NATIONAL RESEARCH UNIVERSITY

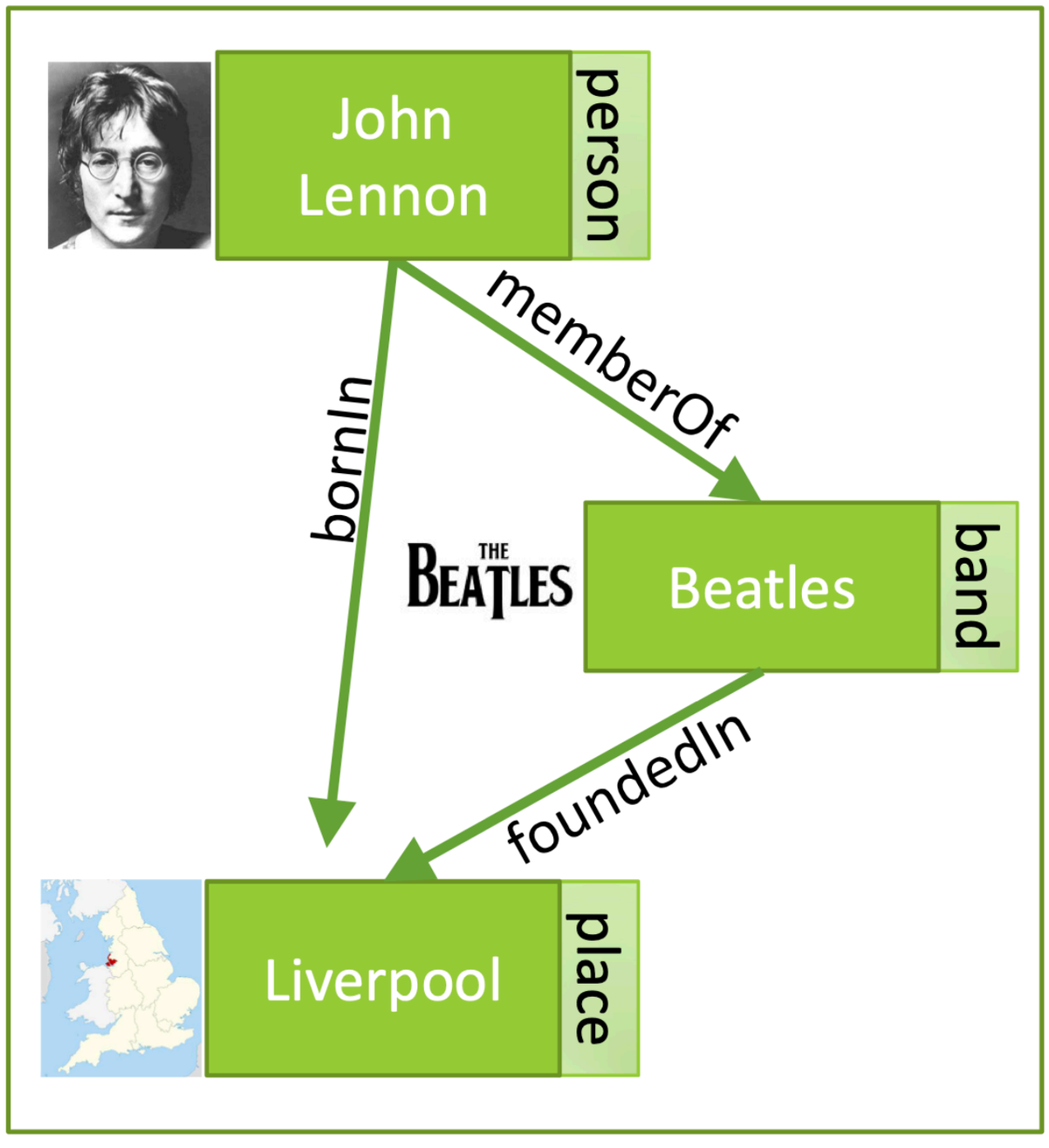
Department of Computer Science

KNOWLEDGE GRAPHS

Network Science

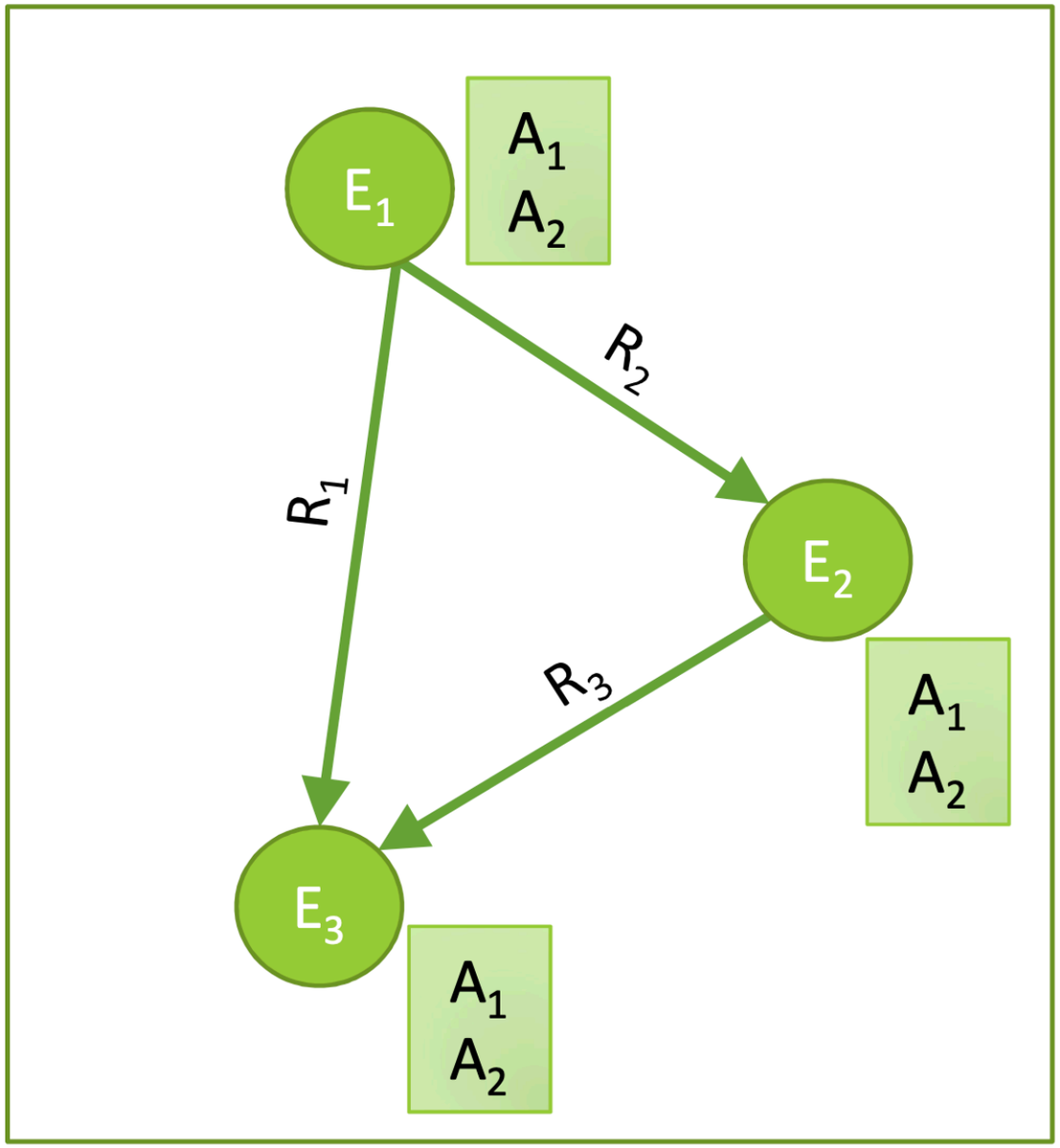
Lecture 16

KNOWLEDGE GRAPHS

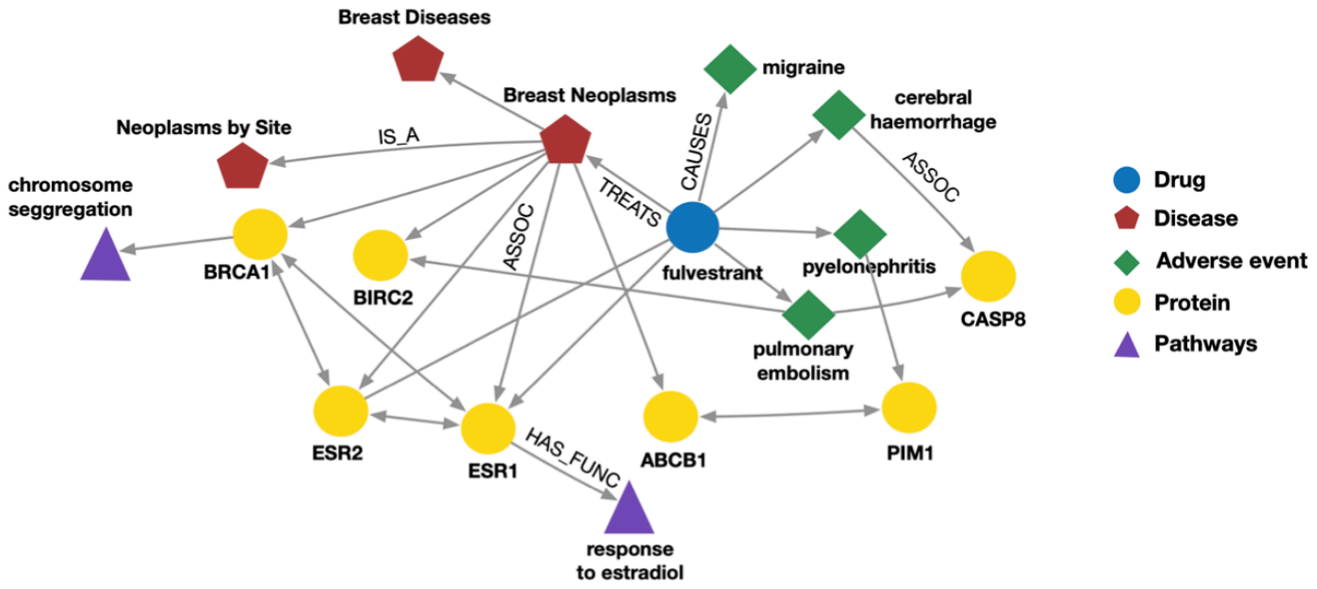


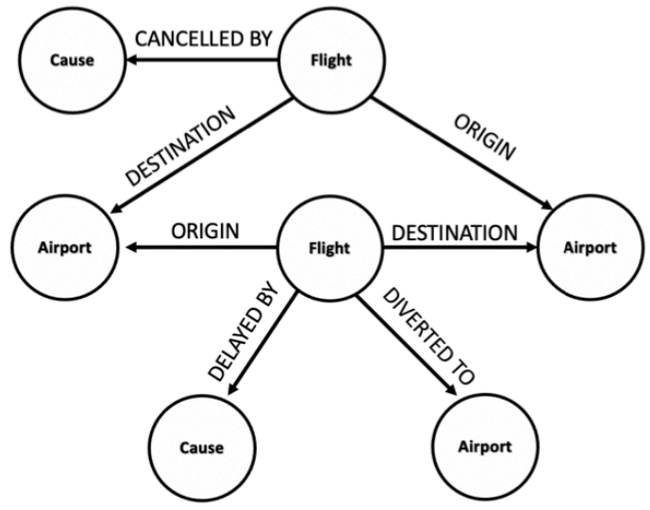
KNOWLEDGE GRAPHS

- Nodes are entities
- Nodes are labeled with attributes (types)
- Edges are typed, captures types of relationships between entities



