OntoLM Tutorial | Chapter 2 of 7

ISWC BARI 2026

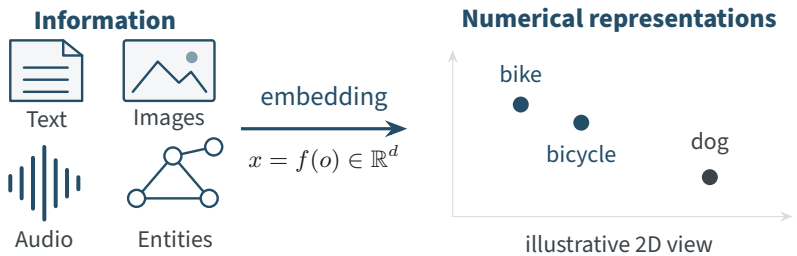
Geometric Embeddings: From Hierarchies to Axioms

Hui Yang
Research Associate
University of Manchester

MANCHESTER 1824
The University of Manchester

University of Exeter

Embeddings map information into vectors



An embedding represents an object as a vector, preserving task-relevant information

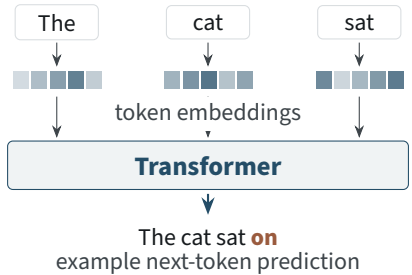
Embeddings support many different tasks

Simple: nearest-neighbour search



- Embed query and product descriptions
- Rank by vector similarity
- Retrieve by meaning, not exact wording

Complex: Transformer language models



- Map each token to a vector
- Use context to update representations
- Generate, translate and summarise

Roadmap: from hierarchies to ontologies

Three parts: hierarchies $\rightarrow$ ontologies $\rightarrow$ open problems

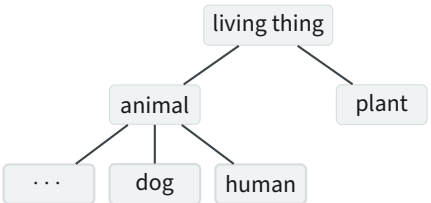
	Part	What it covers	Methods	Time
1	**Hierarchy embeddings**	is-a: asymmetry + transitivity, in low dimensions	order & box embeddings, hyperbolic space & cones, RegD[17]	20 min
2	**EL ontology embeddings**	beyond is-a: conjunction $\sqcap$, existential $\exists r$, plus analysis and training	ELEm[18], ELBE[20], BoxEL[21], Box^{2}EL[22], TransBox[23]	16 min
3	**Other works**	other logics, theoretical analysis, open problems	BoxLitE, FALCON, KB semantics	1 min

1

Hierarchy embeddings

Order and box embeddings
Hyperbolic space
RegD

Our first structured case: a hierarchy



A small taxonomy (example from RegD, KR 2026)

What must the geometry capture?

$C \sqsubseteq D$: "C is a subclass of D" (is-a)

- **Asymmetry**
  dog $\sqsubseteq$ animal, animal $\not\sqsubseteq$ dog
- **Transitivity**
  dog $\sqsubseteq$ animal $\sqsubseteq$ living thing
  $\Rightarrow$ dog $\sqsubseteq$ living thing

- **Where they appear:** plant taxonomy and product categories
  e.g. running shoes $\sqsubseteq$ shoes $\sqsubseteq$ footwear
- **Beyond trees:** a class may have multiple parents (a DAG)

Goal: embed a hierarchy while preserving its is-a relations

Order embeddings[4]: from points to nested regions

in depth

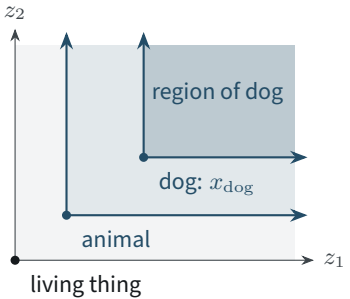
Definition: a region with apex x_u

Each concept u is embedded as an **upper orthant** R_u with apex $x_u \in \mathbb{R}^n_+$:

$$R_u = \{z \in \mathbb{R}^n_+ : z_i \geq (x_u)_i \text{ for all } i\}.$$

Encode $u \sqsubseteq v$ by inclusion:

$$R_u \subseteq R_v.$$



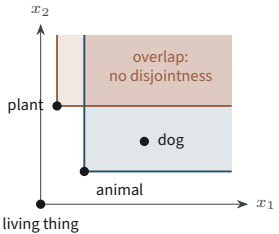
Transitivity comes from set inclusion:

$R_{\text{dog}} \subseteq R_{\text{animal}} \subseteq R_{\text{living thing}} \Rightarrow R_{\text{dog}} \subseteq R_{\text{living thing}}$.

Vendrov et al., ICLR 2016 [4], §2.1; region (cone) view: Vilnis et al., ACL 2018 [5], §3.2

Order embeddings[4]: two limitations

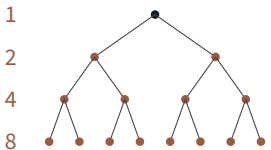
1 No disjointness



Animal and plant always overlap.

2 Crowding

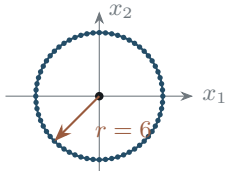
Binary tree (unit edges)



depth 3: $2^3 = 8$ nodes

depth 6: $2^6 = 64$ nodes

2D Euclidean space



64 nodes on the circle

$2\pi r \approx 37.7$

A b-ary tree: b^r **nodes** at depth r (unit edges).

In $\mathbb{R}^n$: sphere area $\propto r^{n-1}$.

Order embeddings: Vendrov et al., ICLR 2016 [4]; no disjointness: Vilnis et al., ACL 2018 [5], §4.2; crowding: Nickel & Kiela, NeurIPS 2017 [7], §2

τBox[16]: extending order embeddings with boxes

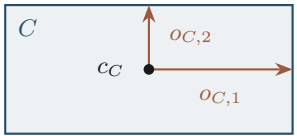
Box definition

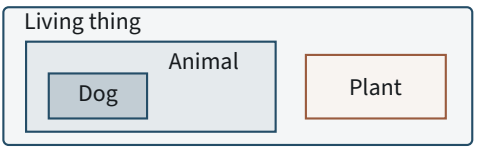
Each concept C is an axis-aligned box:

$$\mathrm{Box}(C) = [\, c_C - o_C,\ c_C + o_C \,].$$

$c_C \in \mathbb{R}^n$: **centre**

$o_C \in \mathbb{R}^n_{\geq 0}$: **offset** (half-width in each dimension)





What boxes fix

Containment: encode is-a.

