Chasing the Tail of Self Supervised Named Entity Disambiguation

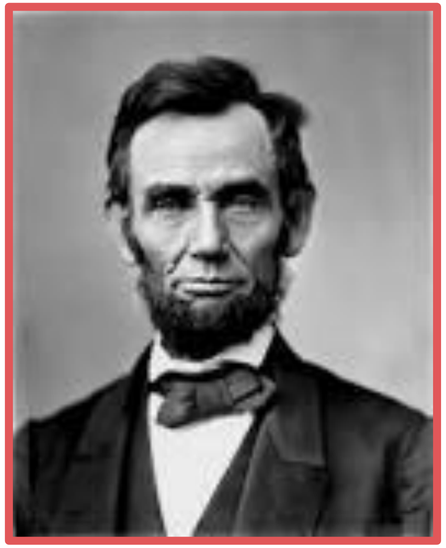
Knowledge Graphs Seminar
April 2021

Laurel Orr,
Megan Leszczynski,
Neel Guha,
Sen Wu,
Simran Arora,
Xiao Ling (Apple),
Chris Ré

Named Entity Disambiguation

Map "strings to things" in a knowledge base

How tall is Lincoln?



Q91

Key part of assistant, search, and information extraction

The Long Tail of NED

HEAD

Washington, DC
Q61

TORSO

Chevrolet, Corvette
Q56166

TAIL

Reddish Potato Beetle
Q14934552

UNSEEN

Sauce! by XXXTENTACION
Q???

Popular

Rare

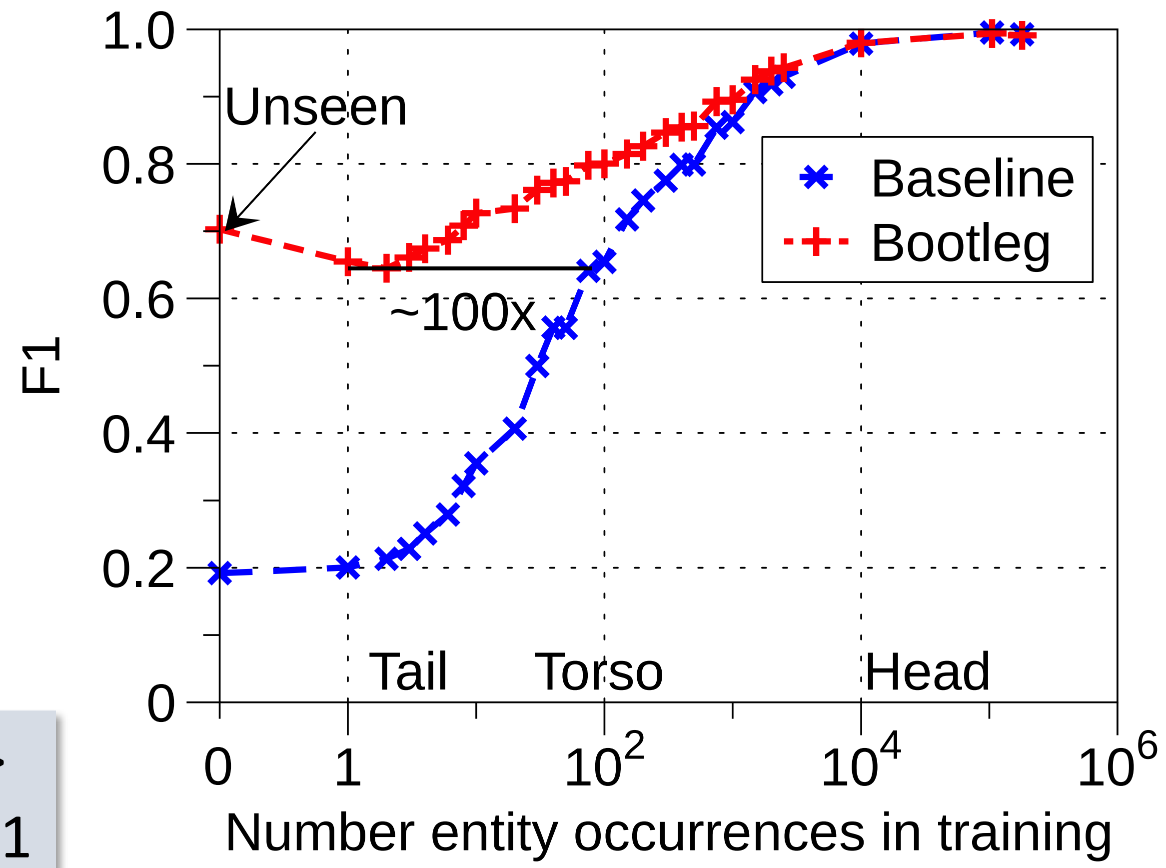
IR and BERT/GPT work great for the HEAD.

The majority of entities are rare!
13% entities have Wikipedia page.
< 1% of songs in Wikidata!

Tail Scalability Challenge

Large number of patterns needed to resolve the tail, making it difficult to scale a system that can learn the patterns.

But the majority of entities are rare (tail).