KNOWLEDGE GRAPHS





Biomedical Knowledge Graphs

Example node: Migraine
Example edge: (fulvestrant, Treats, Breast Neoplasms)
Example node type: Protein
Example edge type (relation): Causes

Event Graphs

Example node: SFO
Example edge: (UA689, Origin, LAX)
Example node type: Flight
Example edge type (relation): Destination

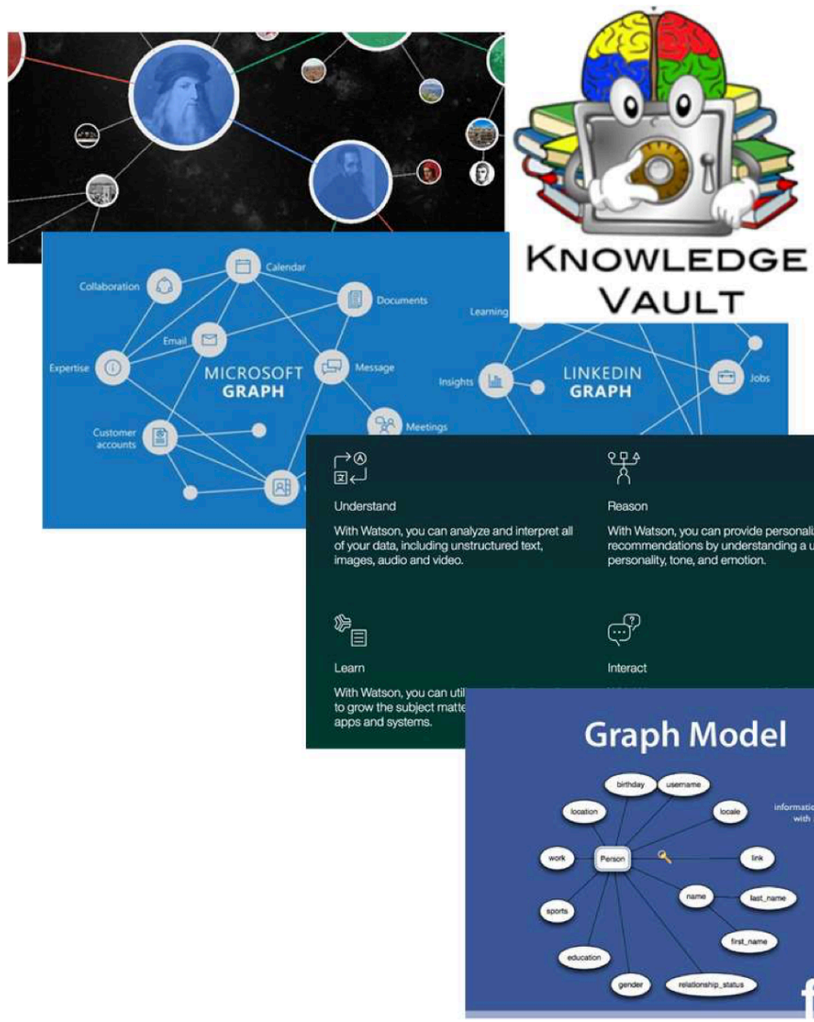
INFORMATION EXTRACTION

Unstructured
Ambiguous
Lots and lots of it!

