ELK-Em: Closure-Aware Embeddings for $\mathcal{EL}^+_\bot$ Ontologies

Naman Singh, Ernesto Jiménez-Ruiz and Tillman Weyde

City St George's, University of London, London, UK

Abstract

$\mathcal{EL}^+_\bot$ is a lightweight description logic designed to capture the expressive requirements of large practical ontologies while supporting tractable reasoning, as implemented by reasoners such as ELK. Recent efforts have developed scalable methods for learning continuous representations of such ontologies for inductive inference, but their coverage of the normalised axiom forms varies, they provide no formal guarantees relative to an inference calculus, and they cannot place individuals with no observed assertions. We present ELK-Em, a novel geometric ontology embedding model that embeds *(i)* concepts as axis-aligned boxes; *(ii)* roles with a new parametrisation that allows coverage of every normalised $\mathcal{EL}^+_\bot$ axiom form, yielding, at zero loss, a model of the ontology and satisfaction of all of ELK's inference rules, with role chains enforced by a stronger sufficient condition; and *(iii)* individuals as points regularised using domain-specific encoders. On GALEN, GO, and Anatomy, ELK-Em ranks the atomic subsumptions ELK derives above non-entailed candidates more accurately than geometric baselines. On CAFA5 human protein-function data it also predicts GO concept assertions for held-out proteins from ESM-2 embeddings; nearest-neighbour transfer scores higher, a gap we locate in box-membership scoring rather than placement. Code and a technical report with full proofs are available at our repository.

Keywords

Ontology Embeddings, $\mathcal{EL}$ Embeddings, Biomedical Applications

1. Introduction

The $\mathcal{EL}$ family of description logics [1] underlies OWL 2 EL [2], designed for ontologies with large numbers of classes and relations while retaining polynomial-time reasoning for core tasks. This makes it suitable for large biomedical ontologies such as the Gene Ontology (GO) [3], SNOMED CT [4] and GALEN [5]. In these logics, domain knowledge is expressed through axioms, divided into an ABox of instance-level assertions such as $Parent(Alex)$ or $hasChild(Alex, Alice)$, and a TBox of concept-level axioms such as $Parent \sqsubseteq Person$.

Reasoners such as ELK [6] perform consistency checking, subsumption and instance checking over large ontologies: the axioms above entail $Person(Alex)$, which ELK derives automatically, letting applications draw on implicit as well as explicit knowledge. Such reasoning is nonetheless limited to the symbolic input. In the absence of ABox assertions for an individual, a reasoner cannot derive its class membership, even if the individual carries labels, comments or external descriptions.

Ontology embedding methods provide one way to support inductive inference over OWL ontologies [7]. Onto2Vec [8], OPA2Vec [9] and OWL2Vec* [10] learn vector representations from ontology-derived corpora, but treat ontologies as graph- or corpus-based inputs and do not impose description-logic semantics as constraints during training. More recent works [11, 12, 13, 14, 15, 16] instead represent concepts, roles and individuals as geometric objects and encode axioms as constraints in the embedding space, tying the representation more closely to description-logic reasoning. Three limitations remain:

- They learn individual embeddings only from observed ABox facts, and so cannot place an individual with no assertions, whatever labels, annotations or domain-specific encodings it carries. This limits applications such as CAFA [17], SNOMED CT entity linking [18] and phenotype annotation [19, 20].

Third Workshop on Knowledge Graphs and Neurosymbolic AI (KG-NeSy 2026), in ISWC 2026 Workshops Joint Proceedings, October 25–26, 2026, Bari, Italy

naman.singh@citystgeorges.ac.uk (N. Singh); ernesto.jimenez-ruiz@citystgeorges.ac.uk (E. Jiménez-Ruiz); t.e.weyde@citystgeorges.ac.uk (T. Weyde)

0009-0000-3659-0266 (N. Singh); 0000-0002-9083-4599 (E. Jiménez-Ruiz); 0000-0001-8028-9905 (T. Weyde)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-05

CEUR-WS.org/Vol-4313/kgnesy-paper14.pdf

- Their geometric designs may impose semantic over-restrictions. As Bourgaux et al. [21] highlight, models might approximate $\exists r.C \sqsubseteq \bot$ as $C \sqsubseteq \bot$, which is not a valid inference and incorrectly penalises every candidate assertion $C(a)$.
- Coverage of the normalised axiom forms varies, with role inclusion and composition treated partially or not at all. Since ABox scores depend on the learned TBox geometry, consequences the geometry does not reflect may lead to poorly scored assertions for individuals.

To address these limitations, we present ELK-Em, a geometric ontology embedding model for $\mathcal{EL}^+_\bot$, the fragment ELK supports, representing concepts as axis-aligned boxes, roles through a parametrisation that allows coverage of all normalised axiom forms, and individuals as points regularised using domain-specific encoders. Boxes are used because, unlike balls, they are closed under intersection [13]. The role parametrisation yields the central guarantee: if every axiom has zero geometric loss, the learned embedding is a model of the ontology, and therefore satisfies every consequence ELK derives from it. Role chains are enforced by a sufficient condition we characterise exactly (Section 4.1). The model is thus a geometric encoding of the logic ELK reasons over, its constructions mirroring ELK's inference rules; the name reflects that target, and the reasoner is not invoked during training.

On GALEN [5], GO [3] and Anatomy (Uberon) [22], ELK-Em ranks the atomic subsumptions ELK derives above non-entailed candidates more accurately than geometric baselines. Using ESM-2 representations [23], it also predicts GO concept assertions for held-out human proteins on CAFA5-derived data [17], whose annotations are withheld during training; we compare against direct sequence-similarity transfer and analyse the results.

The rest of the paper is organised as follows. Section 2 reviews related work, Section 3 the background on $\mathcal{EL}^+_\bot$, and Section 4 presents ELK-Em: its geometric semantics, training, theoretical properties and lexical regularisation. Section 5 evaluates it on ranking entailed subsumptions and on inductive ABox inference, and Section 6 concludes.

2. Related Work

Knowledge graph embedding methods preceded much OWL ontology embedding work and inspired representation learning for structured knowledge [24]. TransE, for example, embeds a triple (h, r, t) so that $\mathbf{h} + \mathbf{r}$ is close to $\mathbf{t}$ for observed triples and distant for corrupted ones [25]. Later work predicts links for entities unseen during training by encoding textual descriptions [26] or relation-level features [27]. All target ABox-level facts, whereas OWL ontologies also contain TBox axioms with class hierarchies, logical constructors and role inclusions.

Ontology embedding methods address this setting [7]. Onto2Vec [8], OPA2Vec [9] and OWL2Vec* [10] learn representations from ontology-derived corpora, built from reasoner-expanded axioms, annotations and lexical metadata, and random walks over the ontology graph. Axioms and constructors therefore serve as context rather than as constraints on the learned representation. These methods also rely on ontology-internal lexical metadata, which is not always the most informative signal for comparing individuals; domain-specific encoders such as ESM-2 for protein sequences [23] can supply a stronger external signal, which we exploit in Section 4.4.