Disjointness: encode empty intersection.

Remaining limitation

Euclidean space still suffers from **crowding**.

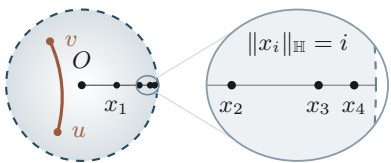
Box embeddings: Vilnis et al., ACL 2018 [5]; τBox: Boratko et al., NeurIPS 2021 [16]

Hyperbolic space: the Poincaré ball

Definition: the Poincaré ball $\mathbb{B}^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$

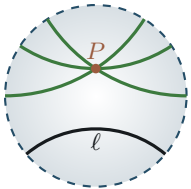
$$d_{\mathbb{H}}(u, v) = \operatorname{arcosh}\left(1 + \frac{2\|u - v\|^2}{(1 - \|u\|^2)(1 - \|v\|^2)}\right)$$

Shortest paths



Orange curve: shortest path from u to v
$\|x\|_{\mathbb{H}} := d_{\mathbb{H}}(O, x)$ (distance from O)
Boundary: **infinitely far**

Parallel lines

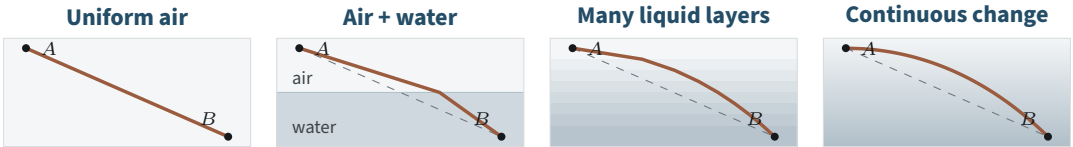


Through P: **infinitely many** lines never meet ℓ
(Euclidean space: exactly one)

Poincaré ball model; distance as in Nickel & Kiela, NeurIPS 2017 [7], §3, Eq. 1

Why can a “straight line” look curved?

An optical analogy for why a shortest route can look curved.



Orange: fastest light path

Dashed: straight shortcut

Light: fastest route

Changing speed can make the fastest route bend.

Hyperbolic space: the same intuition

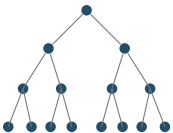
In the Poincaré ball picture, “density” increases towards the boundary, so shortest paths can look curved.

> An intuitive motivation only: not an equivalent model or a definition of hyperbolic space.

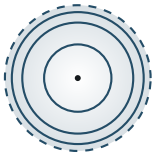
Optical analogy: Feynman Lectures I, Ch. 26; Poincaré model: Nickel & Kiela, 2017 [7]

Why hyperbolic space? Room that grows like a tree

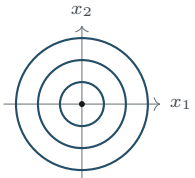
Crowding comes from $\mathbb{R}^n$: trees grow exponentially, Euclidean room only polynomially



Tree
2^ℓ nodes at level ℓ



Poincaré ball
$L(r) = 2\pi \sinh r$



Euclidean plane
$L(r) = 2\pi r$

$\times$ Euclidean $\mathbb{R}^n$: polynomial

Sphere of radius r: area $\propto r^{n-1}$.

Hyperbolic (curvature -1): exponential

Sphere area $\propto \sinh^{n-1} r \sim e^{(n-1)r}$.

Nickel & Kiela, NeurIPS 2017 [7], Section 2

Any tree fits in two hyperbolic dimensions

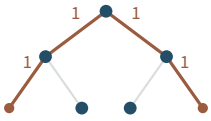
A tree as a **distance space**: vertices are points, path lengths are distances

Theorem (Sarkar, GD 2011 [8])

For every finite weighted tree T and every $\varepsilon > 0$, there are an embedding $f : V(T) \to \mathbb{H}^2$ and a global scale $\tau > 0$ such that

$$\tfrac{1}{1+\varepsilon}\, d_T(u,v) \;\le\; \tfrac{1}{\tau}\, d_{\mathbb{H}}\big(f(u), f(v)\big) \;\le\; d_T(u,v).$$

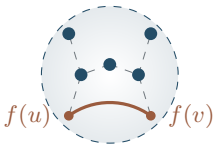
(a) Tree metric $(V(T), d_T)$



$d_T(u,v) = 1+1+1+1 = 4$

f

(b) Embedding in $\mathbb{H}^2$



$d_{\mathbb{H}}(f(u), f(v))$

Sarkar, GD 2011 [8]. Schematic mapping, not the theorem's exact construction.

Poincaré embeddings[7]: an asymmetric score

Score function ($\alpha > 0$)

$$\mathrm{score}(u \sqsubseteq v) = -\big(1 + \underbrace{\alpha(\|v\|_{\mathbb{H}} - \|u\|_{\mathbb{H}})}_{\text{asymmetry term}}\big) \quad \underbrace{d_{\mathbb{H}}(u, v)}_{\text{hyperbolic distance}}$$

$\|x\|_{\mathbb{H}} := d_{\mathbb{H}}(O, x)$: **hyperbolic norm** (distance from the origin).

- $d_{\mathbb{H}}(u, v) = d_{\mathbb{H}}(v, u)$: distance alone gives no direction.
- With the asymmetry term:

$$\|v\|_{\mathbb{H}} < \|u\|_{\mathbb{H}} \quad \Longrightarrow \quad \mathrm{score}(u \sqsubseteq v) > \mathrm{score}(v \sqsubseteq u).$$

The score favours **parents closer to the origin** than their children.

Hyperbolic-radius variant of Nickel & Kiela, NeurIPS 2017 [7], §4.3, Eq. 7

Poincaré embeddings[7]: WordNet results

Reconstruct the transitive closure of WordNet nouns: 743k is-a edges among 82k nouns

Method	Dimension	Space	MAP ↑	Mean rank ↓
Poincaré embeddings[7]	5	Hyperbolic (Poincaré ball)	**0.82**	**4.9**
Euclidean point baseline	200	Euclidean	0.17	1157

MAP ↑

How well related nouns are ranked above unrelated ones.
1 is best.

Mean rank ↓

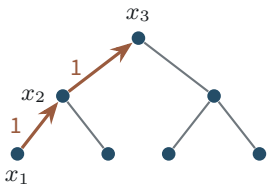
Average position of a related noun.
Lower is better.

> 40× fewer dimensions, far better reconstruction

Nickel & Kiela, NeurIPS 2017 [7], §4.1 and Table 1

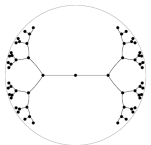
Hyperbolic embeddings keep distances, not order

Sarkar's theorem preserves the tree metric.
It does not define an is-a relation.