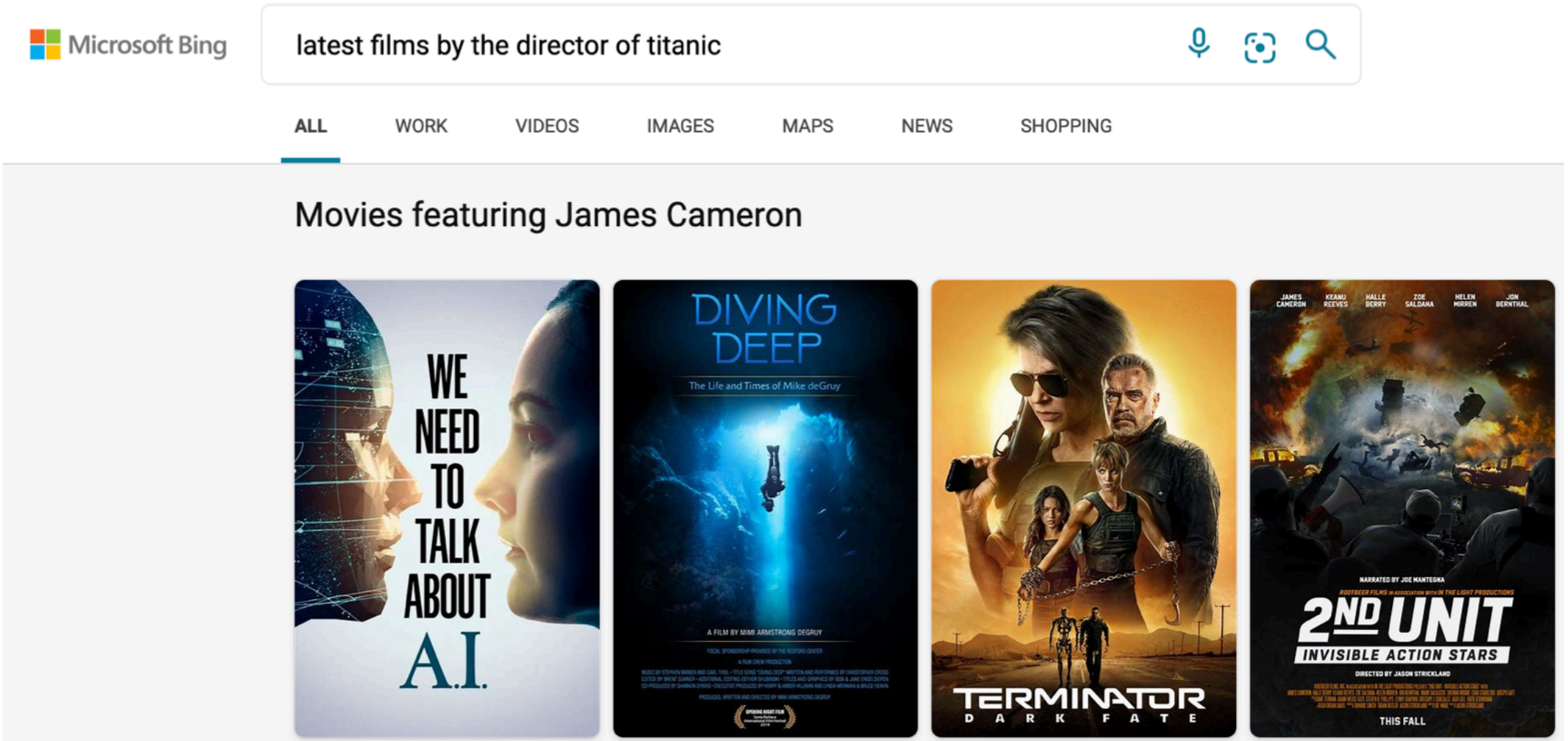
Structured
Precise, Actionable
Specific to the task

KNOWLEDGE GRAPHS

- Google Knowledge Graph
  - Google Knowledge Vault
- Amazon Product Graph
- Facebook Graph API
- IBM Watson
- Microsoft Satori
  - Project Hanover/Literome
- LinkedIn Knowledge Graph
- Yandex Object Answer
- Diffbot, GraphIQ, Maana, ParseHub, Reactor Labs, SpazioDati