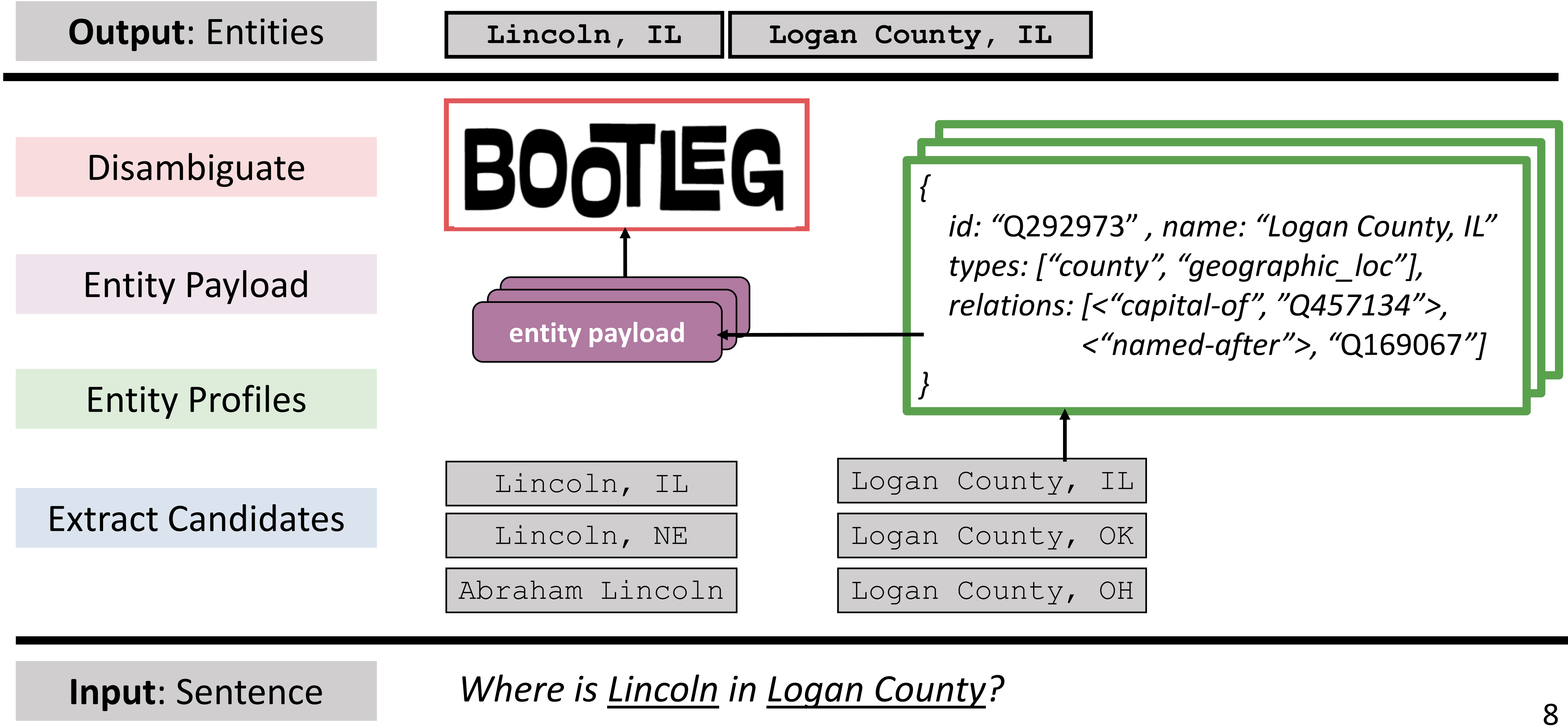
90 million entities in Wikidata -> 90*100 million examples for 60 F1

Subtle reasoning clues are needed for the tail!
(+40 F1 points by encoding these reasoning patterns)

How do you disambiguate the tail?

(Ans: structural information over types and relations)

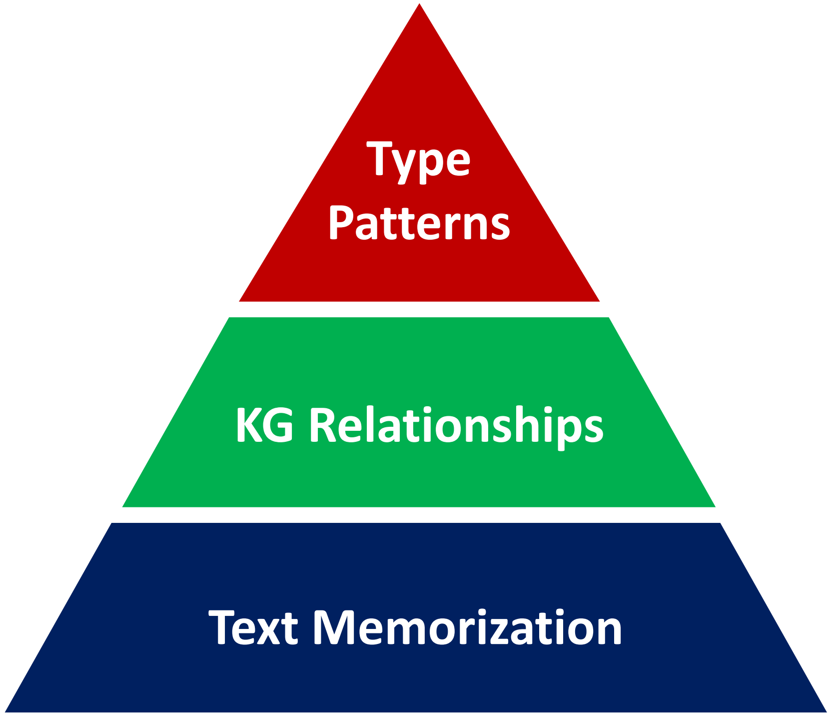
Disambiguation Input & Output



Signals of NED

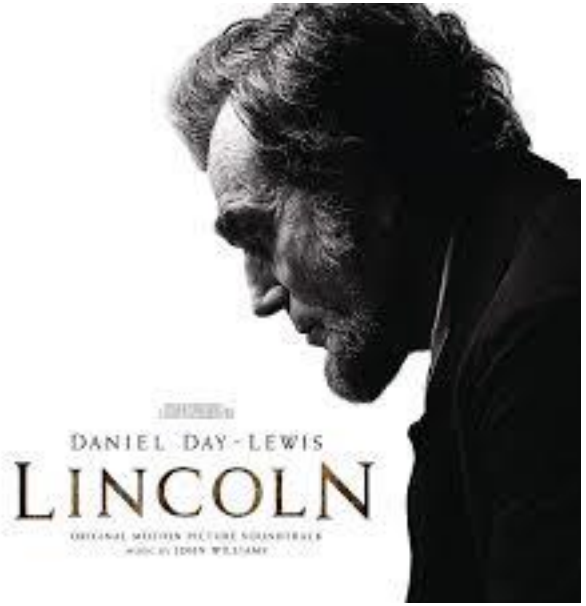
Want to learn **information at level** of

- Individual entities
- Relationships
- Types



Text Memorization

Where is Lincoln in Nebraska



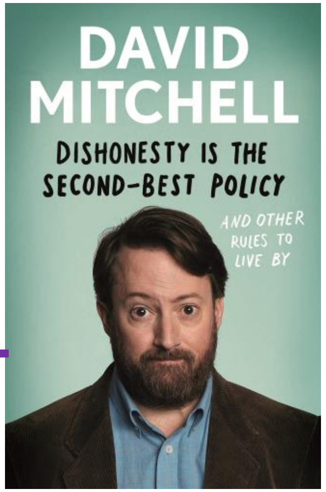
What year did Steven Speilberg make Lincoln?