A second line of work develops geometric models for the $\mathcal{EL}$ family. ELEm uses n-balls for concepts and translations for roles [11], and EmEL^{++} extends this to further constructs [12]. ELBE uses axis-aligned boxes with translational roles, boxes being closed under intersection [13], while BoxEL models roles as affine transformations [14]. Box^{2}EL and TransBox further enrich roles and complex concepts [15, 16], the latter targeting closure under the constructors of the logic. Coverage of the normalised axiom forms varies across these models, with role inclusion, role composition and unsatisfiable existential restrictions treated partially or not at all. They are closer to the model-theoretic structure of the logic than corpus-based embeddings, but establish no guarantees relative to the ELK inference calculus, and individual placement remains tied to observed ABox facts. ELK-Em addresses this gap: geometric modelling of the normalised axiom forms of Section 3 with guarantees relative to ELK inference, and encoder-based placement of individuals with no observed assertions.

3. Background: The Description Logic $\mathcal{EL}^+_\bot$

$\mathcal{EL}^+_\bot$ is the tractable description logic underlying the ELK reasoner [6], closely related to the logical core of the OWL 2 EL profile [2]. Its signature consists of sets N_C, N_R and N_I of *concept*, *role* and *individual names*, and concept descriptions are built with the constructors of Table 1. An ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ consists of a TBox $\mathcal{T}$ of *general concept inclusions* (GCIs) and *role inclusions* (RIs), and an ABox $\mathcal{A}$ of *concept assertions* $C(a)$ and *role assertions* $r(a, b)$. Its members are the *axioms* of $\mathcal{O}$.

Table 1
Syntax and semantics of the $\mathcal{EL}^+_\bot$ constructors and axioms.

Name	Syntax	Semantics
Top	$\top$	$\Delta^\mathcal{I}$
Bottom	$\bot$	$\emptyset$
Concept name	A	$A^\mathcal{I} \subseteq \Delta^\mathcal{I}$
Conjunction	$C \sqcap D$	$C^\mathcal{I} \cap D^\mathcal{I}$
Existential rest.	$\exists r.C$	$\{x \mid \exists y : (x, y) \in r^\mathcal{I} \wedge y \in C^\mathcal{I}\}$
GCI	$C \sqsubseteq D$	$C^\mathcal{I} \subseteq D^\mathcal{I}$
RI	$r_1 \circ \cdots \circ r_k \sqsubseteq s$	$r_1^\mathcal{I} \circ \cdots \circ r_k^\mathcal{I} \subseteq s^\mathcal{I}$
Concept assertion	$C(a)$	$a^\mathcal{I} \in C^\mathcal{I}$
Role assertion	$r(a, b)$	$(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$

An interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consists of a non-empty domain and a function mapping each $a \in \mathsf{N}_I$ to an element of $\Delta^\mathcal{I}$, each $A \in \mathsf{N}_C$ to a subset, and each $r \in \mathsf{N}_R$ to a binary relation, extended to complex concepts as in Table 1. $\mathcal{I}$ *satisfies* α, written $\mathcal{I} \models \alpha$, if the corresponding condition holds, and is a *model* of $\mathcal{O}$ if it satisfies every axiom of $\mathcal{O}$. An axiom α is *entailed* by $\mathcal{O}$, written $\mathcal{O} \models \alpha$, if every model of $\mathcal{O}$ satisfies it.

Normal form. A *normalised* ontology contains only axioms of the forms $A \sqsubseteq B$ (GCI0), $A \sqcap B \sqsubseteq D$ (GCI1), $A \sqsubseteq \exists r.B$ (GCI2), $\exists r.A \sqsubseteq B$ (GCI3), $r \sqsubseteq s$ (RI0) and $r_1 \circ r_2 \sqsubseteq s$ (RI1), where $A, B, D \in \mathsf{N}_C$ and $r, r_1, r_2, s \in \mathsf{N}_R$ [1], the variants GCI0-BOT, GCI1-BOT and GCI3-BOT in which the right-hand side is $\bot$, and atomic assertions $A(a)$ and $r(a, b)$. Every $\mathcal{EL}^+_\bot$ ontology can be transformed in linear time into a normalised ontology that is a conservative extension of it [28].

Classification with ELK. ELK *classifies* an ontology by computing all entailed subsumptions $A \sqsubseteq B$ between named concepts (form GCI0), using the sound and complete inference rules of Figure 1.

$$R_0 \ \frac{}{C \sqsubseteq C} \qquad R_\top \ \frac{}{C \sqsubseteq \top}$$

$$R_\bot \ \frac{E \sqsubseteq \exists r.C \quad C \sqsubseteq \bot}{E \sqsubseteq \bot} \qquad R^-_\sqcap \ \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}$$

$$R^+_\sqcap \ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \qquad R_\exists \ \frac{E \sqsubseteq \exists r.C \quad C \sqsubseteq D}{E \sqsubseteq \exists r.D}$$

$$R_\sqsubseteq \ \frac{C \sqsubseteq D}{C \sqsubseteq E} \ : \ D \sqsubseteq E \in \mathcal{O} \qquad R_H \ \frac{E \sqsubseteq \exists r.C}{E \sqsubseteq \exists s.C} \ : \ r \sqsubseteq s \in \mathcal{O}$$

$$R_\circ \ \frac{E \sqsubseteq \exists r_1.C \quad C \sqsubseteq \exists r_2.D}{E \sqsubseteq \exists s.D} \ : \ r_1 \circ r_2 \sqsubseteq s \in \mathcal{O}$$

Figure 1: Inference rules for classification in $\mathcal{EL}^+_\bot$, following [6]. Expressions above the line are (derived) premises, those below are derived conclusions, and expressions after the colon are side conditions, i.e. fixed axioms of the input ontology whose presence fires the rule.

4. Method

ELK-Em accepts any OWL ontology as input but preserves only $\mathcal{EL}^+_\bot$ TBox axioms, which are normalised before the geometric embeddings are computed. ABox axioms may be given together with the TBox or in independent files, and the geometric embedding can be combined with a lexical embedding, where available, to improve the placement of unseen instances. We apply ELK-Em to ranking the atomic subsumptions ELK derives (Section 5.1) and to class-assertion prediction (Section 5.2). The remainder of this section presents the geometric design, training procedure, scoring, formal guarantees and lexical regularisation. Proofs and a more detailed exposition are given in the technical report [29].

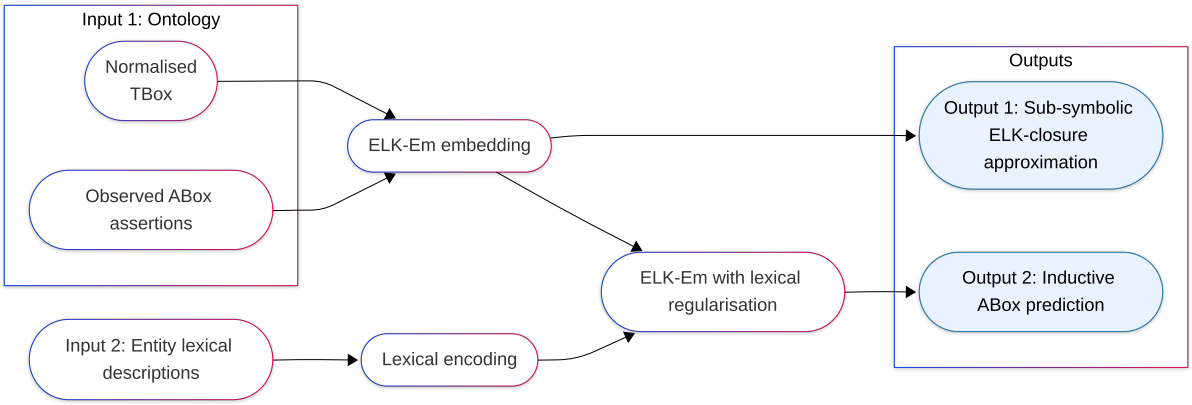


Figure 2: Overview of the ELK-Em pipeline.