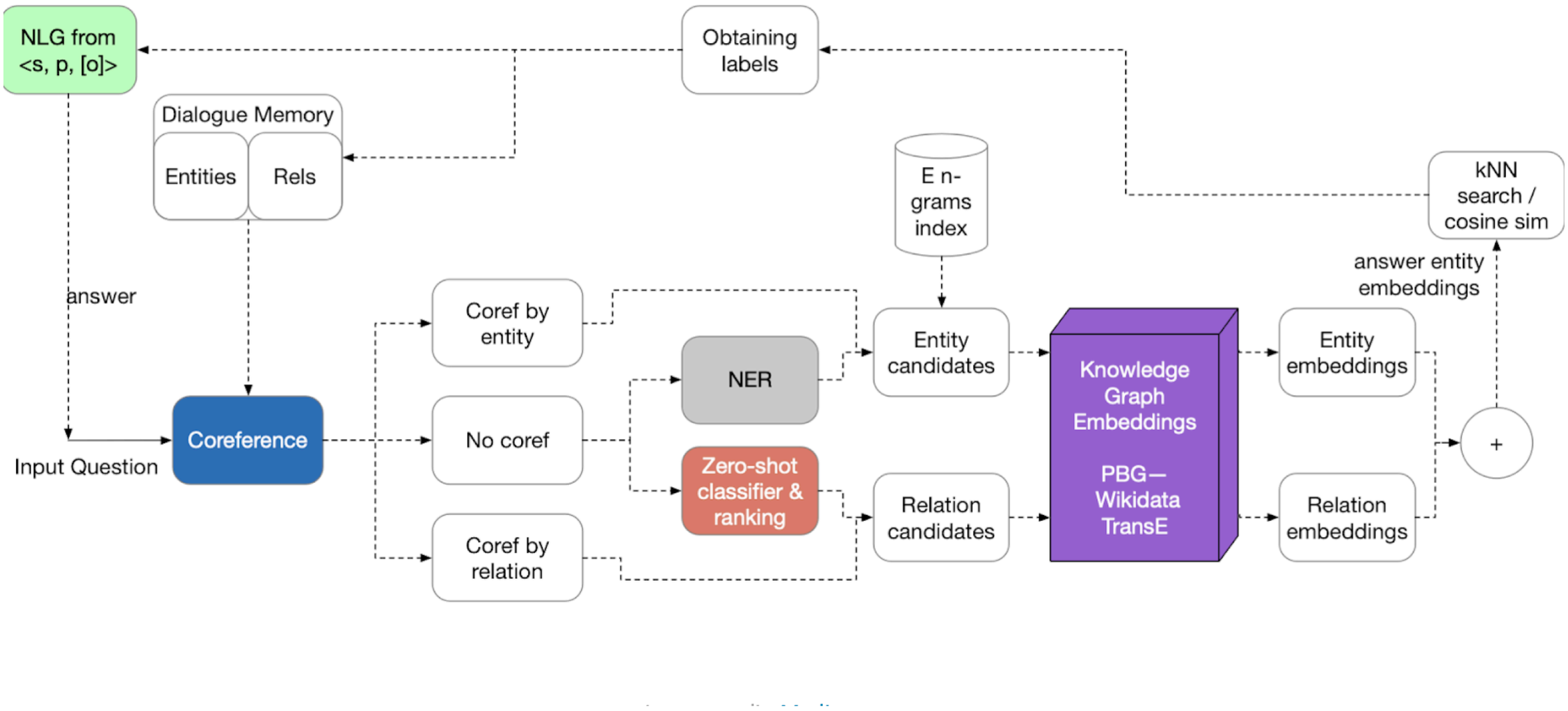


KNOWLEDGE GRAPHS



KNOWLEDGE GRAPHS

- **Question answering and conversation agents**



KNOWLEDGE GRAPHS

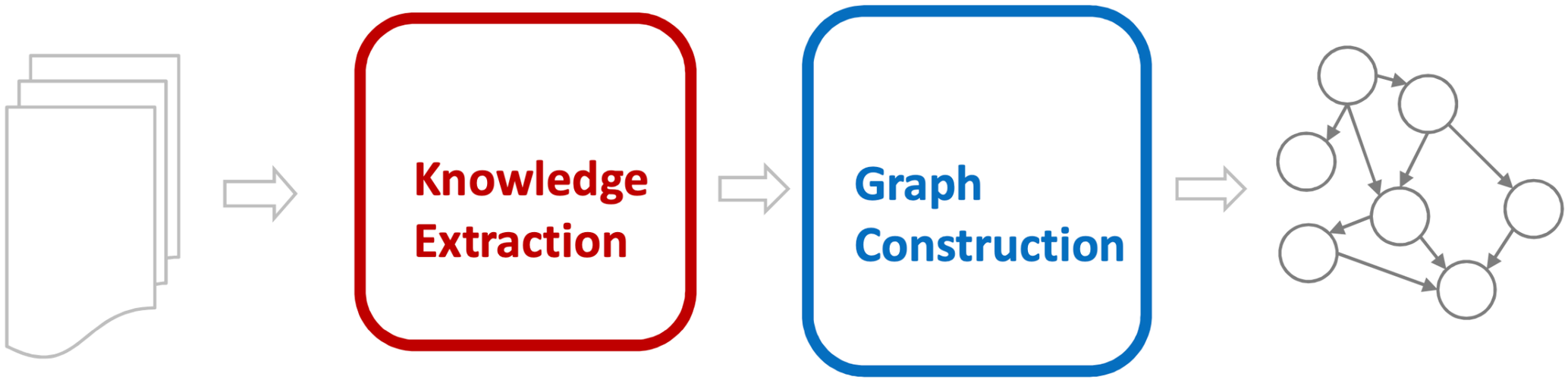
- **Freebase**
  - ~50 million **entities**
  - ~38K **relation types**
  - ~3 billion **facts/triples**

93.8% of persons from Freebase have no place of birth and 78.5% have no nationality!

- **Datasets:** FB15k/FB15k-237
  - A **complete** subset of Freebase, used by researchers to learn KG models

Dataset	Entities	Relations	Total Edges
FB15k	14,951	1,345	592,213
FB15k-237	14,505	237	310,079

KNOWLEDGE GRAPHS



KNOWLEDGE GRAPHS

Knowledge Extraction

- **Who** are the entities (nodes) in the graph?
  - Named Entity Recognition
  - Entity Coreference
- **What** are their attributes and types (labels)?
  - Named Entity Recognition
- **How** are they related (edges)?
  - Relation Extraction
  - Semantic Role Labeling

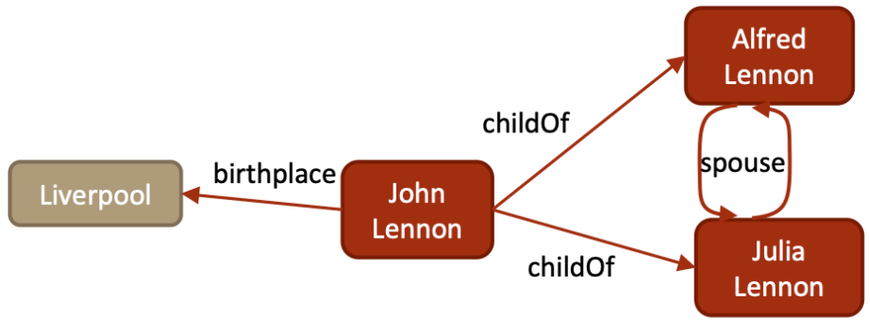
Graph Construction

- **Who** are the entities (nodes) in the graph?
  - Entity Linking
  - Entity Resolution
- **What** are their attributes and types (labels)?
  - Collective Classification
- **How** are they related (edges)?
  - Link Prediction

KNOWLEDGE EXTRACTION

Information Extraction

Entity resolution,
Entity linking,
Relation extraction…



Document

Coreference Resolution…

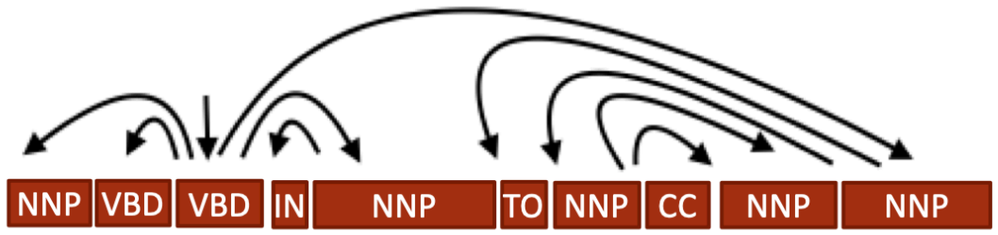
Lennon..
John Lennon…
the Pool
Mrs. Lennon..
.. his mother ..
his father
he
Alfred

Person Location Person Person

John was born in Liverpool, to Julia and Alfred Lennon.