Associate phrases per entity – but this doesn't generalize to other entities

Knowledge Graph Relations

Victoria Mitchell
(poker player, writer)

spouses



David Mitchell

Victoria Mitchell
(runner)

David and Victoria **Mitchell** added spice to their marriage

Associate phrases per relation and use relations to disambiguate – generalizes to rare entities with the same relation

Types Patterns

How tall is Lincoln?

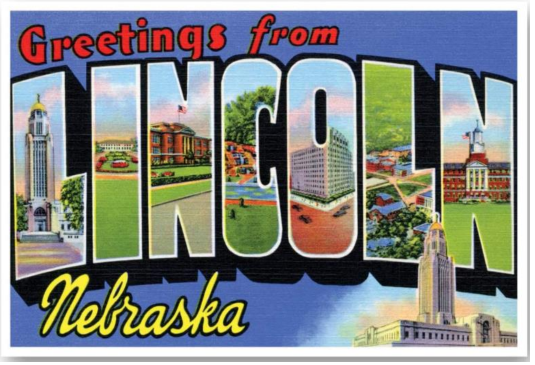
People have heights, not places or brands

What is the cheapest Lincoln?

LINCOLN

Brands have prices, not places or people

How many people are in Lincoln?



Places have populations, not people or brands

Associate phrases per entity type – generalizes to rare entities of the same type

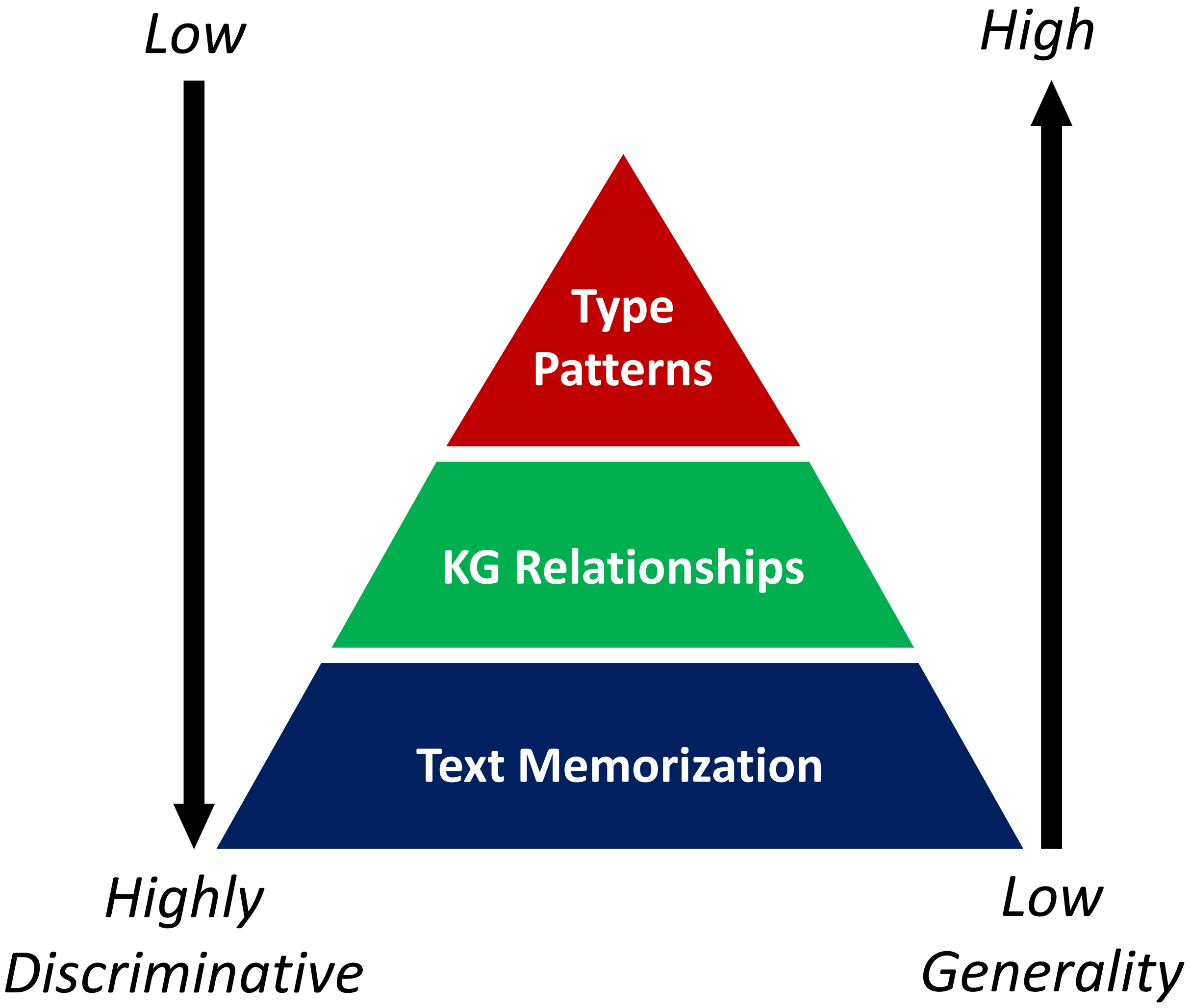
Signals of NED

Want to learn **information at level** of

- Individual entities
- Relationships
- Types

In Wikipedia sample

- 22% of entities have a relationship
  - 126 relationships occur more than 100 times
- 96% of entities have types
  - 2.8K occur more than 100 times