4.1. Geometric Interpretation

An *axis-aligned box* in $\mathbb{R}^n$ is a set $[\mathbf{l}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{l} \leqslant \mathbf{x} \leqslant \mathbf{u}\}$, with the inequalities taken component-wise, and is empty if $u_i < l_i$ for some i. We write $[-\infty, \infty]$ for $\mathbb{R}^n$, $[\infty, -\infty]$ for the canonical empty box, $[\mathbf{l}_x, \mathbf{u}_x] \cap [\mathbf{l}_y, \mathbf{u}_y] = [\max(\mathbf{l}_x, \mathbf{l}_y), \min(\mathbf{u}_x, \mathbf{u}_y)]$ for box intersection, and $[\mathbf{l}_x, \mathbf{u}_x] \oplus [\mathbf{l}_y, \mathbf{u}_y] = [\mathbf{l}_x + \mathbf{l}_y, \mathbf{u}_x + \mathbf{u}_y]$ for box addition.

An ELK-Em embedding is read as an interpretation $\mathcal{E} = (\mathbb{R}^n, \cdot^\mathcal{E})$ assigning

- to each $A \in \mathsf{N}_C$ a box $A^\mathcal{E} = B_A = [\mathbf{l}_A, \mathbf{u}_A]$,
- to each $a \in \mathsf{N}_I$ a point $a^\mathcal{E} = \mathbf{e}_a$ (or equivalently a box $[\mathbf{e}_a, \mathbf{e}_a]$),
- to each $r \in \mathsf{N}_R$ a triple $\langle R_r, E_r, T_r \rangle$ with non-empty range box $R_r = [\mathbf{l}^R_r, \mathbf{u}^R_r]$, non-empty error box $E_r = [\mathbf{l}^E_r, \mathbf{u}^E_r]$, and diagonal affine map $T_r(\mathbf{x}) = \mathbf{m}_r\mathbf{x} + \mathbf{c}_r$ ($\mathbf{m}_r \in \mathbb{R}^n_{>0}, \mathbf{c}_r \in \mathbb{R}^n$), giving, with $T_r(\mathbf{y})$ read as the singleton box $[T_r(\mathbf{y}), T_r(\mathbf{y})]$, the relation

$$r^\mathcal{E} = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in T_r(\mathbf{y}) \oplus E_r \wedge \mathbf{y} \in R_r\}.$$

The interpretation extends to complex concepts following Table 1: $\top$ and $\bot$ are the full and empty boxes, $(C \sqcap D)^\mathcal{E} = C^\mathcal{E} \cap D^\mathcal{E}$, and $(\exists r.C)^\mathcal{E} = T_r(C^\mathcal{E} \cap R_r) \oplus E_r$, written $B_{\exists r.C}$, the points related by r to a point in both $C^\mathcal{E}$ and R_r. Since T_r has positive slope and intersection and $\oplus$ preserve boxes, induction on C shows that every concept description is interpreted as an axis-aligned box, which we write B_C.

The interpretation determines when axioms hold in the embedding space (see Figure 3): $C \sqsubseteq D$ holds when $C^\mathcal{E} \subseteq D^\mathcal{E}$, i.e. $\mathbf{l}_D \leqslant \mathbf{l}_C \wedge \mathbf{u}_C \leqslant \mathbf{u}_D$; $C(a)$ holds when $\mathbf{l}_C \leqslant \mathbf{e}_a \leqslant \mathbf{u}_C$; and $r(a, b)$ holds when $(a^\mathcal{E}, b^\mathcal{E}) \in r^\mathcal{E}$. Each inference rule of Figure 1 then has a counterpart at the level of boxes, established in the propositions below and combined in Theorem 6, beginning with existential restriction, which is monotone in its filler.

Proposition 1. *(Rule $R_\exists$.) For any role r and concept descriptions C, D, if $C^\mathcal{E} \subseteq D^\mathcal{E}$ then $(\exists r.C)^\mathcal{E} \subseteq (\exists r.D)^\mathcal{E}$.*

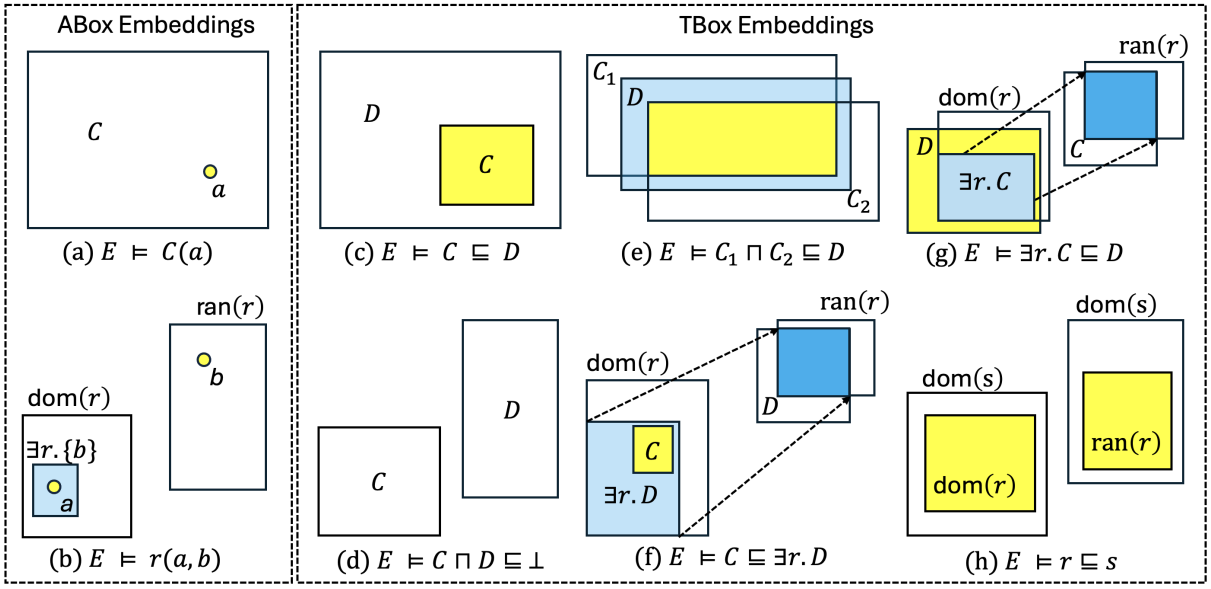


Figure 3: Geometric semantics of ABox axioms (left) and TBox axioms (right) in $\mathcal{EL}^+_\bot$ under ELK-Em embeddings.

