings in concrete alignment tasks warrants systematic investigation. This includes evaluating their utility for candidate correspondence generation, semi-automated class alignment, and expert-in-the-loop alignment workflows, where embeddings serve as a semantic prior rather than a definitive decision mechanism. Third, embedding-based representations open the door to two-dimensional semantic mappings of the AEC ontology ecosystem. Future research should examine the stability, interpretability, and analytical value of such maps for ecosystem-level exploration, comparison, and gap analysis. Together, these directions would further clarify the scope and limits of embedding-based methods and support their principled integration into ontology engineering practice without overextending their claims.

Acknowledgments

Claims expressed in this article do not necessarily have to coincide with the positions of the BBR/BBSR. This work was partially funded by the DFG (German Research Foundation) under EXC 3115 - 533767731.

References

Adolfsson, A., Ackerman, M., and Brownstein, N. C. (2019). To cluster, or not to cluster: An analysis of clusterability methods. Pattern Recognition, 88:13–26.

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. In Goos, G., Hartmanis, J., Van Leeuwen, J., Van Den Bussche, J., and Vianu, V., editors, Database Theory — ICDT 2001, volume 1973, pages 420–434. Springer Berlin Heidelberg, Berlin, Heidelberg.

Akbarieh, A., Osadcha, I., Farghaly, K., Bosché, F., and Puust, R. (2025). Development of a built environment ontology lookup service (be-ols) with a new ontology evaluation. In Proceedings of the 2025 European Conference on Computing in Construction, volume 6 of Computing in Construction, Porto, Portugal. European Council on Computing in Construction.

Alshargi, F., Shekarpour, S., Soru, T., and Sheth, A. P. (2019). Metrics for evaluating quality of embeddings for ontological concepts. In AAAI Spring Symposium Combining Machine Learning with Knowledge Engineering.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When Is "Nearest Neighbor" Meaningful? In Goos, G., Hartmanis, J., Van Leeuwen, J., Beeri, C., and Buneman, P., editors, Database Theory — ICDT'99, volume 1540, pages 217–235. Springer Berlin Heidelberg, Berlin, Heidelberg.

Boggust, A., Carter, B., and Satyanarayan, A. (2022). Embedding Comparator: Visualizing Differences in Global Structure and Local Neighborhoods via Small Multiples. In 27th International Conference on Intelligent User Interfaces, pages 746–766, Helsinki Finland. ACM.

Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M. C., and Gómez-Pérez, A. (2019). Why are ontologies not reused across the same domain? Journal of Web Semantics, 57:100492.

Giaffar, H., Rullán Buxó, C., and Aoi, M. (2024). The effective number of shared dimensions between paired datasets. In Dasgupta, S., Mandt, S., and Li, Y., editors, Proceedings of the 27th International Conference on Artificial Intelligence and Statistics, volume 238 of Proceedings of Machine Learning Research, pages 4249–4257. PMLR.

Hopkins, Brian. and Skellam, J. G. (1954). A New Method for determining the Type of Distribution of Plant Individuals. Annals of Botany, 18(70):213–227.

Khan, A., Shah, Q., Uddin, M. I., Ullah, F., Alharbi, A., Alyami, H., and Gul, M. A. (2020). Sentence embedding based semantic clustering approach for discussion thread summarization. Complexity, 2020(1):4750871.

Maratsi, M. I., Alexopoulos, C., and Charalabidis, Y. (2023). A structured analysis of domain-specific linked open vocabularies (lov): Indicators for interoperability and reusability. In European, Mediterranean, and Middle Eastern Conference on Information Systems, pages 135–152. Springer.

Radovanović, M., Nanopoulos, A., and Ivanović, M. (2009). Nearest neighbors in high-dimensional data: The emergence and influence of hubs. In Proceedings of the 26th Annual International Conference on Machine Learning, pages 865–872, Montreal Quebec Canada. ACM.

Recanatesi, S., Bradde, S., Balasubramanian, V., Steinmetz, N. A., and Shea-Brown, E. (2022). A scale-dependent measure of system dimensionality. Patterns, 3(8):100555.

Reimers, N. and Gurevych, I. (2024). all-minilm-l6-v2: Sentence-transformers model. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2. Accessed: 2025-11-26.

Ristoski, P. and Paulheim, H. (2016). RDF2Vec: RDF Graph Embeddings for Data Mining. In Groth, P., Simperl, E., Gray, A., Sabou, M., Krötzsch, M., Lecue, F., Flöck, F., and Gil, Y., editors, The Semantic Web – ISWC 2016, volume 9981, pages 498–514. Springer International Publishing, Cham.

Tao, C., Shen, T., Gao, S., Zhang, J., Li, Z., Hua, K., Hu, W., Tao, Z., and Ma, S. (2024). Llms are also effective embedding models: An in-depth overview. arXiv preprint arXiv:2412.12591. revised 2025.

Teymurova, S., Jiménez-Ruiz, E., Weyde, T., and Chen, J. (2024). Owl2vec4oa: Tailoring knowledge graph embeddings for ontology alignment. DOI: 10.48550/arXiv.2408.06310.

Wang, P., Xu, J., Xu, B., Liu, C., Zhang, H., Wang, F., and Hao, H. (2015). Semantic clustering and convolutional neural network for short text categorization. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 352–357.

Yu, P., Merrick, L., Nuti, G., and Campos, D. (2024). Arctic-embed 2.0: Multilingual retrieval without compromise.