Arrows: child $\to$ parent





Poincaré disk

Preserved: distances

Sarkar: distortion at most $1 + \varepsilon$.

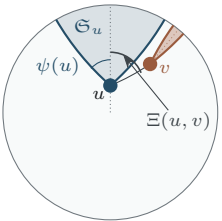
Not preserved: transitive

x_1 is a child of x_2, x_2 is a child of x_3.

$\not\Rightarrow$ x_1 is a child of x_3.

Sarkar, GD 2011 [8]. Visualisation: Stanford DAWN

Entailment cones[14]: hierarchy through nested regions



$\mathfrak{S}_u$ (blue), $\mathfrak{S}_v$ (orange) $\quad v \notin \mathfrak{S}_u$.

Cone geometry

Apex u Half-aperture $\psi(u)$ around the outward radial direction.

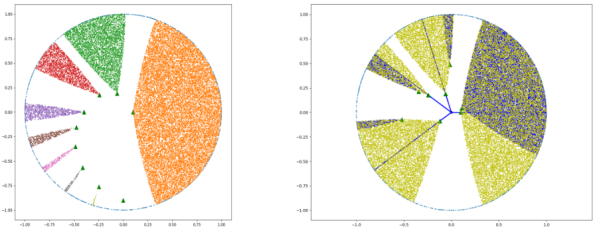
$$\psi(u) = \arcsin\left(K\frac{1-\|u\|^2}{\|u\|}\right), \quad K > 0$$

Entailment test

$\Xi(u, v)$: angle from the outward radius to the geodesic toward v.

$$E(u, v) = \max\big(0, \Xi(u, v) - \psi(u)\big)$$

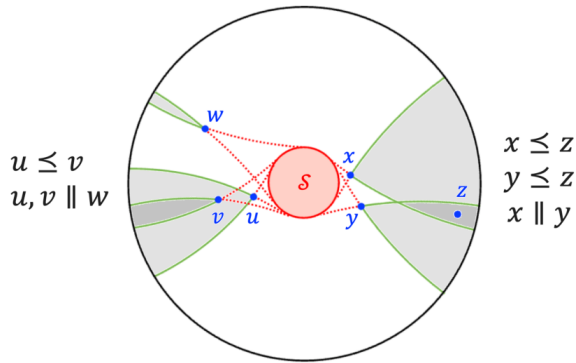
$E = 0$ inside $\quad E > 0$ outside $\quad$ Nesting gives transitivity.



Cone regions and an embedded hierarchy

Ganea et al., ICML 2018 [14], Eqs. 24–28, 33; Fig. 2. Left: $K = 0.2$, $\|u\| = 0.25$; $\|u\|$ is the Euclidean coordinate norm.

Shadow cones[15]: inspired by light and shadow



Original figure: $u \preceq v$ corresponds to $v \sqsubseteq u$.

- **Different construction:** cones inspired by light and shadow.
- **Same principle:** hierarchy through nested regions.

Yu et al., Shadow Cones: A Generalized Framework for Partial Order Embeddings, ICLR 2024 [15], Fig. 1.

Remaining limitations, and our question

× **Remaining limitations of hyperbolic methods**

- Specific cone constructions: incompatible with other region embeddings
- Hyperbolic computation, Riemannian optimisation and high numerical precision (e.g. FP64 / double precision)

Riemannian SGD: Bonnabel, IEEE TAC 2013 [10]; Bécigneul & Ganea, ICLR 2019 [11]. Precision: Sala et al., ICML 2018 [9]; Yu & De Sa, NeurIPS 2019 [12]; Mishne et al., ICML 2023 [13].

Can **Euclidean regions** get **hyperbolic expressivity**?

Remaining limitations, and our question

× **Remaining limitations of hyperbolic methods**

- Specific cone constructions: incompatible with other region embeddings
- Hyperbolic computation, Riemannian optimisation and high numerical precision (e.g. FP64 / double precision)

Riemannian SGD: Bonnabel, IEEE TAC 2013 [10]; Bécigneul & Ganea, ICLR 2019 [11]. Precision: Sala et al., ICML 2018 [9]; Yu & De Sa, NeurIPS 2019 [12]; Mishne et al., ICML 2023 [13].

Can **Euclidean regions** get **hyperbolic expressivity**?

RegD[17]: a Euclidean framework

1. **Plug-and-play**: works with existing region methods (balls, boxes)
2. **Euclidean realisation**: hyperbolic expressivity, standard computation

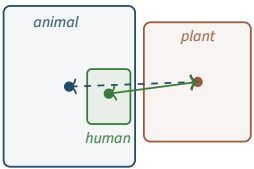
Goal: Euclidean regions + hyperbolic expressivity

What Euclidean regions lack: P1 and P2

Two key properties of hyperbolic space that Euclidean regions lack

P1: layer discrimination

A descendant can be arbitrarily farther from an outside concept than its ancestor



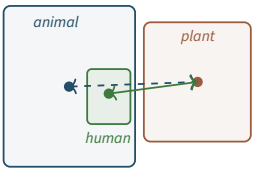
Euclidean centres: no gap guaranteed

What Euclidean regions lack: P1 and P2

Two key properties of hyperbolic space that Euclidean regions lack

P1: layer discrimination

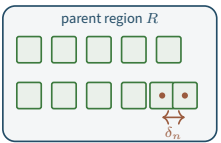
A descendant can be arbitrarily farther from an outside concept than its ancestor



Euclidean centres: no gap guaranteed

P2: child distinguishability

A parent can hold arbitrarily many children, pairwise far apart



Euclidean packing: children crowd together

Yang & Chen, KR 2026 [17], §3

RegD[17] (1): depth dissimilarity

in depth

Definition: depth dissimilarity for regions R_1, R_2

$$d_{\mathrm{dep}}(R_1, R_2) = g\left(\frac{\|P(R_1) - P(R_2)\|_p^p}{f(R_1)f(R_2)}\right)$$

P: parameters | $f(R) > 0$, region size $\to 0 \Rightarrow f(R) \to 0$ | g increasing, $g(0) = 0$.

E.g. $R = B(\mathbf{c}, r)$: $P(R) = (\mathbf{c}, r), \quad f(R) = r, \quad g(x) = x.$

Hyperbolic distance in the Poincaré ball $\mathbb{B}^n$

$$d_{\mathbb{H}}(u, v) = \operatorname{arcosh}\left(1 + \frac{2\|u - v\|^2}{(1 - \|u\|^2)(1 - \|v\|^2)}\right)$$

Yang & Chen, KR 2026 [17], §3.1, Prop. 1; Nickel & Kiela, NeurIPS 2017 [7], Eq. 1.