Sentence

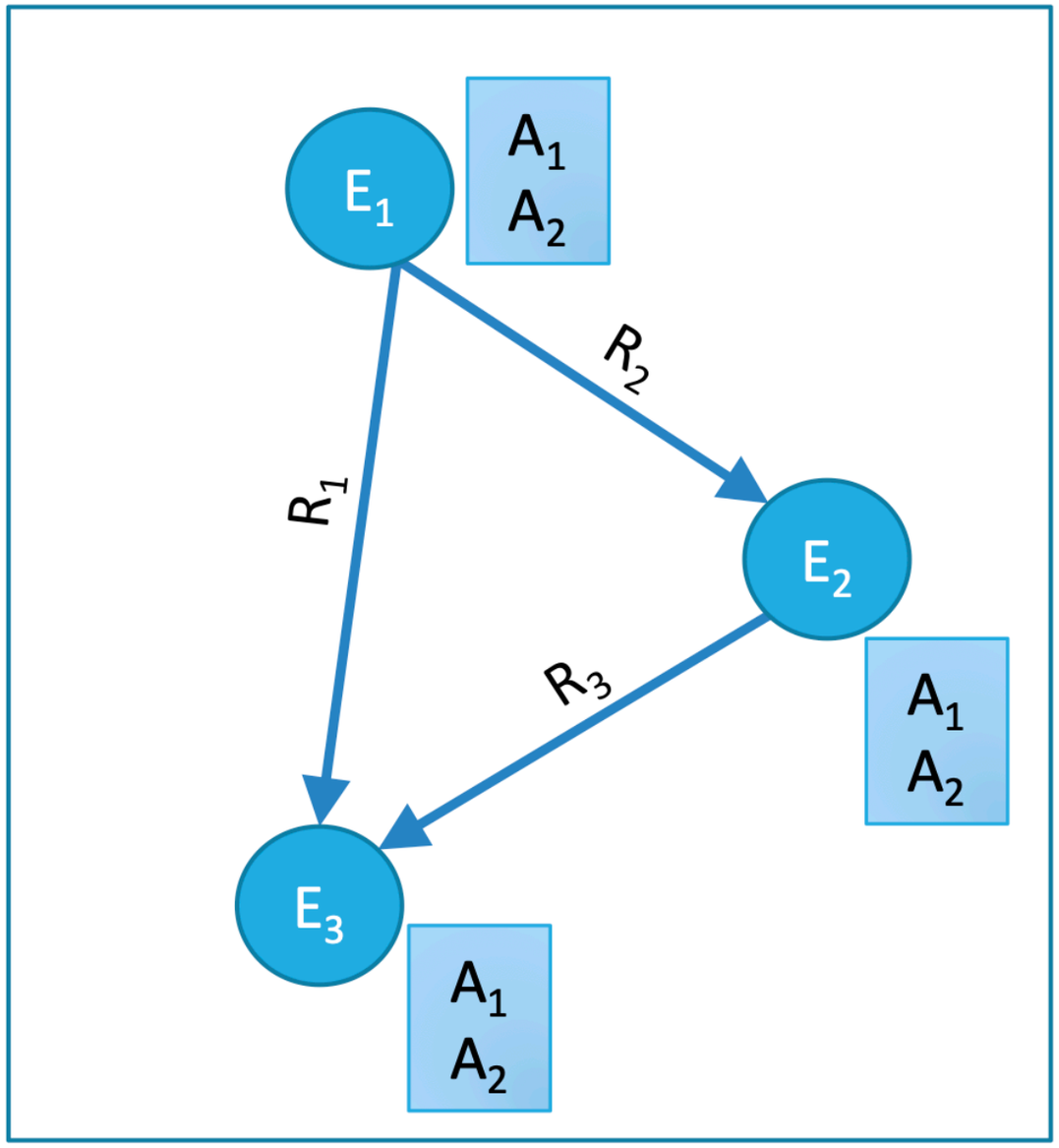
Dependency Parsing,
Part of speech tagging,
Named entity recognition…



John was born in Liverpool, to Julia and Alfred Lennon.

KNOWLEDGE GRAPHS CONSTRUCTION

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?



KNOWLEDGE GRAPHS

- **Publicly available KGs:**
  - FreeBase, Wikidata, Dbpedia, YAGO, NELL, etc.

- **Common characteristics:**
  - **Massive**: millions of nodes and edges
  - **Incomplete**: many true edges are missing

Given a massive KG, enumerating all the possible facts is intractable!

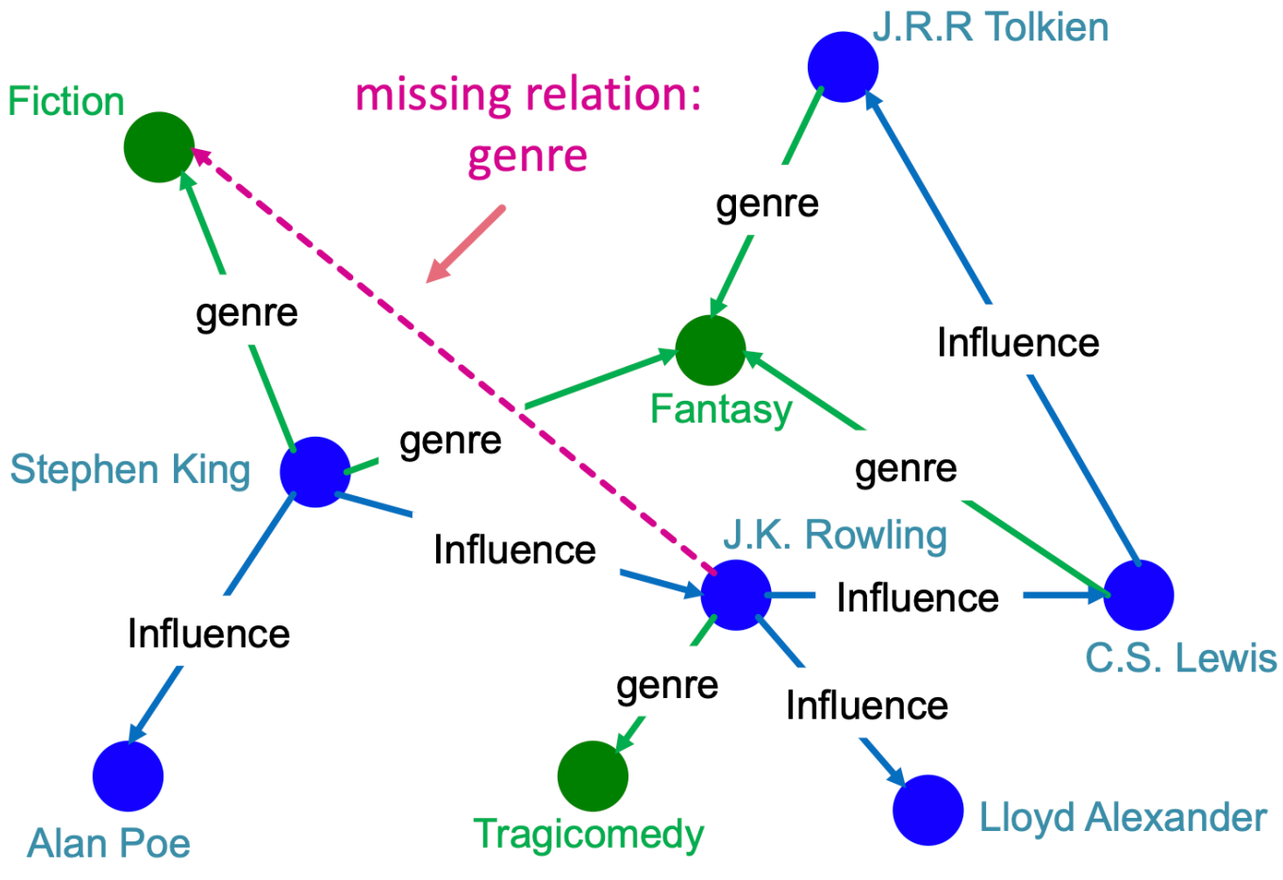
⇨

Can we predict plausible BUT missing links?