A role inclusion $r \sqsubseteq s$ holds when $r^{\mathcal{E}} \subseteq s^{\mathcal{E}}$, a containment reducible to the role parameters.

Proposition 2. *For $r, s \in \mathsf{N}_R$ the following are equivalent:* (1) $r^{\mathcal{E}} \subseteq s^{\mathcal{E}}$; (2) $R_r \subseteq R_s$ *and* $T_r(\mathbf{y}) \oplus E_r \subseteq T_s(\mathbf{y}) \oplus E_s$ *for all* $\mathbf{y} \in R_r$; (3) *the same for* $\mathbf{y} \in \{\mathbf{l}^R_r, \mathbf{u}^R_r\}$.

Proposition 3. *If $r^{\mathcal{E}} \subseteq s^{\mathcal{E}}$, then* (1) $(a^{\mathcal{E}}, b^{\mathcal{E}}) \in r^{\mathcal{E}}$ *implies* $(a^{\mathcal{E}}, b^{\mathcal{E}}) \in s^{\mathcal{E}}$; *and* (2) *(Rule R_H)* $(\exists r.C)^{\mathcal{E}} \subseteq (\exists s.C)^{\mathcal{E}}$ *for every concept C.*

For a chain axiom $r_1 \circ r_2 \sqsubseteq s$, satisfaction is containment of the composite $r_1^{\mathcal{E}} \circ r_2^{\mathcal{E}} = \{(\mathbf{x}, \mathbf{z}) \mid \exists \mathbf{y} : (\mathbf{x}, \mathbf{y}) \in r_1^{\mathcal{E}} \wedge (\mathbf{y}, \mathbf{z}) \in r_2^{\mathcal{E}}\}$ in $s^{\mathcal{E}}$.

Proposition 4. *Let* $\max(r_1 \circ r_2) = \langle R_{r_2}, \mathbf{m}_{r_1} E_{r_2} \oplus E_{r_1}, T_{r_1} \circ T_{r_2} \rangle$[1] *and let* $T_{r_2}(R_{r_2}) \oplus E_{r_2} \subseteq R_{r_1}$ *be the* compatibility condition, *which places the domain* $(\exists r_2.\top)^{\mathcal{E}}$ *of* r_2 *within the range* R_{r_1} *of* r_1. *Then* (1) $r_1^{\mathcal{E}} \circ r_2^{\mathcal{E}} \subseteq \max(r_1 \circ r_2)^{\mathcal{E}}$; (2) *equality holds iff the compatibility condition holds;* (3) *if the compatibility condition fails,* $r_1^{\mathcal{E}} \circ r_2^{\mathcal{E}}$ *is not of the form* $\langle R, E, T \rangle$.

Proposition 5. *If* $\max(r_1 \circ r_2)^{\mathcal{E}} \subseteq s^{\mathcal{E}}$, *then* (1) $\mathcal{E} \models r_1 \circ r_2 \sqsubseteq s$; *and* (2) *(Rule $R_\circ$) if* $X^{\mathcal{E}} \subseteq (\exists r_1.C)^{\mathcal{E}}$ *and* $C^{\mathcal{E}} \subseteq (\exists r_2.D)^{\mathcal{E}}$, *then* $X^{\mathcal{E}} \subseteq (\exists s.D)^{\mathcal{E}}$.

The true composite is in general not of the form $\langle R, E, T \rangle$ (Prop. 4.3), so Proposition 2 does not apply to it and containment in s cannot be reduced to a check on parameters. We therefore work with the over-approximation $\max(r_1 \circ r_2)$, which is of that form, and enforce $\max(r_1 \circ r_2)^{\mathcal{E}} \subseteq s^{\mathcal{E}}$ in place of the chain axiom. By Proposition 5 this is sufficient, entailing both $\mathcal{E} \models r_1 \circ r_2 \sqsubseteq s$ and rule $R_\circ$, though it is strictly stronger than the axiom where the over-approximation is strict (Prop. 4.2). The technical report [29] also gives an alternative under-approximation for which $R_\circ$ fails.

Satisfaction. An interpretation $\mathcal{E}$ *satisfies* a normalised axiom according to the conditions above: $\mathcal{E} \models C(a)$ iff $a^{\mathcal{E}} \in C^{\mathcal{E}}$, $\mathcal{E} \models r(a, b)$ iff $(a^{\mathcal{E}}, b^{\mathcal{E}}) \in r^{\mathcal{E}}$, $\mathcal{E} \models C \sqsubseteq D$ iff $C^{\mathcal{E}} \subseteq D^{\mathcal{E}}$ for the non-$\bot$ forms GCI0–GCI3, $\mathcal{E} \models r \sqsubseteq s$ iff $r^{\mathcal{E}} \subseteq s^{\mathcal{E}}$, and $\mathcal{E} \models r_1 \circ r_2 \sqsubseteq s$ iff $r_1^{\mathcal{E}} \circ r_2^{\mathcal{E}} \subseteq s^{\mathcal{E}}$, for which we enforce the sufficient condition above. The bottom forms are covered by the same inclusion condition, since $\bot^{\mathcal{E}} = \varnothing$, but are enforced through box disjointness (Section 4.2).

[1] We write $\mathbf{m}E$ for the box $[\mathbf{ml}^E,\ \mathbf{mu}^E]$.

Table 2
Loss terms for the normalised $\mathcal{EL}^+_\bot$ axiom forms. Overlap terms ($\ell_{\cap^+}$) in GCI1 and GCI3 prevent degenerate solutions; $\ell_{\mathrm{RI0}}(\max(r_1 \circ r_2), s)$ applies the RI0 loss to $\max(r_1 \circ r_2)$ of Prop. 4. Unsatisfiable concepts ($A \sqsubseteq \bot$) are fixed to the empty box rather than trained.

	Axiom	**Loss**
GCI0	$C \sqsubseteq D$	$\ell^\gamma_{\subseteq}(B_C, B_D)$
(GCI0-BOT)	$C \sqsubseteq \bot$	$-$ (fixed $B_C = [\infty, -\infty]$)
GCI1	$C_1 \sqcap C_2 \sqsubseteq D$	$\ell^\gamma_{\subseteq}(B_{C_1} \cap B_{C_2}, B_D) + \ell^\gamma_{\cap^+}(B_{C_1}, B_{C_2})$
(GCI1-BOT)	$C_1 \sqcap C_2 \sqsubseteq \bot$	$\ell^\gamma_{\cap^-}(B_{C_1}, B_{C_2})$
GCI2	$C \sqsubseteq \exists r.D$	$\ell^\gamma_{\subseteq}(B_C, B_{\exists r.D})$
GCI3	$\exists r.C \sqsubseteq D$	$\ell^\gamma_{\subseteq}(B_{\exists r.C}, B_D) + \ell^\gamma_{\cap^+}(R_r, B_C)$
(GCI3-BOT)	$\exists r.C \sqsubseteq \bot$	$\ell^\gamma_{\cap^-}(R_r, B_C)$
RI0	$r \sqsubseteq s$	$\ell^\gamma_{\subseteq}(R_r, R_s) + \sum_{\mathbf{y} \in \{\mathbf{l}^R_r, \mathbf{u}^R_r\}} \ell^\gamma_{\subseteq}\big(T_r(\mathbf{y}) \oplus E_r, T_s(\mathbf{y}) \oplus E_s\big)$
RI1	$r_1 \circ r_2 \sqsubseteq s$	$\ell_{\mathrm{RI0}}(\max(r_1 \circ r_2), s)$

4.2. Model Training

We train the embedding by minimising a geometric loss whose zero set is contained in the satisfaction conditions above. All losses are margin-based with a scalar margin $\gamma > 0$, added component-wise where the argument is a vector, and follow previous work in being distance-based [15, 16].