RegD[17] (1): depth dissimilarity

in depth

Definition: depth dissimilarity for regions R_1, R_2

$$d_{\text{dep}}(R_1, R_2) = g\left(\frac{\|P(R_1) - P(R_2)\|_p^p}{f(R_1)f(R_2)}\right)$$

P: parameters | $f(R) > 0$, region size $\to 0 \Rightarrow f(R) \to 0$ | g increasing, $g(0) = 0$.

E.g. $R = B(\mathbf{c}, r)$: $P(R) = (\mathbf{c}, r), \quad f(R) = r, \quad g(x) = x.$

Proposition 1: exact hyperbolic distance

For **Euclidean balls**, suitable f, g (with $p = 2$) make depth dissimilarity **exactly equal to hyperbolic distance** in one higher dimension.

Yang & Chen, KR 2026 [17], §3.1, Prop. 1; Nickel & Kiela, NeurIPS 2017 [7], Eq. 1.

RegD[17] (1): what depth dissimilarity guarantees

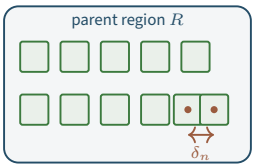
in depth

Theorem 1: d_{dep} realises P1 and P2 (balls and boxes alike)

P1: every small enough $R' \subseteq R_2$ has $d_{\mathrm{dep}}(R_1, R') > d_{\mathrm{dep}}(R_1, R_2) + \Delta$, for any $\Delta > 0$.

P2: every region contains arbitrarily many children, pairwise $d_{\mathrm{dep}} > M$, for any $M > 0$.

P2 example: fit more children, yet keep them distinguishable



Euclidean centres crowd together.

$$d_{\mathrm{dep}}(R_i, R_j) = g\left(\frac{\|P(R_i) - P(R_j)\|_p^p}{f(R_i) f(R_j)}\right)$$

Step 1. Shrink to fit. Make regions small enough to fit any number N inside the parent.

Step 2. Shrink again to separate. Even with tiny parameter gaps, shrink sizes further $\Rightarrow$ pairwise $d_{\mathrm{dep}} > \delta_n$.

Yang & Chen, KR 2026 [17], §3.1, Theorem 1. P2 example uses $g(x) = x$.

RegD[17] (2): boundary dissimilarity

Boundary dissimilarity: is R_2 inside R_1?

$$d_{\mathrm{bd}}(R_1, R_2) = \begin{cases} -\min_{\mathbf{x}_2 \in R_2} d(R_1^c, \mathbf{x}_2), & R_2 \subseteq R_1, \\ \max_{\mathbf{x}_2 \in R_2 \setminus R_1} d(R_1, \mathbf{x}_2), & \text{otherwise.} \end{cases}$$

Inside: minus the cost to move part of R_2 out of R_1.

Outside: the cost to move all of R_2 into R_1.

Yang & Chen, KR 2026 [17], §3.2, Def. of boundary dissimilarity and ball example.

RegD[17] (2): boundary dissimilarity

Boundary dissimilarity: is R_2 inside R_1?

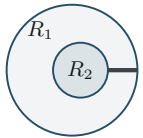
$$d_{\mathrm{bd}}(R_1, R_2) = \begin{cases} -\min_{\mathbf{x}_2 \in R_2} d(R_1^c, \mathbf{x}_2), & R_2 \subseteq R_1, \\ \max_{\mathbf{x}_2 \in R_2 \setminus R_1} d(R_1, \mathbf{x}_2), & \text{otherwise.} \end{cases}$$

Inside: minus the cost to move part of R_2 out of R_1.

Outside: the cost to move all of R_2 into R_1.

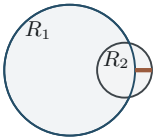
Example: balls $R_i = B(\mathbf{c}_i, r_i)$

$d_{\mathrm{bd}}(R_1, R_2) = \|\mathbf{c}_1 - \mathbf{c}_2\|_2 + r_2 - r_1$



$d_{\mathrm{bd}}(R_1, R_2) < 0$

contained



$d_{\mathrm{bd}}(R_1, R_2) > 0$

violation

$d_{\mathrm{bd}}(R_1, R_2) \le 0 \iff R_2 \subseteq R_1$ (zero: contained and touching).

Yang & Chen, KR 2026 [17], §3.2, Def. of boundary dissimilarity and ball example.

RegD[17] (3): the score function

Poincaré embeddings[7]: points in $\mathbb{B}^n$

$$-\mathrm{score}(u \sqsubseteq v) \;=\; \big(1 + \underbrace{\alpha\,(\|v\| - \|u\|)}_{\text{direction: norm}}\big) \cdot \underbrace{d_{\mathbb{H}}(u, v)}_{\text{hyperbolic distance}}$$

RegD[17]: regions in $\mathbb{R}^n$

$$E(u \sqsubseteq v) \;=\; \underbrace{d_{\mathrm{bd}}(R_v, R_u)}_{\text{direction: inclusion}} \;+ \lambda \cdot \underbrace{d_{\mathrm{dep}}(R_v, R_u)}_{\text{hyperbolic-like distance}}$$

Yang & Chen, KR 2026 [17], §3.3 and Table 1; Nickel & Kiela, NeurIPS 2017 [7], Eq. 7

RegD[17] (3): the score function

Poincaré embeddings[7]: points in $\mathbb{B}^n$

$$-\text{score}(u \sqsubseteq v) \;=\; \big(1 + \underbrace{\alpha\,(\|v\| - \|u\|)}_{\text{direction: norm}}\big) \cdot \quad \underbrace{d_{\mathbb{H}}(u, v)}_{\text{hyperbolic distance}}$$

RegD[17]: regions in $\mathbb{R}^n$

$$E(u \sqsubseteq v) \;=\; \underbrace{d_{\text{bd}}(R_v, R_u)}_{\text{direction: inclusion}} \;+ \lambda \cdot \quad \underbrace{d_{\text{dep}}(R_v, R_u)}_{\text{hyperbolic-like distance}}$$

- Lower is better.
- Predict $u \sqsubseteq v$ when $E \leq t$ (threshold tuned on validation).
- λ **tunes the distance term**: $\lambda = 0 \rightarrow$ pure inclusion
  $\lambda > 0 \rightarrow$ adds hyperbolic behaviour.