KNOWLEDGE GRAPH COMPLETION

- For a given (**head**, **relation**), we predict missing **tails**.
  - (Note this is slightly different from link prediction task)

Example task: predict the tail "Science Fiction" for ("J.K. Rowling", "genre")

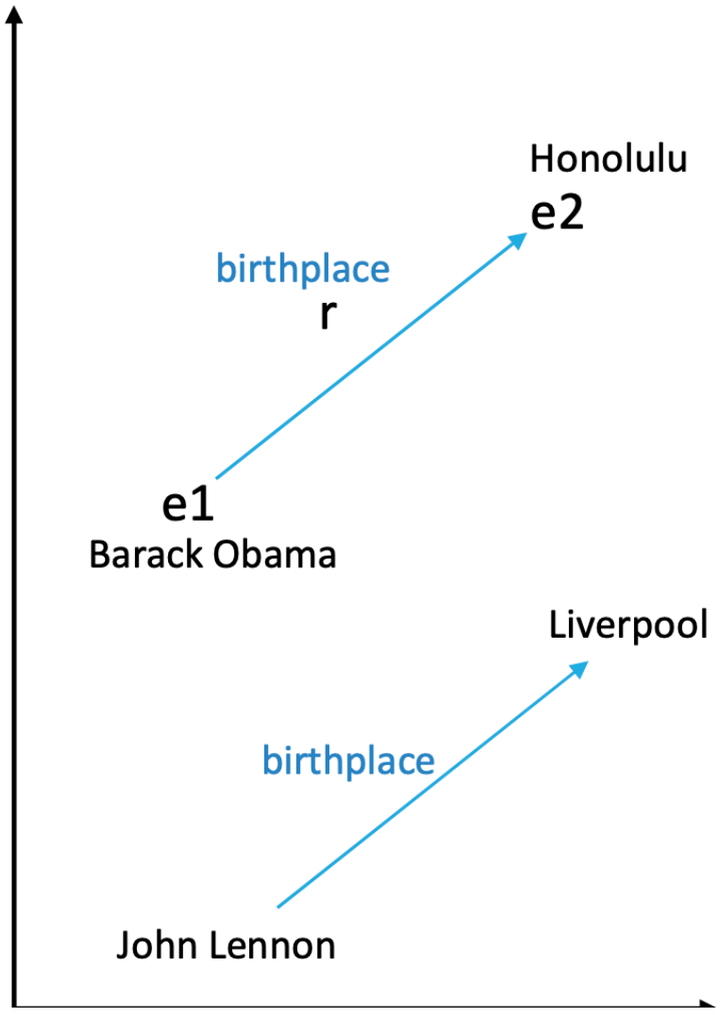


KG EMBEDDINGS

- Edges in KG are represented as **triples** (h, r, t)
  - head (h) has relation (r) with tail (t)
- **Key Idea:**
  - Model entities and relations in the embedding/vector space $\mathbb{R}^d$.

- Given a true triple (h, r, t), the goal is that the embedding of (h, r) **should be close** to the embedding of t.
  - How to embed (h, r)?
  - How to define closeness?

KG EMBEDDINGS



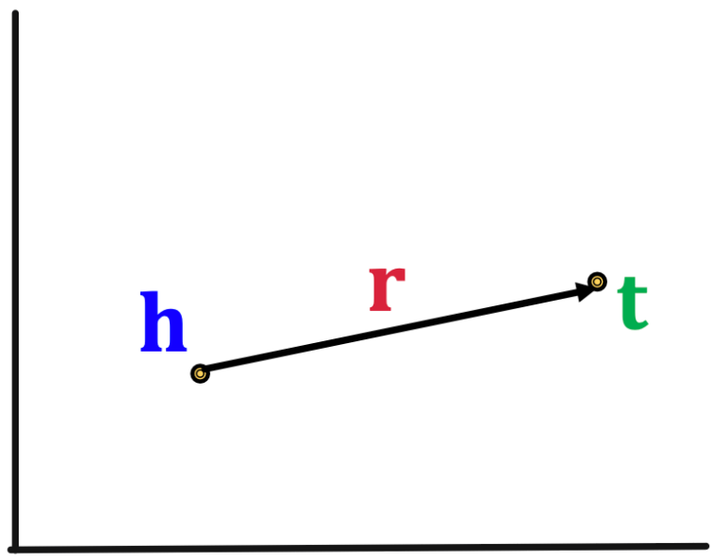
the functional relation induced by the labeled edges corresponds to a translation of the embeddings,

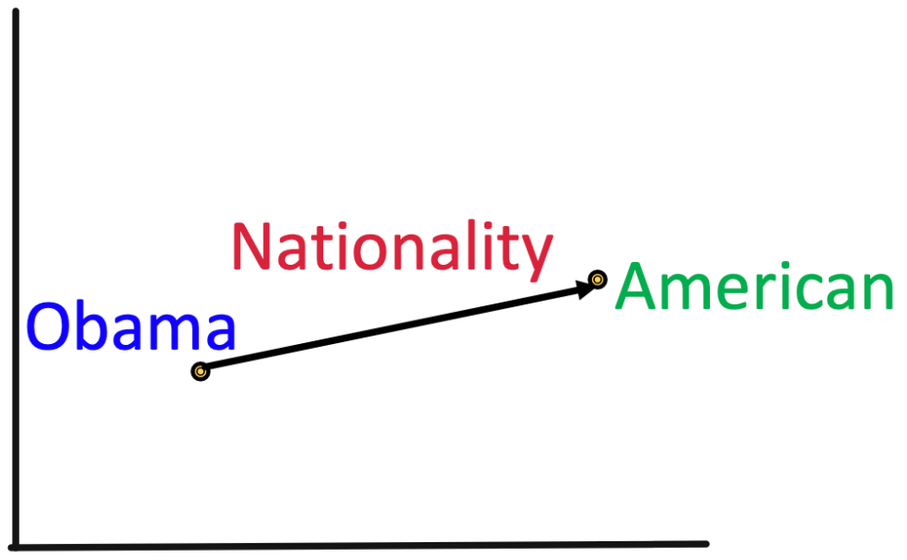
KG EMBEDDING: TRANSE

- **Translation Intuition**:
  For a triple (h, r, t), $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$,
  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ if the given fact is true
  else $\mathbf{h} + \mathbf{r} \neq \mathbf{t}$

embedding vectors will appear in boldface

Scoring function: $f_r(h, t) = -||\mathbf{h} + \mathbf{r} - \mathbf{t}||$





KG EMBEDDING: TRANSE

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

Entities and relations are initialized uniformly, and normalized

loop

  $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

  $S_{batch} \leftarrow$ sample(S, b) // sample a minibatch of size b

  $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

  for $(h, \ell, t) \in S_{batch}$ **do**

    $(h', \ell, t') \leftarrow$ sample$(S'_{(h,\ell,t)})$ // sample a corrupted triplet