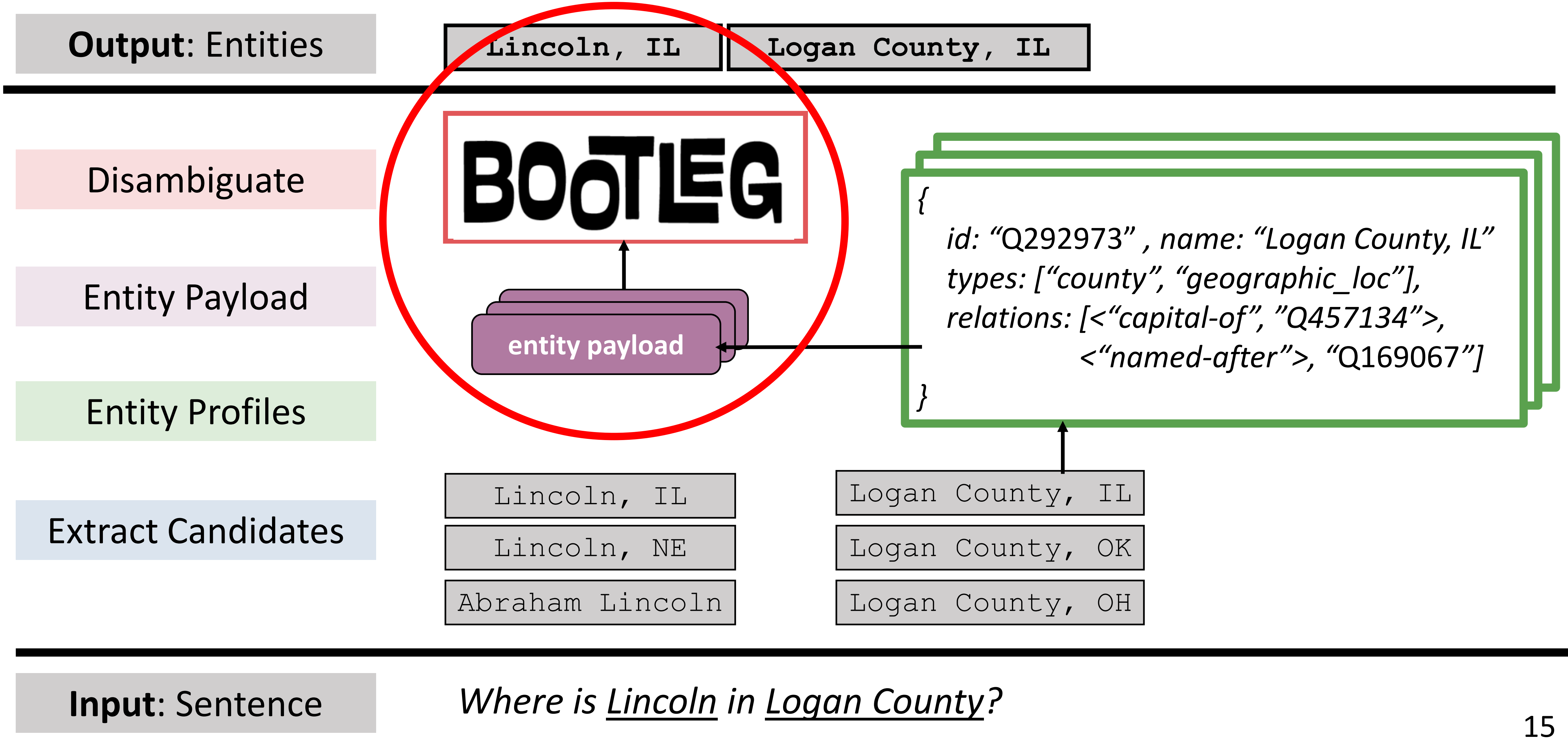


Takeaway: learning patterns of types and relations will help generalize to the tail

How do you learn the entity, type, and KG signals?

(Ans: using embeddings)

Disambiguation Input & Output



Disambiguation Input & Output

Entity Payload

Entity Profiles

{
id: “Q292973” , name: “Logan County, IL”
types: [“county”, “geographic_loc”],
relations: [<“capital-of”, ”Q457134”>,
<“named-after”>, “Q169067”]
}