Yang & Chen, KR 2026 [17], §3.3 and Table 1; Nickel & Kiela, NeurIPS 2017 [7], Eq. 7

RegD[17] experiments (1): DAG inference

Train on basic edges, predict held-out inferred edges

F1 (%)	dims	best baseline	RegD[17]
Mammal	2 / 5	57.7 / 69.4	64.9 / 71.8
WordNet Noun	5 / 10	45.2 / 52.2	58.4 / 59.1
MCG	5 / 10	48.4 / 55.5	50.7 / 58.5
Hearst	5 / 10	41.6 / 46.8	42.8 / 49.6

- Baselines: entailment / shadow cones, order emb., τBox, ELBE, ELEm
- RegD best in **8/8** settings
- Largest gain $+13.2$ (Noun, $d{=}5$)

Yang & Chen, KR 2026 [17], Table 2

Hierarchy embeddings: a comparison

Method	Representation	Transitivity	Disjointness	No crowding (low dim.)	Euclidean computation
Order embeddings[4]	orthants in $\mathbb{R}^n_+$	✓	×	×	✓
τBox[16]	boxes in $\mathbb{R}^n$	✓	✓	×	✓
Poincaré embeddings[7]	points in $\mathbb{B}^n$	×	×	✓	×
Hyperbolic cones	cones in $\mathbb{H}^n$	✓	✓	✓	×
RegD[17]	any regions in $\mathbb{R}^n$	partial*	✓	✓	✓

* Full transitivity only for $\lambda = 0$ (pure inclusion, but then no hyperbolic room).

Summary of the methods above; RegD: Yang & Chen, KR 2026 [17], §3.3

2

EL-family ontology embeddings

EL-family normal forms
From points to regions
ELEm $\rightarrow$ boxes $\rightarrow$ TransBox

EL-family normal forms

From hierarchies to ontology axioms between concepts

Normal forms (templates)

NF1 $A \sqsubseteq B$

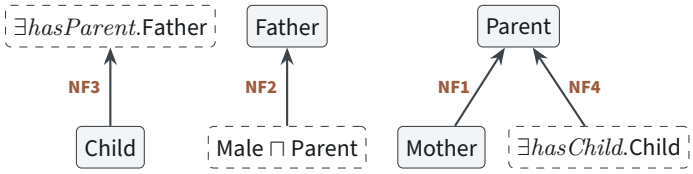
NF2 $A_1 \sqcap A_2 \sqsubseteq B$

NF3 $A \sqsubseteq \exists r.B$

NF4 $\exists r.B \sqsubseteq A$

A, A_i, B: concept names
r: role

Family example: a hierarchy of (complex) concepts



Arrows: $\sqsubseteq$ (sub $\rightarrow$ super)
Dashed boxes: complex concepts

Beyond basic EL: two role-axiom forms used here

Role inclusion: $r \sqsubseteq s$
Example: $hasMother \sqsubseteq hasParent$

Role-chain inclusion: $r_1 \circ r_2 \sqsubseteq s$
Example: $hasParent \circ hasParent \sqsubseteq hasGrandparent$

Normal forms: Baader et al., An Introduction to Description Logic, 2017 [1], Ch. 6; family ontology: Yang et al., WWW 2025 [23], Fig. 4a

Compare KG embeddings to ontology embeddings

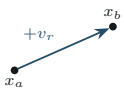
Basic idea: extend KG embeddings from **points** to **regions**

KG embedding (TransE[2])

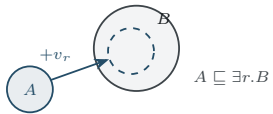
- Tuples $r(a, b)$ between **instances**
- Instance $\mapsto$ point

Ontology embedding

- Axioms $A \sqsubseteq B$ between **concepts**
- Concepts **include** each other $\Rightarrow$ concept $\mapsto$ **region**



KG: points



ontology: regions

TransE: Bordes et al., NeurIPS 2013 [2]

ELEm[18]: balls and translations

in depth

Concept $A \mapsto \mathrm{Ball}(A) := B(c_A, r_A)$ (centre, radius)
Role $r \mapsto$ translation vector v_r

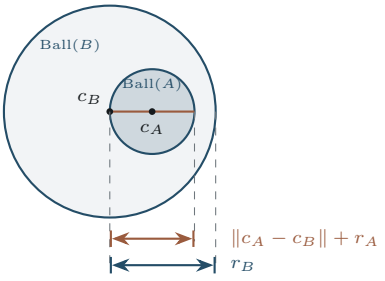
NF1 $A \sqsubseteq B$**:** $\mathrm{Ball}(A) \subseteq \mathrm{Ball}(B)$

$$\mathrm{Ball}(A) \subseteq \mathrm{Ball}(B) \iff \|c_A - c_B\| + r_A \leq r_B$$

Loss: $[\,\|c_A - c_B\| + r_A - r_B - \gamma\,]_+$ (margin γ)

Encode: $(\exists r.B)^{\mathcal{I}} = \mathrm{Ball}(B) - v_r$.

NF3 $A \sqsubseteq \exists r.B$: $\mathrm{Ball}(A) + v_r \subseteq \mathrm{Ball}(B)$.
NF4 $\exists r.B \sqsubseteq A$: $\mathrm{Ball}(B) - v_r \subseteq \mathrm{Ball}(A)$.



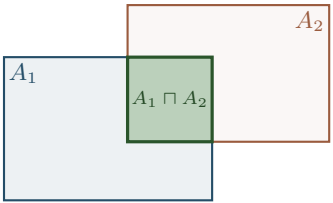
orange $\leq$ blue $\Rightarrow$ ball containment

NF2 $A_1 \sqcap A_2 \sqsubseteq B$: $\mathrm{Ball}(A_1) \cap \mathrm{Ball}(A_2)$ is generally not a ball $\Rightarrow$ **approximate** loss.
Zero loss need not imply $\mathrm{Ball}(A_1) \cap \mathrm{Ball}(A_2) \subseteq \mathrm{Ball}(B)$.

Kulmanov et al., IJCAI 2019 [18], §3.2, Eqs. 1–4

ELBE[20]: intersectional closure

brief



The box picture from ELBE

Roles use translations: $(\exists r.B)^{\mathcal{I}} = \mathrm{Box}(B) - v_r$.

NF2 $A_1 \sqcap A_2 \sqsubseteq B$

$$\mathrm{Box}(A_1) \cap \mathrm{Box}(A_2) \subseteq \mathrm{Box}(B)$$

Conjunction is represented exactly:

the intersection of two boxes is a box (or empty).