Negative sampling with triplet that does not appear in the KG

    $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

  end for

  Update embeddings w.r.t.

$$\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$$

d represents distance (negative of score)

positive sample

negative sample

end loop

Contrastive loss: favors lower distance (or higher score) for valid triplets, high distance (or lower score) for corrupted ones

CONNECTIVITY RELATION PATTERN

:

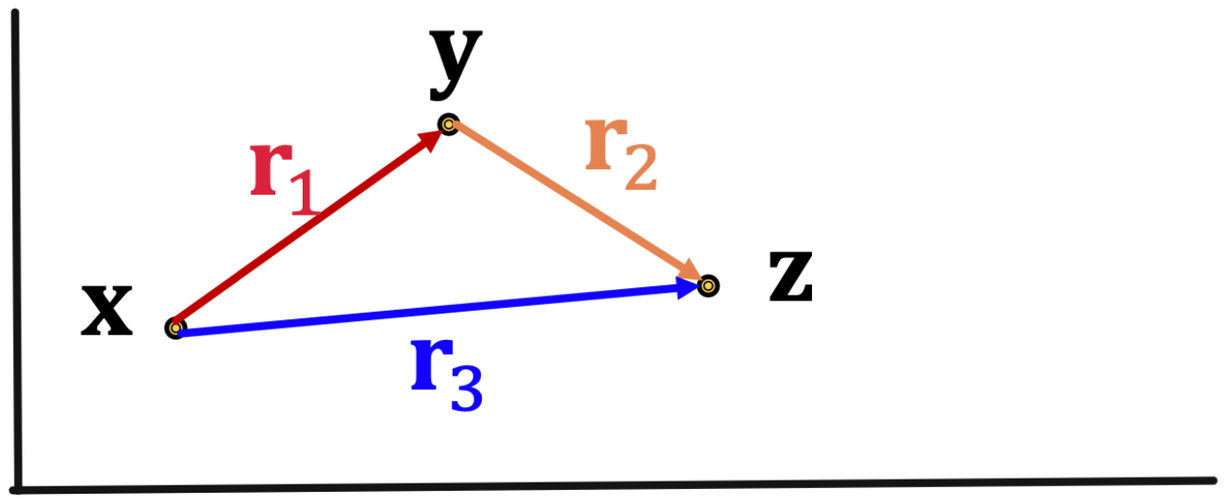
- Symmetric (Antisymmetric): family, roommates $r(h,t) \Rightarrow r(t,h)$
- Inverse Relations: adviser advisee $r_2(h,t) \Rightarrow r_1(t,h)$
- Composition (transitive): family ties $r_1(x,y) \wedge r_2(y,z) \Rightarrow r_3(x,z)$
- 1-N relations $r(h,t_1), r(h,t_2), \ldots, r(h,t_n)$

- **Composition (Transitive) Relations:**

$$r_1(x, y) \wedge r_2(y, z) \Rightarrow r_3(x, z) \quad \forall x, y, z$$

  - **Example**: My mother's husband is my father.
- **TransE** can model composition relations ✓

$$\mathbf{r}_3 = \mathbf{r}_1 + \mathbf{r}_2$$

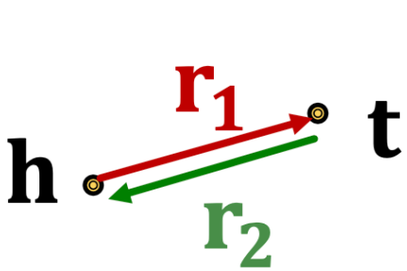


KG EMBEDDING: TRANSE

- **Inverse Relations:**

$$r_2(h,t) \Rightarrow r_1(t,h)$$

  - **Example** : (Advisor, Advisee)
- **TransE** can model inverse relations ✓
  - $\mathbf{h} + \mathbf{r}_2 = \mathbf{t}$, we can set $\mathbf{r}_1 = -\mathbf{r}_2$