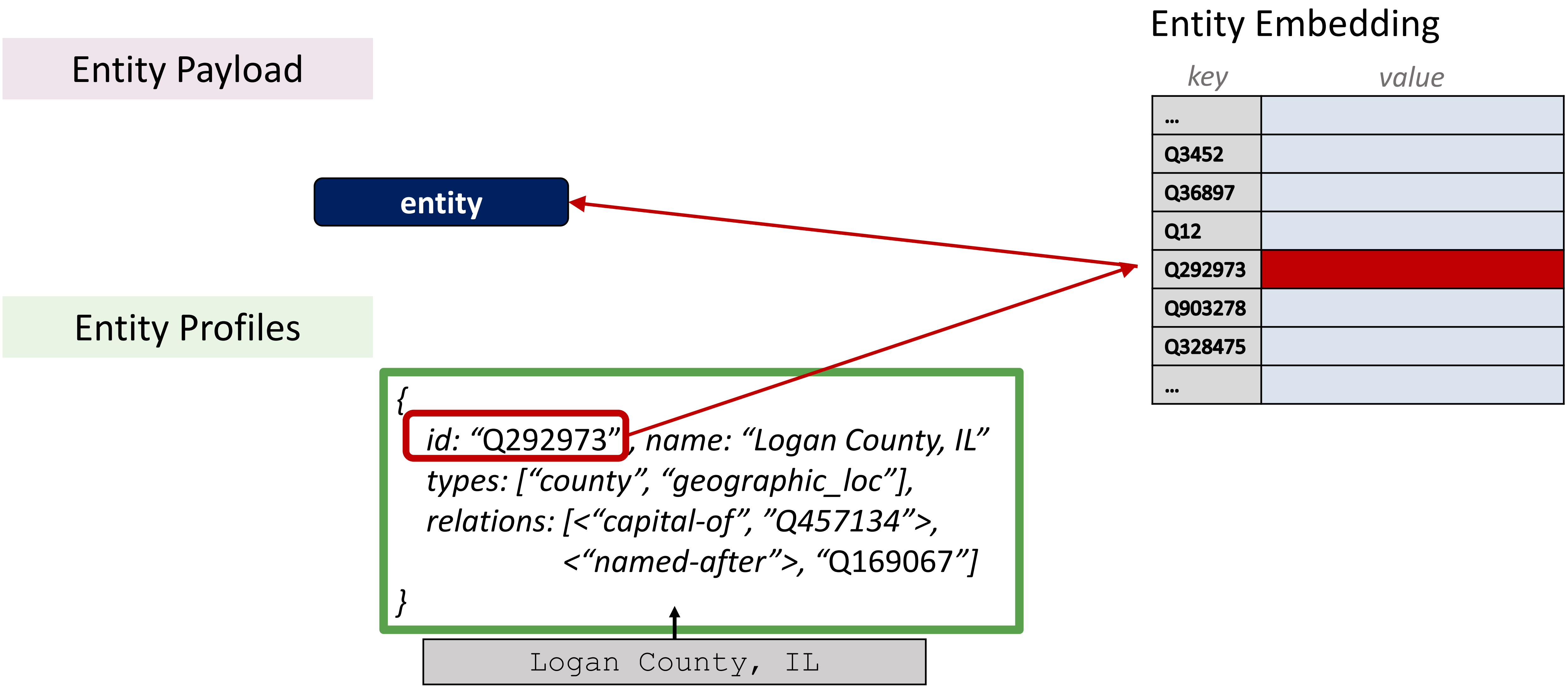
`Logan County, IL`

Entity Embedding

key	*value*
…	
Q3452	
Q36897	
Q12	
Q292973	
Q903278	
Q328475	
…	

For each candidate, we use the entity profile to extract (learned) embeddings.

Disambiguation Input & Output

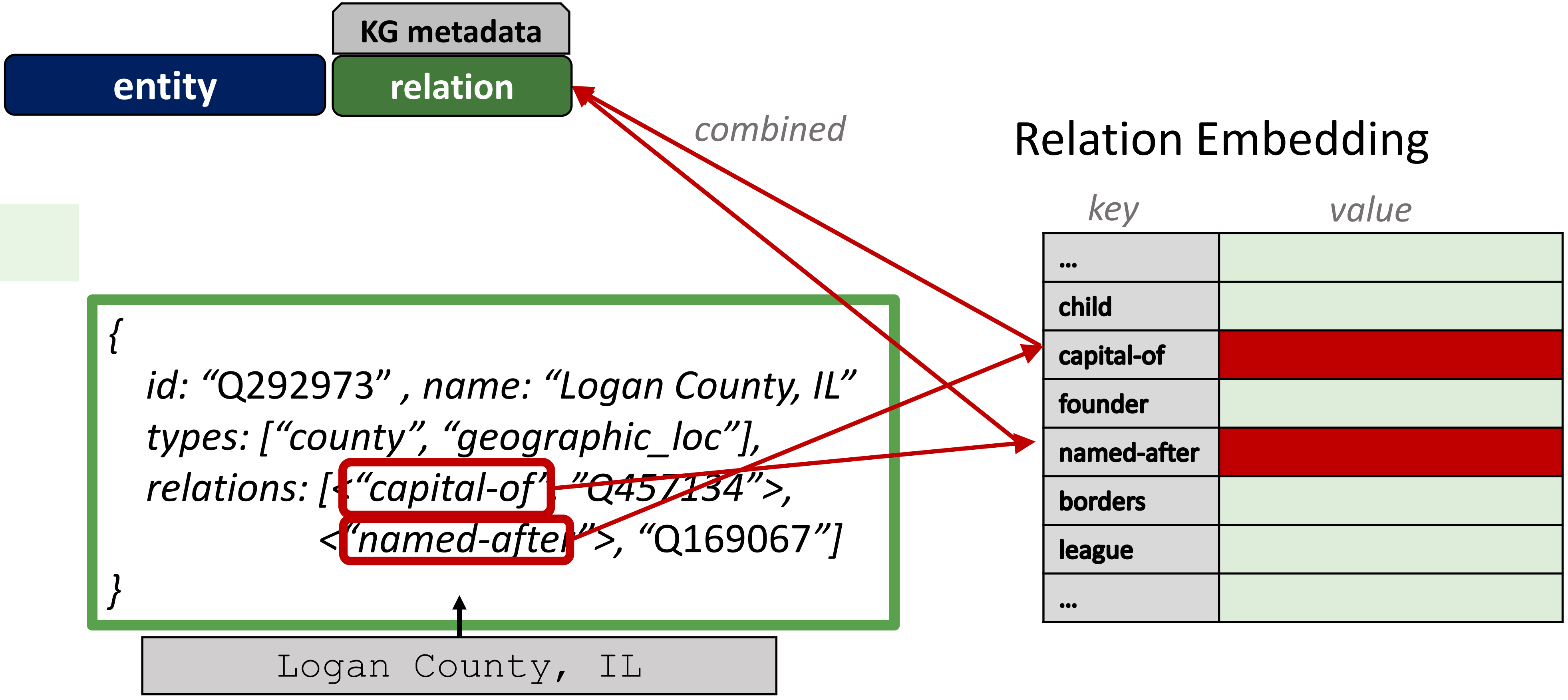


For each candidate, we use the entity profile to extract (learned) embeddings.