Peng et al., ELBE, arXiv 2022 [20], §4.2–4.3, Eqs. 2–6

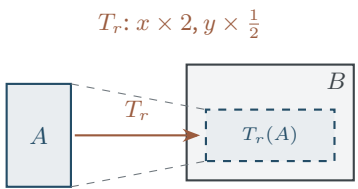
BoxEL[21]: roles as affine maps

Role definition (invertible case)

$$T_r(x) = D_r x + b_r, \quad D_r = \mathrm{diag}(d_{r,1}, \ldots, d_{r,n}), \quad d_{r,i} > 0; \qquad r^{\mathcal{I}} = \{(x, T_r(x)) : x \in \mathbb{R}^n\}.$$

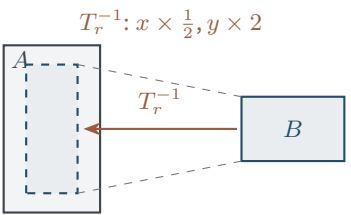
Encode $\exists r.B$ as $T_r^{-1}(\mathrm{Box}(B))$: take the preimage of B under the role map.

NF3 $A \sqsubseteq \exists r.B$



$$T_r\big(\mathrm{Box}(A)\big) \subseteq \mathrm{Box}(B)$$

NF4 $\exists r.B \sqsubseteq A$



$$T_r^{-1}\big(\mathrm{Box}(B)\big) \subseteq \mathrm{Box}(A)$$

Xiong et al., ISWC 2022 [21], §4.1, §4.4, Eqs. 9–10; adapted from Fig. 2(f), illustrative scales

Beyond functions: roles can be many-to-many

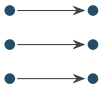
One transformation per role

ELEm[18] **/ ELBE**: one translation vector

$$x_b = x_a + v_r$$

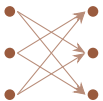
BoxEL[21]: one affine map per role.

$\Rightarrow$ Each input has **one output**.



One output per input

each input $\mapsto$ exactly one output



Many-to-many

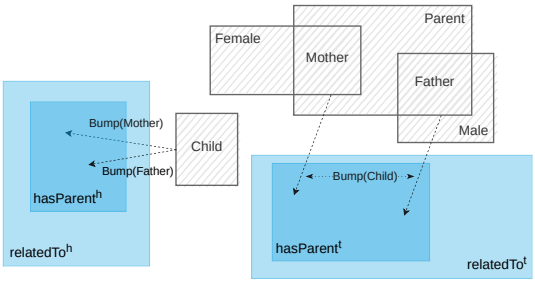
each input $\mapsto$ several outputs,
each output $\leftarrow$ several inputs
e.g. colleagueOf, hasParent

Next: **Box**^{2}**EL**[22] $\rightarrow$ head / tail boxes + bumps
TransBox[23] $\rightarrow$ boxes of translations

ELEm: Kulmanov et al., 2019 [18]; ELBE: Peng et al., 2022 [20]; BoxEL: Xiong et al., 2022 [21].

Box^{2}EL[22]: concept boxes and role position boxes

Each concept: a box and one learned bump.
Box^{2}EL: $A \mapsto (\mathrm{Box}(A), b_A)$



Stripes: concept boxes Blue: role head / tail boxes
Arrows: bumps

Each role: head and tail boxes.
$r \mapsto (H_r, T_r)$.

NF3 training: $A \sqsubseteq \exists r.B$

$$\mathrm{Box}(A) + b_B \subseteq H_r$$
$$\mathrm{Box}(B) + b_A \subseteq T_r$$

NF4 training: $\exists r.B \sqsubseteq A$

Step 1. Cover the existential region:
For any $x \in (\exists r.B)^{\mathcal{I}}$, NF3's head condition gives

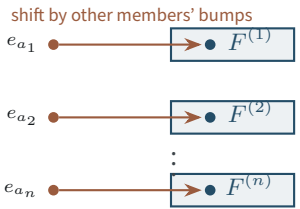
$$x + b_B \in H_r \;\Rightarrow\; x \in H_r - b_B.$$

Step 2. Require containment in A:
$H_r - b_B \subseteq \mathrm{Box}(A)$ (sufficient condition).

BoxE: the source of Box^{2}EL's point-and-bump idea

$F(a_1, \ldots, a_n)$ is a relation fact: F names the relation, and each a_i is a participating entity.

- **Example:** $F = \text{TeamMembers}$
  Each a_i is one team member.
- **Each member a_i:** a point e_{a_i} and one learned bump b_{a_i}.
- **Relation F:** one box $F^{(i)}$ for each argument position i.



One box per argument position

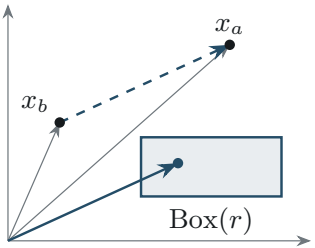
$F(a_1, \ldots, a_n)$: $e_{a_i} + \sum_{j \neq i} b_{a_j} \in F^{(i)}$ for every i. Other members' bumps only.

Abboud et al., NeurIPS 2020 (BoxE), §4, Eq. 1

TransBox[23]: roles as boxes of translations

in depth

A single translation vector → a **region of possible translations**



$r(a, b)$ iff $x_a - x_b \in \mathrm{Box}(r)$

A box for each role

Pair (a, b) satisfies r:
its displacement lies inside the role box

Many valid displacements

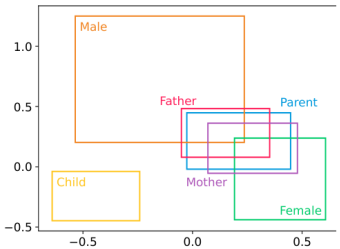
The same individual can relate to several others.

⇒ **many-to-many roles**

Concepts = regions of individuals
Roles = regions of translations

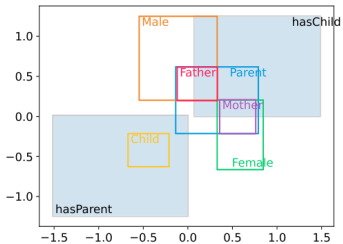
TransBox[23]: an embedding example

Family ontology (TransBox paper): a child has a mother *and* a father, and Male and Female are disjoint



(b) ELBE

ELBE: Father and Mother overlap



(d) TransBox

TransBox: they are disjoint (roles are the **large** shaded boxes)

[screenshot] Yang et al., WWW 2025 [23], Fig. 4 (b, d)

TransBox[23]: sparse intersections

Box are sparse in high dimensions spaces.

Challenge: Theorem 4.5

Centers: $[-1, 1]^n$ Offsets: $(0, a_0]^n$ ($a_0 < 1$).

$$P(B \cap B' \neq \emptyset) \leq \left(a_0 - \tfrac{1}{4}a_0^2\right)^n \quad \Longrightarrow \quad n = 50,\ a_0 \uparrow 1 :\ P < 6 \times 10^{-7}.$$

Strategy: Extend the space.