KG EMBEDDING: TRANSE

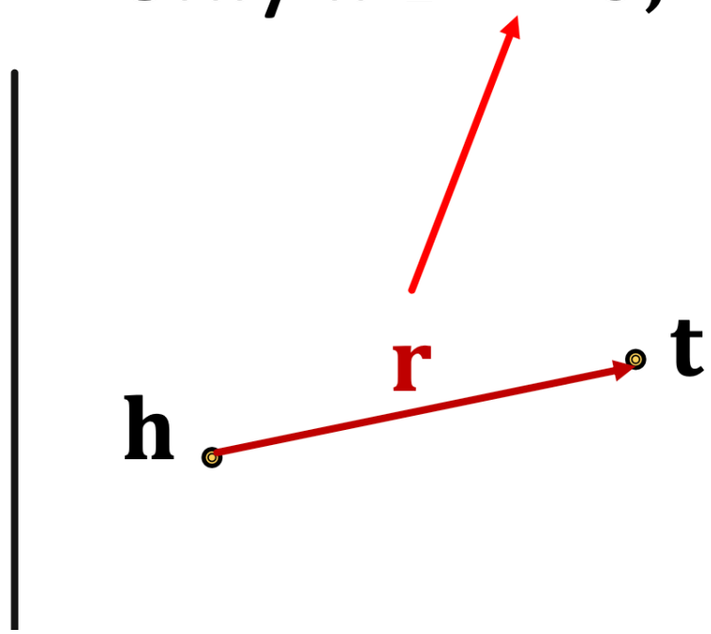
- **Symmetric Relations:**

$$r(h,t) \Rightarrow r(t,h) \quad \forall h,t$$

  - **Example**: Family, Roommate

- **TransE cannot** model symmetric relations ✘

only if $\mathbf{r} = 0,\ \mathbf{h} = \mathbf{t}$



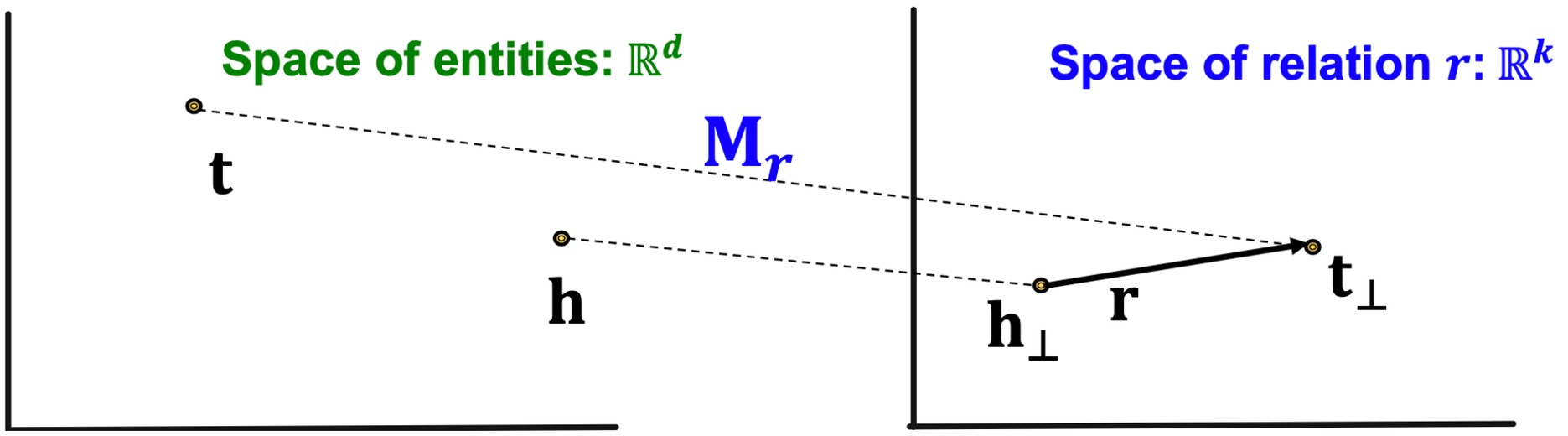
For all h, t that satisfy $r(h,t)$, $r(t,h)$ is also True, which means $\|\mathbf{h}+\mathbf{r}-\mathbf{t}\| = 0$ and $\|\mathbf{t}+\mathbf{r}-\mathbf{h}\| = 0$. Then $\mathbf{r} = 0$ and $\mathbf{h} = \mathbf{t}$, however h and t are two different entities and should be mapped to different locations.

KG EMBEDDINGS: TRANSR

- **TransR**: model **entities** as vectors in the entity space $\mathbb{R}^d$ and model each **relation** as vector in relation space $\mathbf{r} \in \mathbb{R}^k$ with $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ as the **projection matrix**.

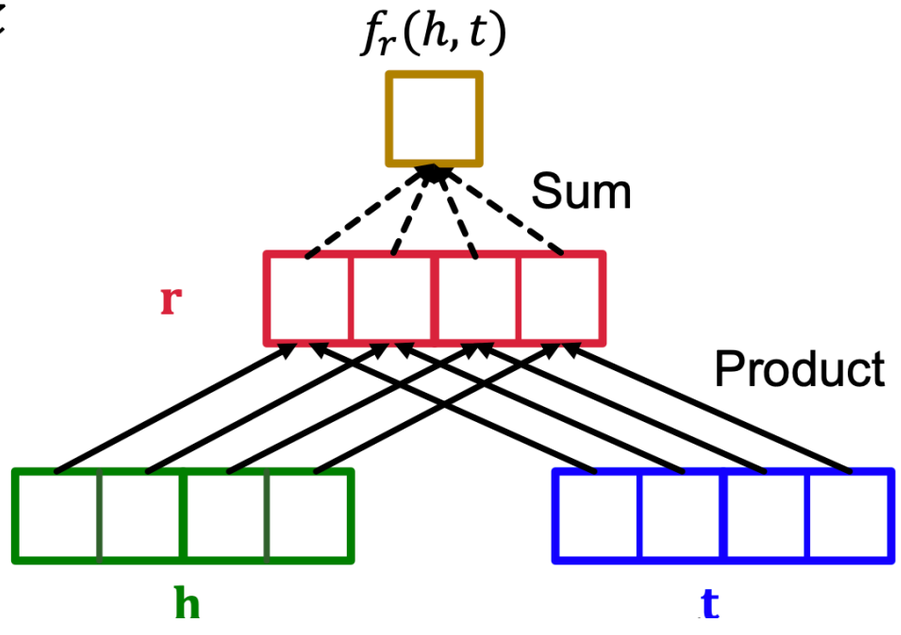
Use $\mathbf{M}_r$ to project from entity space $\mathbb{R}^d$ to relation space $\mathbb{R}^k$!

- $\mathbf{h}_\perp = \mathbf{M}_r\mathbf{h},\ \mathbf{t}_\perp = \mathbf{M}_r\mathbf{t}$
- **Score function**: $f_r(h,t) = -||\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp||$



KG EMBEDDINGS:DISTMULT

- **So far:** The scoring function $f_r(h,t)$ is **negative of L1 / L2 distance** in **TransE** and **TransR**
- Another line of KG embeddings adopt **bilinear** modeling
- **DistMult**: Entities and relations using vectors in $\mathbb{R}^k$
- **Score function:** $f_r(h,t) =< \mathbf{h}, \mathbf{r}, \mathbf{t} > = \sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \mathbf{t}_i$
- $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$



KG EMBEDDINGS COMPARISON

Model	Score	Embedding	Sym.	Antisym.	Inv.	Compos.	1-to-N
TransE	$-\|\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|\|$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{R}^k$	✗	✓	✓	✓	✗
TransR	$-\|\|\mathbf{W}_r\mathbf{h}+\mathbf{r}-\mathbf{W}_r\mathbf{t}\|\|$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{R}^k$, $\mathbf{W}_r\in\mathbb{R}^k$	✓	✓	✓	✗	✓
DistMult	$<\mathbf{h},\mathbf{r},\mathbf{t}>$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{R}^k$	✓	✗	✗	✗	✓
ComplEx	$\mathrm{Re}(<\mathbf{h},\mathbf{r},\bar{\mathbf{t}}>)$	$\mathbf{h},\mathbf{t},\mathbf{r}\in\mathbb{C}^k$	✓	✓	✓	✗	✓

REFERENCES

- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In Advances in neural information processing systems, 2787–2795. 2013.
- Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint, 2014.
- Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In International Conference on Machine Learning, 2071–2080. 2016.
- Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. Holographic embeddings of knowledge graphs. In AAAI, 1955–1961. 2016.
- Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In Procs of AAAI. 2018.

REFERENCES

- Stanovsky, Gabriel, Julian Michael, Luke Zettlemoyer, and Ido Dagan. "Supervised open information extraction." In Proceedings of the 2018 NACL, pp. 885-895. 2018.
- Xiong, Wenhan, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. "One-shot relational learning for knowledge graphs." arXiv preprint arXiv:1808.09040. 2018.
- Bosselut, Antoine, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. "COMET: Commonsense transformers for automatic knowledge graph construction." arXiv preprint arXiv:1906.05317. 2019.
- Liu, Zhibin, Zheng-Yu Niu, Hua Wu, and Haifeng Wang. "Knowledge aware conversation generation with explainable reasoning over augmented graphs." arXiv preprint arXiv:1903.10245. 2019.
- Al-Moslmi, Tareq, Marc Gallofré Ocaña, Andreas L. Opdahl, and Csaba Veres. "Named entity extraction for knowledge graphs: A literature overview." IEEE Access 8, p. 32862-32881. 2020.

NATIONAL RESEARCH
UNIVERSITY

www.text Phone: +X (XXX) XXX XXXX Address: TextTextTextTextTextTextTextTextTextTextTextTextText