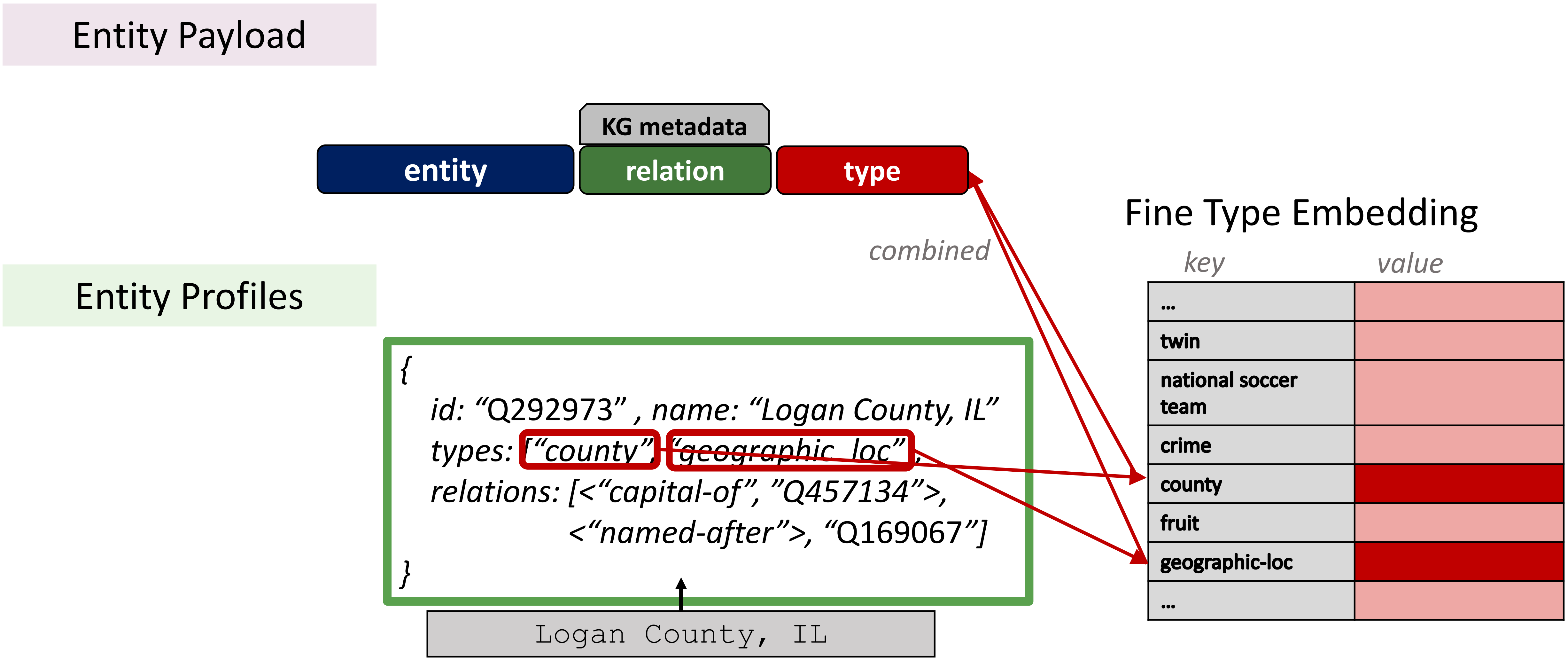
Disambiguation Input & Output

Entity Payload

Entity Profiles



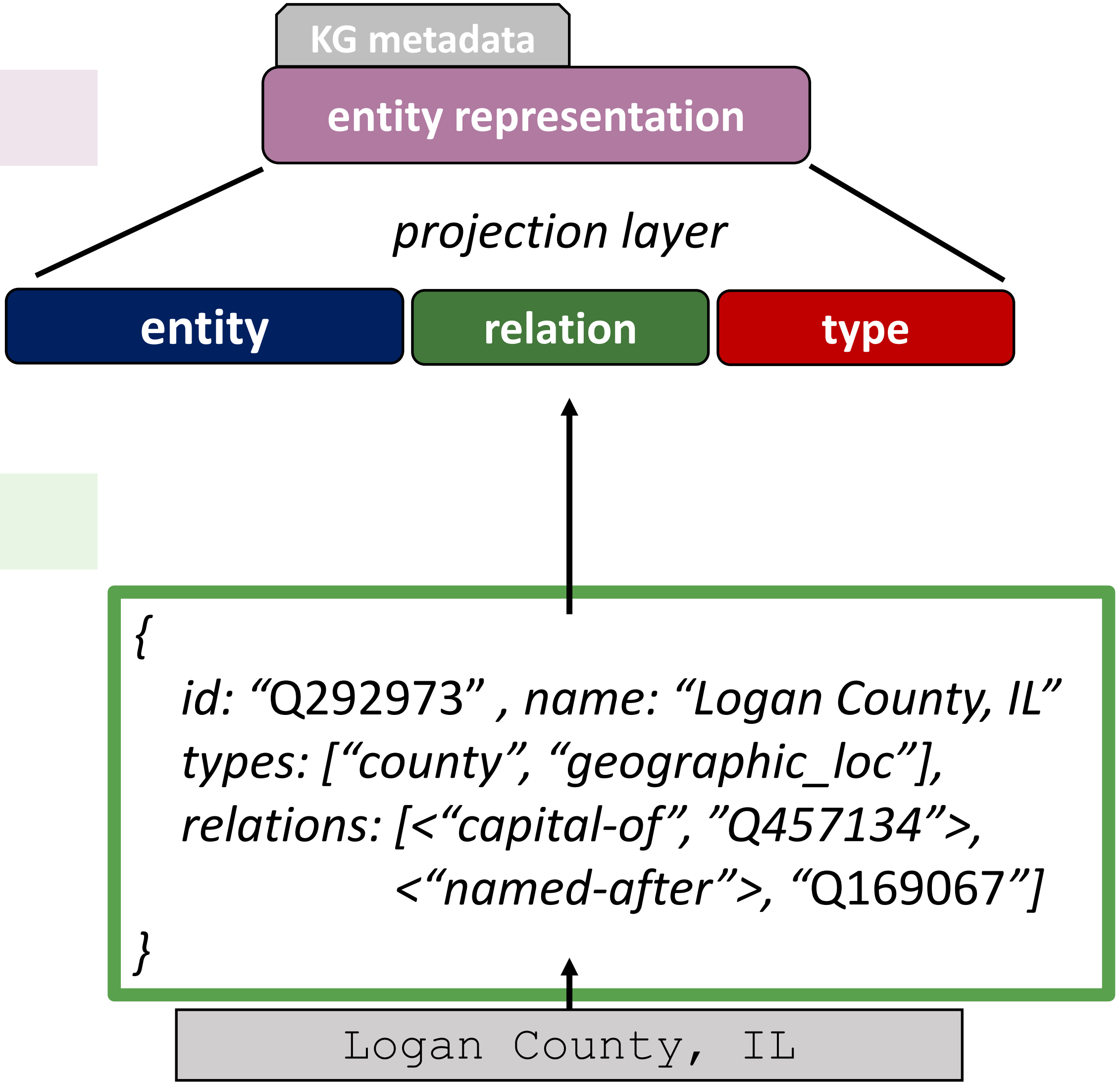
Disambiguation Input & Output

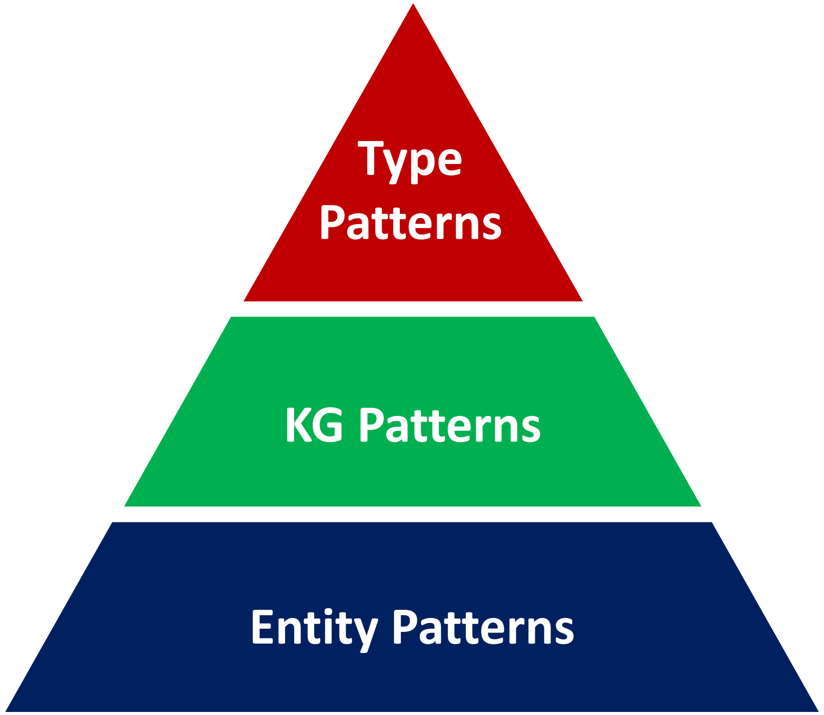