Add an empty coordinate value:

$$\mathbb{R}^n \longrightarrow (\mathbb{R} \cup \{\emptyset\})^n$$

$$\widetilde{B} = \prod_{j=1}^{n} (I_j \cup \{\emptyset\})$$

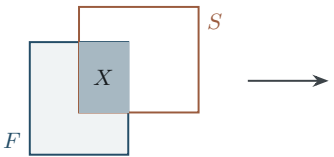
- If one axis is disjoint, its coordinate becomes $\{\emptyset\}$. Other axes may overlap.
- An adapted loss for compute the inclusion in this case.

Yang et al., WWW 2025 [23], §4.4–4.5, Thm. 4.5

Analysis (1): composing a complex concept

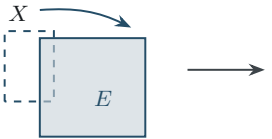
We can build **Person** $\sqcap \exists \mathit{hasChild}.(\textbf{Female} \sqcap \textbf{Student})$ in three steps

1 Intersect



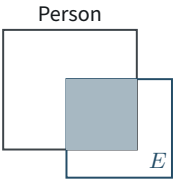
$X = \mathrm{Box}(F) \cap \mathrm{Box}(S)$
Female students

2 Apply the role



$E = X \oplus \mathrm{Box}(\mathit{hasChild})$
Those with a female student child

3 Intersect again



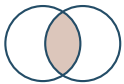
$\mathrm{Box}(\mathsf{Person}) \cap E$
The complete concept

Closure: every step returns a box $\Rightarrow$ any nesting can be built
Methods differ in steps 1, 2

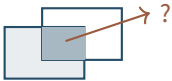
Yang et al., WWW 2025 [23], §3–4; $\oplus$: Minkowski sum; F = Female, S = Student

Analysis (2): conjunctions and NF4

Method	1 $C \sqcap D$	2 $\exists r.B$	How step 2 is built
ELEm[18]	✗	✓	translate a ball
ELBE[20]	✓	✓	translate a box
BoxEL[21]	✓	✓	affine preimage of a box
Box^{2}EL[22]	◐	✓	NF4: $H_r - b_B \subseteq \mathrm{Box}(A)$
TransBox[23]	✓	✓	box sum $\oplus$



ELEm, step 1: ball ∩ ball is a lens



Box^{2}EL, step 1: box intersection exists, but no bump for $C \sqcap D$

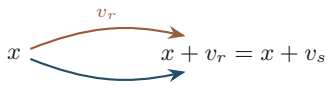
Kulmanov et al., 2019 [18]; Peng et al., 2022 [20]; Xiong et al., 2022 [21]; Jackermeier et al., 2024 [22]; Yang et al., WWW 2025 [23], §3

Analysis (3): how the methods handle role axioms

Compare the geometric condition with what each role representation can express.

Method	Role	$r \sqsubseteq s$	$r_1 \circ r_2 \sqsubseteq s$
ELEm, ELBE[18,20]	translation v_r	◐ only $v_r = v_s$	◐ only $v_{r_1} + v_{r_2} = v_s$
BoxEL[21]	affine map T_r	◐ only $T_r = T_s$	◐ only $T_{r_2} \circ T_{r_1} = T_s$
Box^{2}EL[22]	head, tail boxes	✓ $H_r \subseteq H_s, T_r \subseteq T_s$	◐ $H_{r_1} \subseteq H_s$, $T_{r_2} \subseteq T_s$
TransBox[23]	box of translations	✓ $\mathrm{Box}(r) \subseteq \mathrm{Box}(s)$	◐ $\mathrm{Box}(r_1) \oplus \mathrm{Box}(r_2) \subseteq \mathrm{Box}(s)$

ELEm / ELBE: one vector per role



$r \sqsubseteq s$ forces the two maps to coincide.

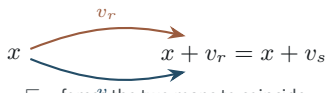
Jackermeier et al., WWW 2024 [22], §3.1–3.2; Yang et al., WWW 2025 [23], §4.1, Prop. 4.2; Mondal et al., AAAI-MAKE 2021 [19], §3.2.3 (EmEL++)

Analysis (3): how the methods handle role axioms

Compare the geometric condition with what each role representation can express.

Method	Role	$r \sqsubseteq s$	$r_1 \circ r_2 \sqsubseteq s$
ELEm, ELBE[18,20]	translation v_r	◐ only $v_r = v_s$	◐ only $v_{r_1} + v_{r_2} = v_s$
BoxEL[21]	affine map T_r	◐ only $T_r = T_s$	◐ only $T_{r_2} \circ T_{r_1} = T_s$
Box^{2}EL[22]	head, tail boxes	✓ $H_r \subseteq H_s, T_r \subseteq T_s$	◐ $H_{r_1} \subseteq H_s$, $T_{r_2} \subseteq T_s$
TransBox[23]	box of translations	✓ $\mathrm{Box}(r) \subseteq \mathrm{Box}(s)$	◐ $\mathrm{Box}(r_1) \oplus \mathrm{Box}(r_2) \subseteq \mathrm{Box}(s)$

ELEm / ELBE: one vector per role



$r \sqsubseteq s$ forces the two maps to coincide.

Can't handle transitive role ($r \circ r \sqsubseteq r$)

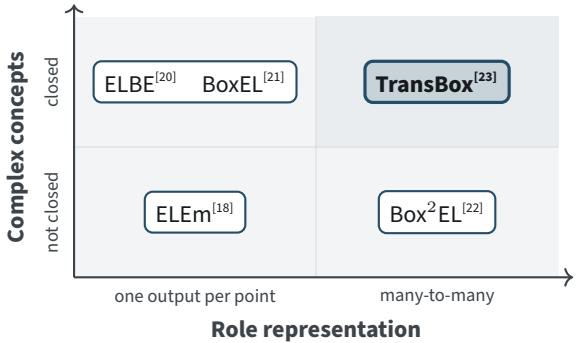
Box^{2}EL: $H_r \subseteq H_r, T_r \subseteq T_r$
Always true ⇒ every role is transitive.

TransBox: $\mathrm{Box}(r) \oplus \mathrm{Box}(r) \subseteq \mathrm{Box}(r)$
$\Rightarrow \mathrm{Box}(r) = \{\mathbf{0}\}$ or $\mathbb{R}^n$.

Jackermeier et al., WWW 2024 [22], §3.1–3.2; Yang et al., WWW 2025 [23], §4.1, Prop. 4.2; Mondal et al., AAAI-MAKE 2021 [19], §3.2.3 (EmEL++)

Analysis (4): capabilities at a glance

Two separate capabilities: composing complex concepts vs. many-to-many roles



One output per point

Translation or affine map:
each point $\mapsto$ **one** point

$\Rightarrow$ a child cannot reach both a mother and a father box

Many-to-many

Box^{2}EL: head / tail boxes
TransBox: a **box of translations**.