Loss primitives. For boxes $B_x = [\mathbf{l}_x, \mathbf{u}_x]$ and $B_y = [\mathbf{l}_y, \mathbf{u}_y]$, write centres $\mathbf{c} = (\mathbf{u} + \mathbf{l})/2$, offsets $\mathbf{o} = (\mathbf{u} - \mathbf{l})/2$, and the per-dimension separation $d(B_x, B_y) = |\mathbf{c}_x - \mathbf{c}_y| - \mathbf{o}_x - \mathbf{o}_y$, with d_i its i-th component. We use three primitives:

$$\ell^\gamma_{\subseteq}(B_x, B_y) = \tfrac{1}{n}\big\|\mathrm{ReLU}\big(d(B_x, B_y) + 2\mathbf{o}_x + \gamma\big)\big\|_1, \tag{1}$$

$$\ell^\gamma_{\cap^+}(B_x, B_y) = \tfrac{1}{n}\big\|\mathrm{ReLU}\big(d(B_x, B_y) + \gamma\big)\big\|_1, \tag{2}$$

$$\ell^\gamma_{\cap^-}(B_x, B_y) = \mathrm{ReLU}\big(\gamma - \max_i d_i(B_x, B_y)\big), \tag{3}$$

penalising violation of $B_x \subseteq B_y$, driving the boxes to overlap, and driving them disjoint respectively. Writing ℓ^0 for the margin-free version ($\gamma = 0$) of each, zero loss on a primitive implies the corresponding box condition (technical report [29]).

Per-axiom losses. Each normal form is assigned the loss in Table 2, built so that its margin-free version is zero when the corresponding satisfaction condition holds. GCI0-BOT carries no loss, as B_A is fixed to the empty box. For GCI3-BOT the disjoint primitive enforces $B_C \cap R_r = \varnothing$, hence $(\exists r.C)^{\mathcal{E}} = \varnothing$, without forcing $B_C = \varnothing$, avoiding the over-restriction in prior box models that approximate this form as $C \sqsubseteq \bot$ [21]. The TBox loss $\mathcal{L}_T = \sum_{\alpha \in \mathcal{T}} \ell(\alpha)$ sums these over all normalised axioms, optimised over mini-batches.

Scoring. At evaluation time we set $\gamma = 0$ in every primitive occurring in a per-axiom loss ℓ, writing ℓ^0 for the result, and score an axiom by

$$s(\alpha) \;=\; \exp\big(-\ell^0(\alpha)\big) \in (0, 1], \tag{4}$$

which is 1 exactly when the enforced geometric condition holds and decreases smoothly with the degree of violation. This is the quantity ranked in Section 5.1.

Negative sampling. To avoid degenerate solutions we contrast each positive axiom with corrupted ones. For a positive GCI0 axiom $C \sqsubseteq D$ we generate $k_{\mathrm{neg}} \geqslant 1$ negatives by replacing D with a uniformly sampled concept name, without filtering corruptions that happen to be entailed; for GCI2 we corrupt C or D with equal probability, keeping r fixed. Collecting the margin-free scores of Eq. (4) in $\boldsymbol{\sigma} \in \mathbb{R}^N$ and the binary labels in $\mathbf{t} \in \{0, 1\}^N$, with sample means $\bar{\sigma}, \bar{t}$, we use a Pearson correlation loss

$$\mathcal{L}_{\mathrm{contrast}} \;=\; 1 - \frac{(\boldsymbol{\sigma} - \bar{\sigma})^\top (\mathbf{t} - \bar{t})}{\|\boldsymbol{\sigma} - \bar{\sigma}\|\,\|\mathbf{t} - \bar{t}\| + \varepsilon}, \tag{5}$$

where $\varepsilon > 0$ prevents division by zero. This separates positives from corruptions without an additional margin hyperparameter. It is used only when training on TBox data alone (Section 5.1), where axiom scores are the ranked quantity; with ABox supervision the assertion losses below already supply a contrastive signal.

ABox training. Concept assertions are trained as a special case of GCI0, an individual being embedded as the degenerate box $[\mathbf{e}_a, \mathbf{e}_a]$, which reduces membership to box inclusion. Since ontologies typically provide only positive assertions, we generate negatives by corrupting the class of a positive assertion, filtering out observed training positives. With $\ell_{C(a)} = \ell^{\gamma}_{\subseteq}([\mathbf{e}_a, \mathbf{e}_a], B_C)$ for a positive assertion and $\mathrm{ReLU}(\delta - \ell_{C(a)})$ for a corrupted one, where $\delta = -\log \tau$ for a target-score threshold $\tau \in (0, 1)$, the ABox loss sums both terms over the sets $\mathcal{A}^+$ and $\mathcal{A}^-$ of observed and corrupted assertions. Role assertions admit the same treatment, since $\mathcal{E} \models r(a, b)$ is an inclusion between $[\mathbf{e}_a, \mathbf{e}_a]$ and $T_r([\mathbf{e}_b, \mathbf{e}_b]) \oplus E_r$ together with $\mathbf{e}_b \in R_r$, both expressible with $\ell^{\gamma}_{\subseteq}$. We restrict supervision to concept assertions here, matching the inductive task of Section 5.2; role-assertion supervision, such as training on protein–protein interaction data alongside functional annotations, is left to future work.

4.3. Theoretical Properties

Bourgaux et al. [21] show that, among the geometric embedding methods they survey, BoxEL is the only one with a soundness guarantee [14], and it does not cover role inclusion or composition; to our knowledge no prior method relates the learned embedding to closure under ELK's inference rules. ELK-Em provides both, for all normalised forms and all rules of Figure 1. Proofs and complexity are in the technical report [29].

Theorem 6. *Let $\mathcal{O}$ be a normalised ontology, and let $\mathcal{E}$ be an embedding such that every TBox axiom and every positive ABox assertion in $\mathcal{O}$ has zero geometric loss. Then $\mathcal{E}$ is a model of $\mathcal{O}$.*

Corollary 7. *Under the assumptions of Theorem 6, $\mathcal{E} \models \alpha$ for every axiom α derivable from $\mathcal{O}$ by the rules of Figure 1.*

The question of completeness is open: the RI1 condition is sufficient but not necessary, so not every consistent ontology need admit a zero-loss embedding.