Disambiguation Input & Output

Entity Payload

Entity Profiles

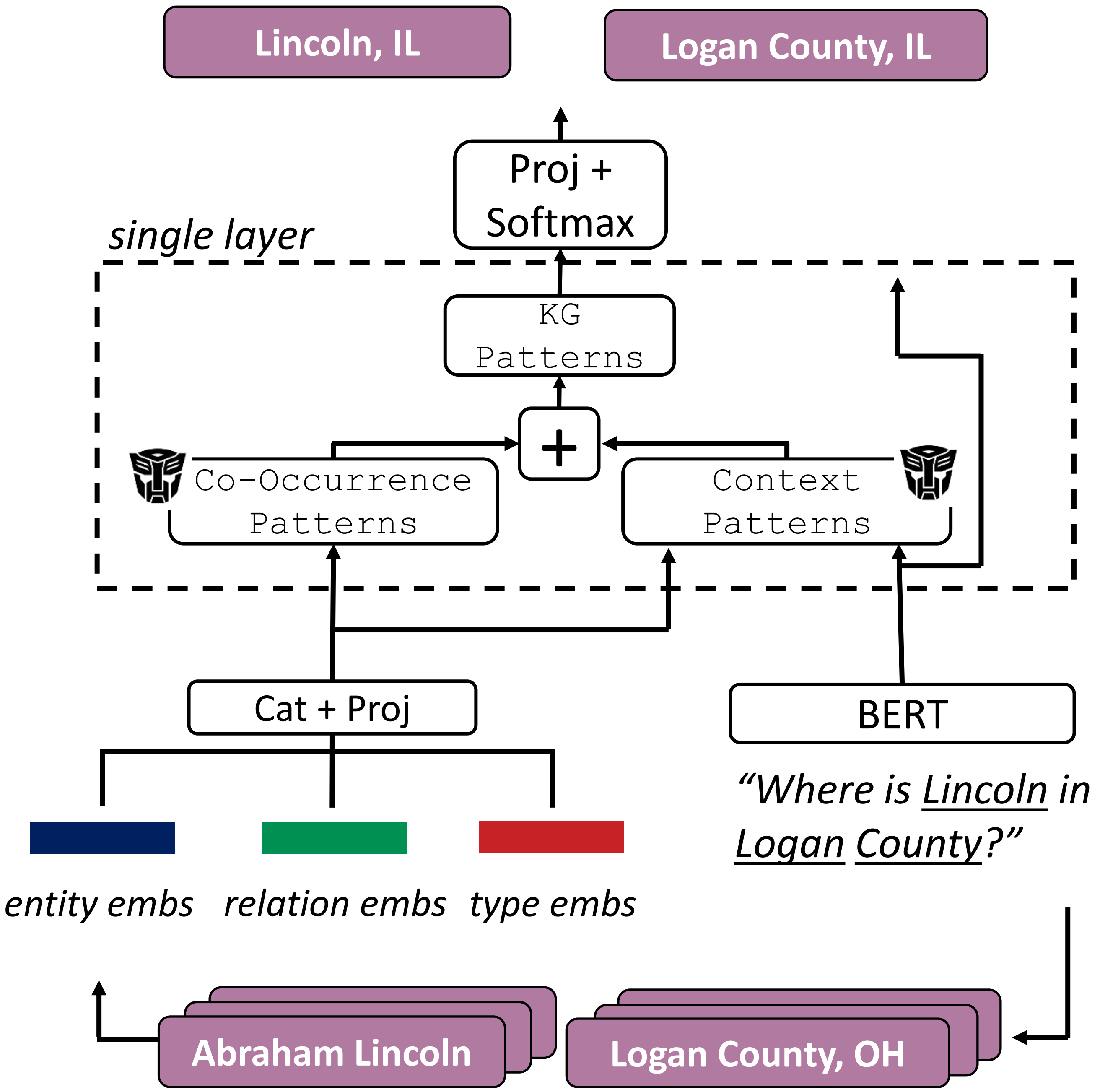




The entity payload has embeddings mapping to each level of the hierarchy.