Right column: many-to-many roles and role axioms. **Training: see backup.**

ELEm [18], ELBE [20], BoxEL [21], Box^{2}EL [22], TransBox [23] (2019–2025)

3

Other works

Beyond the EL family
Theoretical analysis
Open problems

Other works: models and theory at a glance

brief

Different supported constructors — not a strict inclusion chain

DL-Lite$_{\mathcal{H}}$

BoxLitE[32] KR'26
convex training (SOCP)
✓ weakly faithful

EL family

DF-$\mathcal{EL}^{++}$ [33] KDD'26
fuzzy truth degrees

DELE[27] NAI'26
closure-aware negatives

$\mathcal{ALC}$ ($\neg$, $\sqcup$, $\forall$)

FALCON[24] preprint
fuzzy models

CatE[25] NAI'25
concept lattice + order emb.

Axis-aligned cones[26] JAIR'23
$\neg$ via polar cones

Theoretical analysis and results across ontology families

KB embedding semantics[29] KR'24
faithfulness, entailment closure, expressiveness

Strong faithfulness for ELH[30] TGDK'24
existence proof for convex geometric models

Discussion: hierarchy embeddings

Hyperbolic methods work well, but

- hard to implement: Riemannian geometry needs manifold-aware optimisation and numerical care.
- hard to reuse: special constructions (e.g. cones) bring method-specific order tests.

RegD[17]: a first step, hyperbolic in $\mathbb{R}^n$

- Transitivity is only partly supported: $\lambda = 0$ gives exact inclusion. $\lambda > 0$ adds depth but weakens that guarantee.
- Open score design: a sum $d_{\mathrm{bd}} + \lambda d_{\mathrm{dep}}$, a product (as in Poincaré), or another combination?

Hyperbolic optimisation: Mishne et al., ICML 2023 [13]; cones: Ganea et al., ICML 2018 [14]; RegD: Yang & Chen, KR 2026 [17], §3.3

Discussion: ontology embeddings

Problems

1. **Role axioms**: a role hierarchy is still weak.
   In particular, role chains $r_1 \circ r_2 \sqsubseteq s$.
2. **Beyond the EL family**: little work on $\mathcal{ALC}$, and mostly theoretical.
   Negation is the main problem
3. **Real-world applications**: move beyond axiom-ranking benchmarks.
   E.g., clinical ontology completion and ontology-guided biological prediction.
4. **Lack of transfer abilities**

Next (Chapter 3): language models bring in the labels

Ontology-embedding survey [3]; role-method comparison above; broader logics: FALCON [24], CatE [25].

References (1/2): foundations and hierarchies

[1] Baader, Horrocks, Lutz, Sattler. An Introduction to Description Logic. Cambridge University Press, 2017.

[2] Bordes, Usunier, García-Durán, Weston, Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. NeurIPS 2013.

[3] Chen, Mashkova, Zhapa-Camacho, Hoehndorf, He, Horrocks. Ontology Embedding: A Survey of Methods, Applications and Resources. IEEE TKDE, 2025.

[4] Vendrov et al. Order-Embeddings of Images and Language. ICLR 2016.

[5] Vilnis, Li, Murty, McCallum. Probabilistic Embedding of Knowledge Graphs with Box Lattice Measures. ACL 2018.

[7] Nickel & Kiela. Poincaré Embeddings for Learning Hierarchical Representations. NeurIPS 2017.

[8] Sarkar. Low Distortion Delaunay Embedding of Trees in Hyperbolic Plane. GD 2011.

[9] Sala, De Sa, Gu, Ré. Representation Tradeoffs for Hyperbolic Embeddings. ICML 2018.

[10] Bonnabel. Stochastic Gradient Descent on Riemannian Manifolds. IEEE Trans. Automatic Control, 2013.

[11] Bécigneul & Ganea. Riemannian Adaptive Optimization Methods. ICLR 2019.

[12] Yu & De Sa. Numerically Accurate Hyperbolic Embeddings Using Tiling-Based Models. NeurIPS 2019.

[13] Mishne, Wan, Wang, Yang. The Numerical Stability of Hyperbolic Representation Learning. ICML 2023.

[14] Ganea, Bécigneul, Hofmann. Hyperbolic Entailment Cones for Learning Hierarchical Embeddings. ICML 2018.

[15] Yu et al. Shadow Cones: A Generalized Framework for Partial Order Embeddings. ICLR 2024.

[16] Boratko et al. Capacity and Bias of Learned Geometric Embeddings for Directed Graphs (τBox). NeurIPS 2021.

[17] Yang & Chen. RegD: Hierarchical Embeddings via Dissimilarity between Arbitrary Euclidean Regions. KR 2026.

References (2/2): ontology embeddings and theory

[18] Kulmanov et al. EL Embeddings: Geometric Construction of Models for the Description Logic EL++ (ELEm). IJCAI 2019.

[19] Mondal et al. EmEL++: Embeddings for EL++ Description Logic. AAAI-MAKE 2021.

[20] Peng et al. Description Logic EL++ Embeddings with Intersectional Closure (ELBE). arXiv 2202.14018, 2022.

[21] Xiong et al. Faithful Embeddings for EL++ Knowledge Bases (BoxEL). ISWC 2022.

[22] Jackermeier, Chen, Horrocks. Dual Box Embeddings for the Description Logic EL++ (Box^{2}EL). WWW 2024.

[23] Yang, Chen, Sattler. TransBox: EL++-closed Ontology Embedding. WWW 2025.

[24] Hinnerichs et al. FALCON: Scalable Reasoning over Inconsistent ALC Ontologies. arXiv:2208.07628v5, 2024.

[25] Zhapa-Camacho & Hoehndorf. Lattice-Based ALC Ontology Embeddings with Saturation (CatE). Neurosymbolic AI, 2025.

[26] Özçep, Leemhuis, Wolter. Embedding Ontologies in the Description Logic ALC by Axis-Aligned Cones. JAIR 2023.

[27] Mashkova, Zhapa-Camacho, Hoehndorf. DELE: Deductive EL++ Embeddings for Knowledge Base Completion. Neurosymbolic AI, 2026.

[29] Bourgaux et al. Knowledge Base Embeddings: Semantics and Theoretical Properties. KR 2024.

[30] Lacerda et al. Strong Faithfulness for ELH Ontology Embeddings. TGDK 2024.

[32] Lourenço, Morgan, Ozaki, Pavlović, Sallinger. BoxLitE: A Faithful Knowledge Base Embedding Based on Convex Optimization. KR 2026.

[33] Zhao. Fast and Faithful: Scalable Neuro-Symbolic Learning and Reasoning with Differentiable Fuzzy EL++. KDD 2026.