4.4. Lexical Regularisation

Many prediction settings involve individuals that lack ABox assertions but carry informative non-logical descriptions, which are the primary signal for prediction: CAFA predicts GO terms for proteins from sequence data [17], the SNOMED CT Entity Linking Challenge assigns concepts to clinical discharge summaries [18], and rare-disease resources such as CADA and PhEval support phenotype-driven prediction from clinical descriptions [19, 20]. Across these settings, external descriptions carry information that can position a new individual relative to a curated TBox.

We use lexical regularisation to transfer this information into the embedding space; we take *lexical* to cover any textual description of an individual, such as an amino acid sequence or a clinical narrative, rather than only the ontology labels and annotations the term often denotes. Let $a_1, \ldots, a_m$ be a batch of individuals, $\mathbf{z}_i \in \mathbb{R}^d$ an external representation of a_i available at prediction time, produced by a task-specific encoder such as ESM-2 for protein sequences [23] or a Sentence-BERT-style encoder for clinical text [30], and $\mathbf{e}_i \in \mathbb{R}^n$ its learned point embedding. Writing $\bar{s}^{\mathrm{ext}}_{ij}$ for the cosine similarity of the external representations rescaled to $[0, 1]$, and s^{geo}_{ij} for a geometric similarity decaying with embedding distance, the regularisation loss aligns the two:

$$\bar{s}^{\mathrm{ext}}_{ij} = \frac{1}{2}\Big(1 + \frac{\mathbf{z}_i^{\top}\mathbf{z}_j}{\|\mathbf{z}_i\|_2\|\mathbf{z}_j\|_2}\Big), \quad s^{\mathrm{geo}}_{ij} = \exp\Big(-\frac{\lambda}{n}\|\mathbf{e}_i - \mathbf{e}_j\|_2^2\Big), \quad \mathcal{L}_{\mathrm{lex}} = \frac{\sum_{i<j}\big(\bar{s}^{\mathrm{ext}}_{ij} - s^{\mathrm{geo}}_{ij}\big)^2}{\binom{m}{2}}, \tag{6}$$

with $\lambda > 0$ controlling the decay rate and n normalising by the embedding dimension. Cosine similarity is insensitive to the magnitudes of the external representations, so individuals with similar descriptions

are placed close together. For individuals with ABox assertions the term acts as an auxiliary regulariser; for inductive individuals with no observed assertions it provides the sole placement signal.

5. Evaluation

5.1. Ranking Entailed Atomic Subsumptions

We evaluate how well the axiom scores of Eq. (4) separate the entailed atomic subsumptions $A \sqsubseteq B$ (form GCI0) derived by ELK from non-entailed candidates, following the setup of Box^{2}EL [15]. The task measures *entailment fidelity*: how far the geometry aligns with the entailment relation, not whether it reproduces classification, which a reasoner already performs exactly.

Datasets and baselines. We use the benchmark and protocol of Box^{2}EL [15] over the biomedical ontologies GALEN, GO and Anatomy (Uberon), with class and role signatures 23,141/950, 45,893/8 and 106,258/166 respectively. Entailed subsumptions are computed with ELK [6]; after removing those already asserted in the training axioms, the benchmark retains 18,138 for GALEN, 3,309 for GO and 2,130 for Anatomy, split 90/10 into test and validation. Baseline results for ELEm [11], EmEL^{++} [12], BoxEL [14], ELBE [13] and Box^{2}EL are reproduced from Table 6 of [15], who retrain all five under this protocol rather than adopting published numbers. TransBox [16] does not report results under this protocol and is not included.

Protocol. For each held-out subsumption $C \sqsubseteq D$ the query concept C is fixed and the entire concept signature ranked in descending order of subsumption score. ELK-Em ranks by $s(\alpha)$ of Eq. (4); the baselines rank by their own GCI0 scoring functions, based on the distance between concept-box centres. We report Mean Rank (MR), Mean Reciprocal Rank (MRR), Median Rank (Med), AUC and Hits@k for $k \in \{1, 10, 100\}$ under the unfiltered protocol of [15], retained for comparability with the published baselines. ELK-Em uses a single configuration across all three ontologies: $n = 50$, Adam, learning rate 0.01 decayed by 0.1 after 5,000 epochs, margin $\gamma = 0.1$, batch size 1,024, 10,000 epochs, with only the number of corruptions per positive differing ($k_{\text{neg}} = 5$ for GO, 3 for GALEN and Anatomy).

Table 3
Ranking of entailed atomic subsumptions on GALEN, GO, and Anatomy. Baseline results are from Table 6 of [15] and use $n = 200$ ($n = 25$ for BoxEL) with per-dataset hyperparameter tuning; ELK-Em uses $n = 50$ and a single configuration across all three ontologies. Both ELK-Em and the baselines are means over five random seeds; deviations are shown for ELK-Em only, as [15] reports means alone. Deviations below 0.005 display as 0.00. Best value per column and ontology in bold, with ties at the displayed precision bolded jointly.

	Model	H@1 ↑	H@10 ↑	H@100 ↑	Med ↓	MRR ↑	MR ↓	AUC ↑
GALEN	ELEm	0.00	0.04	0.20	1807	0.01	4405	0.81
	EmEL^{++}	0.00	0.04	0.18	2049	0.01	4634	0.81
	BoxEL	0.00	0.00	0.01	6906	0.00	7925	0.67
	ELBE	0.00	0.06	0.16	1785	0.02	3974	0.84
	Box^{2}EL	**0.01**	**0.09**	0.24	1003	**0.03**	2833	0.88
	ELK-Em	0.00$\pm$0.00	0.08$\pm$0.00	**0.35**$\pm$0.00	**226**$\pm$3	**0.03**$\pm$0.00	**2323**$\pm$30	**0.90**$\pm$0.00
GO	ELEm	0.00	0.04	0.22	1629	0.02	7377	0.84
	EmEL^{++}	0.00	0.04	0.19	1346	0.01	6557	0.86
	BoxEL	0.00	0.00	0.13	1085	0.00	5359	0.88
	ELBE	0.00	0.06	0.21	935	0.02	3846	0.92
	Box^{2}EL	0.00	0.08	0.49	107	**0.04**	1689	0.96
	ELK-Em	0.00$\pm$0.00	**0.10**$\pm$0.01	**0.71**$\pm$0.01	**43**$\pm$1	**0.04**$\pm$0.00	**529**$\pm$51	**0.99**$\pm$0.00
Anatomy	ELEm	0.00	0.07	0.28	901	0.02	7958	0.93
	EmEL^{++}	0.00	0.07	0.26	1576	0.02	10976	0.90
	BoxEL	0.01	0.10	0.24	838	0.04	9156	0.92
	ELBE	0.00	0.08	0.32	336	0.03	2312	0.98
	Box^{2}EL	0.01	0.09	0.44	152	0.04	1599	**0.99**
	ELK-Em	**0.03**$\pm$0.00	**0.20**$\pm$0.01	**0.50**$\pm$0.01	**104**$\pm$3	**0.08**$\pm$0.00	**1368**$\pm$95	**0.99**$\pm$0.00

Checkpoints are selected by median rank on the validation split. Training takes 6–7 minutes per run on a single A100 with 1.7 GB peak host memory, giving models of 2.6M, 4.6M and 10.7M parameters. Results are means over five random seeds.

Results. Two features of the benchmark shape Table 3. The unfiltered protocol leaves alternative asserted or entailed superclasses in the candidate list, so the ranking space mixes false negatives, genuine non-entailments and consistent non-consequences. A single competing superclass displaces the target from the top of the ranking, which is why Hits@1 is low for all models. Role compositions are also sparse (58, 6 and 31 axioms), so the results say little about RI1 in isolation, though a single composition licenses $R_\circ$ derivations across every pair of existential restrictions it connects. With that in mind, ELK-Em improves on the geometric baselines across all three ontologies, most consistently in median and mean rank, the only exceptions being Hits@1 and Hits@10 on GALEN, using a single configuration at $n = 50$ against baselines tuned per dataset at $n = 200$, with seed variance small relative to the gaps.

5.2. Inductive Protein Function Prediction

We report preliminary results on inductive ABox prediction: assigning GO concepts to proteins that contribute no assertions during training. The aim is to test whether such individuals can be placed at all, not to offer competitive protein function prediction results. This is the setting that motivates lexical regularisation, and it lies outside the reach of a reasoner, which derives nothing for an individual with no assertions. Prior geometric models learn individual embeddings from observed ABox facts alone, so to our knowledge this setting has not been evaluated for them.

Dataset and baseline. We use a CAFA5-derived benchmark over human proteins and Gene Ontology annotations [17], comprising 25,125 proteins split by protein identity into 20,100 training, 2,512 validation and 2,513 test proteins. Validation and test proteins are held out at the individual level, so no protein–function assertion for them appears in the training ABox. Their ESM-2 representations are available during training, so the setting is inductive with respect to assertions rather than to individuals. The candidate set comprises the 43,248 GO concepts of the ontology, with no filtering by sub-ontology or annotation frequency, and external representations are 480-dimensional ESM-2 (t12, 35M) embeddings mean-pooled over residues. Sequence-similarity transfer has been a reference method in CAFA evaluations since the first challenge and remains difficult to improve on [31]; we compare against transferring annotations from the k_{nn} nearest training proteins in ESM-2 space, with no ontology embedding involved. Both methods consume the same vectors, annotations, candidate set and threshold selection, so the comparison isolates the scoring mechanism.

Protocol. We follow the protein-centric CAFA protocol [31], propagating predictions and ground truth through the GO hierarchy before scoring. Thresholds are swept over 71 values evenly spaced in $[0.3, 1.0]$, and the threshold maximising validation F_1 is applied unchanged to the test split; for the baseline, k_{nn} is also selected on validation. We therefore report F_1 at a transferred operating point rather than CAFA $F_{\max}$, which maximises over thresholds on the evaluation set itself; the values below are correspondingly not directly comparable to published $F_{\max}$ figures. ELK-Em is trained in dimension $n = 100$ for 15,000 epochs with Adam at learning rate 0.01, held constant for the first third of training and then cosine-decayed to 10^{-4}. TBox and ABox axioms are batched at 1,024, with margin $\gamma = 0.05$, two corrupted assertions per positive and target-score threshold $\tau = 0.4$. Lexical regularisation uses batches of 2,000 individuals with decay rate $\lambda = 1.0$, and the ABox positive, negative and lexical terms

Table 4
Inductive protein-function prediction, means over three seeds (± standard deviation).

Setting	Precision	Recall	F_1
ELK-Em w/o lexical regularisation	$0.0616_{\pm 0.0010}$	$0.8371_{\pm 0.0044}$	$0.1148_{\pm 0.0017}$
ELK-Em	$0.2208_{\pm 0.0143}$	$0.6135_{\pm 0.0083}$	$0.3246_{\pm 0.0165}$
ESM-2 nearest-neighbour transfer	0.6083	0.5788	0.5932

are weighted 2.0, 0.5 and 10.0, the last selected on validation from $\{1, 10, 50, 200\}$. Results are means over three random seeds.

Results. Table 4 reports the effect of lexical regularisation. Without it, held-out proteins have no evidence tying them to the training proteins: their embeddings remain close to initialisation and many are assigned broad, high-level GO concepts, giving high recall at very low precision. With it, proteins are positioned by sequence similarity to annotated training proteins and F_1 improves roughly threefold, so inductive placement is possible for individuals with no observed assertions. The baseline nonetheless reaches $F_1 = 0.59$ and outperforms ELK-Em here. The gap is not attributable to the weight on $\mathcal{L}_{\text{lex}}$, which peaks at the value used above. Nor does $\mathcal{L}_{\text{lex}}$ fail to transfer external structure, though it transfers it unevenly: rank correlation between $\bar{s}^{\text{ext}}$ and s^{geo} is 0.96 between test and training individuals, while top-k_{nn} neighbourhood overlap is only 10–27%. Global ordering is preserved, local structure is not, which accounts for part of the difference between nearest-neighbour transfer in ESM-2 space and in the learned space (0.43). The remainder lies in the score: box-membership scoring on the same embeddings reaches 0.32. It saturates, assigning 1 to every satisfied assertion, and is non-competitive, each of the 4.3×10^4 candidates being tested against its own box in isolation, so an individual may score 1 on arbitrarily many concepts where neighbour transfer distributes fixed weight among them. Both follow from membership being a predicate: what makes the embedding a model is also what makes its scores poor for ranking a large candidate set. Further diagnostics are in the technical report [29].

6. Conclusions and Future Work

We presented ELK-Em, a geometric embedding model for $\mathcal{EL}^+_\bot$ ontologies. Its parametrisation of roles allows coverage of all normalised axiom forms and yields two guarantees at zero loss: the embedding is a model of the ontology, and it satisfies every consequence derivable by ELK's inference rules, with role chains enforced through a sufficient condition on an over-approximation of the composite. On GALEN, GO and Anatomy, ELK-Em ranks the atomic subsumptions ELK derives above non-entailed candidates more accurately than geometric baselines. It also places individuals that contribute no assertions during training: on CAFA5-derived human protein-function data, lexical regularisation with ESM-2 representations improves F_1 roughly threefold, though direct sequence-similarity transfer still outperforms it, the difference lying in the scoring rather than the placement.

In the near future, we plan to *(i)* study whether closer capture of $\mathcal{EL}^+_\bot$ semantics improves general TBox prediction beyond the entailed subsumptions evaluated here, *(ii)* develop a graded score compatible with the model-theoretic semantics, *(iii)* extend lexical regularisation to concepts and roles with unseen TBox axioms, and *(iv)* design metrics that assess deductive-closure approximation more accurately, such as precision and recall against the ELK closure in place of ranking under an unfiltered protocol. Code, technical report, and resources are available under the Apache-2.0 licence at our github repository.

Acknowledgments

Ernesto was partially supported by the project GUARD. The authors thank Michael Akintunde for general support during the preparation of this work.

Declaration on Generative AI

During the preparation of this work, the authors used *Claude Opus 5* (Anthropic) in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] F. Baader, S. Brandt, C. Lutz, Pushing the EL envelope, in: Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005, p. 364–369.

[2] B. Motik, B. Cuenca Grau, I. Horrocks, Z. Wu, A. Fokoue, C. Lutz, OWL 2 Web Ontology Language Profiles, W3C Recommendation, World Wide Web Consortium, 2012. URL: https://www.w3.org/TR/owl2-profiles/.

[3] S. Aleksander, J. Balhoff, S. Carbon, et al., The Gene Ontology knowledgebase in 2026, Nucleic Acids Research 54 (2025). doi:10.1093/nar/gkaf1292.

[4] International Health Terminology Standards Development Organisation - IHTSDO, SNOMED CT, 2014. URL: http://www.ihtsdo.org/snomed-ct/.

[5] A. Rector, J. Rogers, Ontological and Practical Issues in Using a Description Logic to Represent Medical Concept Systems: Experience from GALEN, in: Reasoning Web, 2006, pp. 197–231. doi:10.1007/11837787_9.

[6] Y. Kazakov, M. Krötzsch, F. Simančík, The Incredible ELK, J. Autom. Reason. 53 (2014) 1–61. doi:10.1007/s10817-013-9296-3.

[7] J. Chen, O. Mashkova, F. Zhapa-Camacho, R. Hoehndorf, Y. He, I. Horrocks, Ontology Embedding: A Survey of Methods, Applications and Resources, IEEE Transactions on Knowledge and Data Engineering 37 (2025) 4193–4212. doi:10.1109/tkde.2025.3559023.

[8] F. Z. Smaili, X. Gao, R. Hoehndorf, Onto2Vec: joint vector-based representation of biological entities and their ontology-based annotations, Bioinformatics 34 (2018) i52–i60. doi:10.1093/bioinformatics/bty259.

[9] F. Z. Smaili, X. Gao, R. Hoehndorf, OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction, Bioinformatics 35 (2019) 2133–2140. doi:10.1093/bioinformatics/bty933.

[10] J. Chen, P. Hu, E. Jimenez-Ruiz, O. M. Holter, D. Antonyrajah, I. Horrocks, OWL2Vec*: Embedding of OWL Ontologies, Machine Learning 110 (2021) 1813–1845. doi:10.1007/s10994-021-05997-6.

[11] M. Kulmanov, W. Liu-Wei, Y. Yan, R. Hoehndorf, EL Embeddings: Geometric Construction of Models for the Description Logic EL++, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, 2019, pp. 6103–6109. doi:10.24963/ijcai.2019/845.

[12] S. Mondal, S. K. Bhatia, R. Mutharaju, EmEL++: Embeddings for EL++ Description Logic, in: AAAI Spring Symposium on Combining Machine Learning with Knowledge Engineering, 2021.

[13] X. Peng, Z. Tang, M. Kulmanov, K. Niu, R. Hoehndorf, Description Logic EL++ Embeddings with Intersectional Closure, 2022. URL: https://arxiv.org/abs/2202.14018. arXiv:2202.14018.

[14] B. Xiong, N. Potyka, T.-K. Tran, M. Nayyeri, S. Staab, Faithful Embeddings for EL++ Knowledge Bases, in: The Semantic Web – ISWC 2022, volume 13489 of *Lecture Notes in Computer Science*, Springer, Cham, 2022, pp. 22–38. doi:10.1007/978-3-031-19433-7_2.

[15] M. Jackermeier, J. Chen, I. Horrocks, Dual Box Embeddings for the Description Logic EL++, in: Proceedings of the ACM Web Conference 2024 (WWW '24), Association for Computing Machinery, New York, NY, USA, 2024, pp. 2250–2258. doi:10.1145/3589334.3645648.

[16] H. Yang, J. Chen, U. Sattler, TransBox: EL++-closed Ontology Embedding, in: Proceedings of the ACM Web Conference 2025 (WWW '25), Association for Computing Machinery, New York, NY, USA, 2025, pp. 22–34. doi:10.1145/3696410.3714672.

[17] I. Friedberg, P. Radivojac, C. D. Paolis, D. Piovesan, P. Joshi, W. Reade, A. Howard, CAFA 5 Protein Function Prediction, https://kaggle.com/competitions/cafa-5-protein-function-prediction, 2023. Kaggle.

[18] R. Davidson, W. Hardman, G. Amit, et al., SNOMED CT entity linking challenge, Journal of the American Medical Informatics Association 32 (2025) 1397–1406. doi:10.1093/jamia/ocaf104.

[19] C. Peng, S. Dieck, A. Schmid, et al., CADA: phenotype-driven gene prioritization based on a

case-enriched knowledge graph, NAR Genomics and Bioinformatics 3 (2021) lqab078. doi:10.1093/nargab/lqab078.

[20] Y. Bridges, V. Souza, K. Cortes, et al., Towards a standard benchmark for phenotype-driven variant and gene prioritisation algorithms: PhEval - Phenotypic inference Evaluation framework, BMC Bioinformatics 26 (2025). doi:10.1186/s12859-025-06105-4.

[21] C. Bourgaux, R. Guimarães, R. Koudijs, V. Lacerda, A. Ozaki, Knowledge Base Embeddings: Semantics and Theoretical Properties, in: Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning, 2024, pp. 823–833. doi:10.24963/kr.2024/77.

[22] C. J. Mungall, C. Torniai, G. V. Gkoutos, S. E. Lewis, M. A. Haendel, Uberon, an integrative multi-species anatomy ontology, Genome Biology 13 (2012) R5. doi:10.1186/gb-2012-13-1-r5.

[23] Z. Lin, H. Akin, R. Rao, et al., Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379 (2023) 1123–1130. doi:10.1126/science.ade2574.

[24] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge Graph Embedding: A Survey of Approaches and Applications, IEEE Transactions on Knowledge and Data Engineering 29 (2017) 2724–2743. doi:10.1109/TKDE.2017.2754499.

[25] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13, Curran Associates Inc., Red Hook, NY, USA, 2013, p. 2787–2795.

[26] D. Daza, M. Cochez, P. Groth, Inductive Entity Representations from Text via Link Prediction, in: Proceedings of the Web Conference 2021 (WWW '21), Association for Computing Machinery, New York, NY, USA, 2021, pp. 798–808. doi:10.1145/3442381.3450141.

[27] G. A. Gesese, H. Sack, M. Alam, RAILD: Towards Leveraging Relation Features for Inductive Link Prediction In Knowledge Graphs, in: Proceedings of the 11th International Joint Conference on Knowledge Graphs (IJCKG '22), Association for Computing Machinery, New York, NY, USA, 2022, pp. 82–90. doi:10.1145/3579051.3579066.

[28] C. Lutz, D. Walther, F. Wolter, Conservative extensions in expressive description logics, in: Proceedings of the 20th International Joint Conference on Artifical Intelligence, IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007, p. 453–458.

[29] N. Singh, E. Jiménez-Ruiz, T. Weyde, ELK-Em: Closure-aware embeddings for $\mathcal{EL}^+_\bot$ ontologies — technical report, https://github.com/city-artificial-intelligence/elk-em, 2026.

[30] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. doi:10.18653/v1/D19-1410.

[31] Y. Jiang, T. R. Oron, W. T. Clark, et al., An expanded evaluation of protein function prediction methods shows an improvement in accuracy, Genome Biology 17 (2016). doi:10.1186/s13059-016-1037-6.