Bootleg Architecture

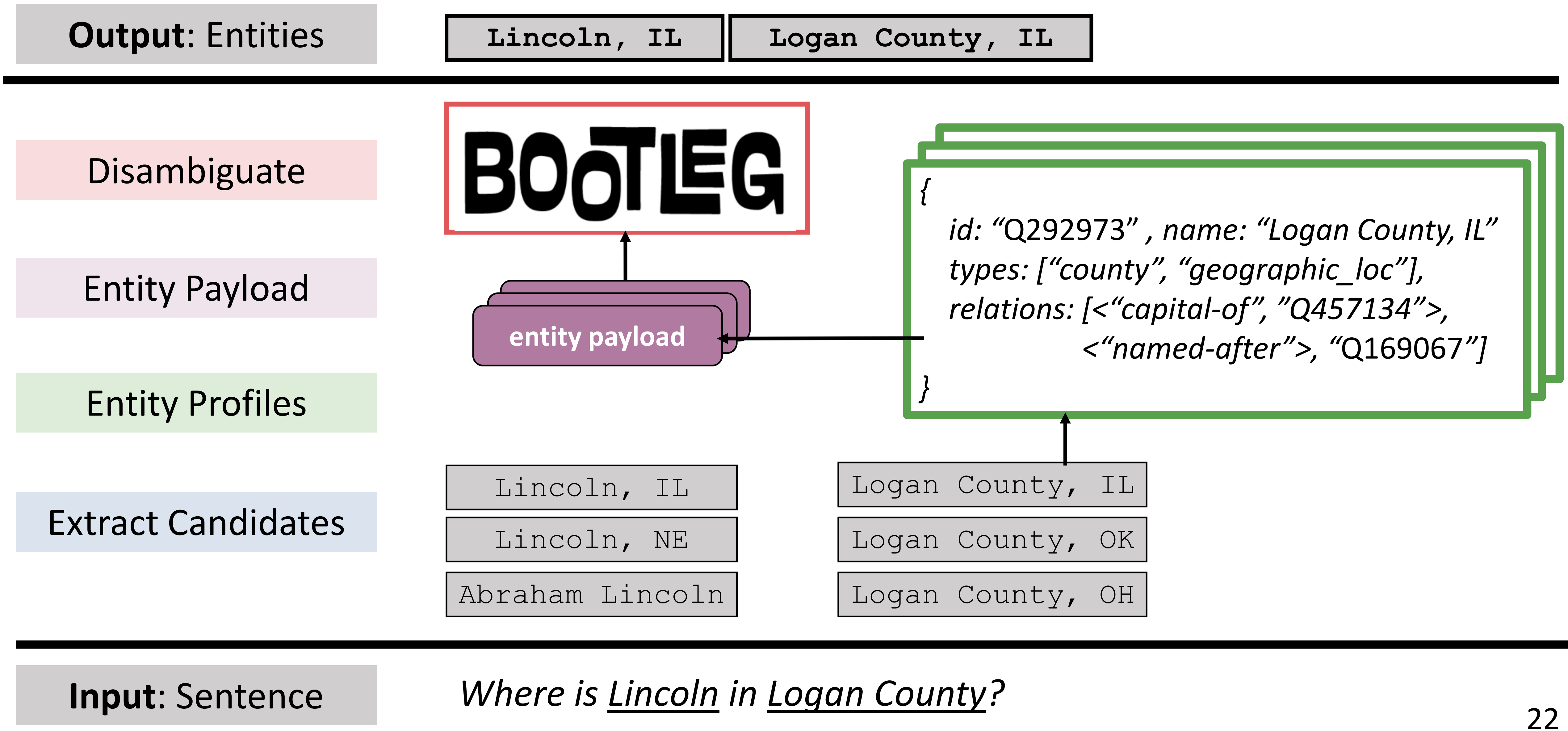
Use simple Transformer building blocks



Architecture supports reasoning over each pattern

KG module allows for related entities to transfer representation

Disambiguation Input & Output



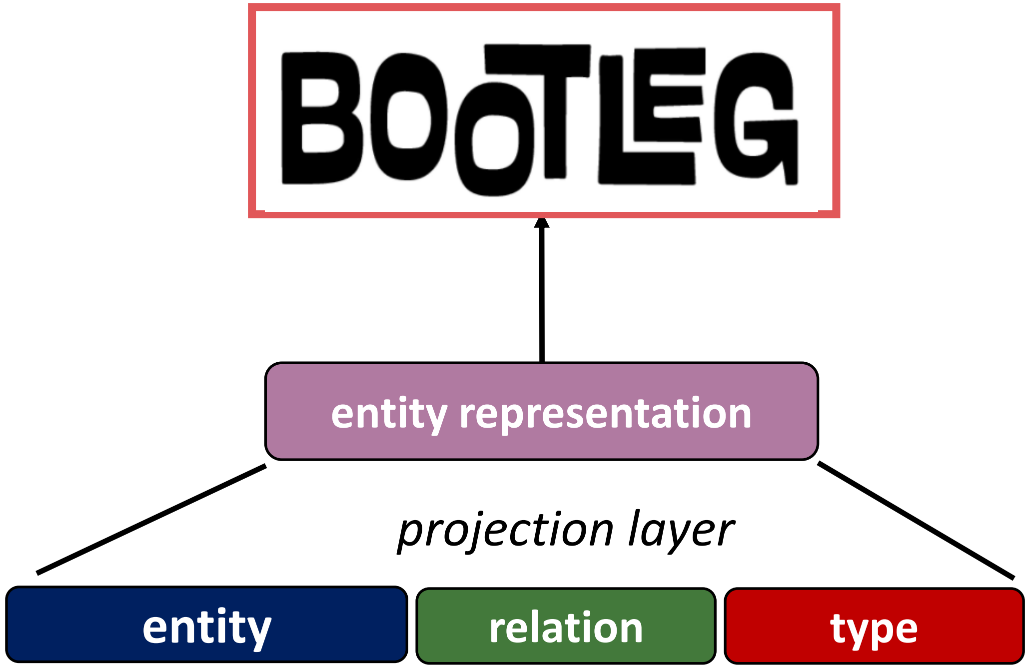
Training: how do you encourage learning structural patterns?

(Ans: regularization and training set refinement)

Inverse Hierarchical Regularization

Want an inductive bias to push toward the tail and more general signals.

But we do not want to lose performance over popular entities.

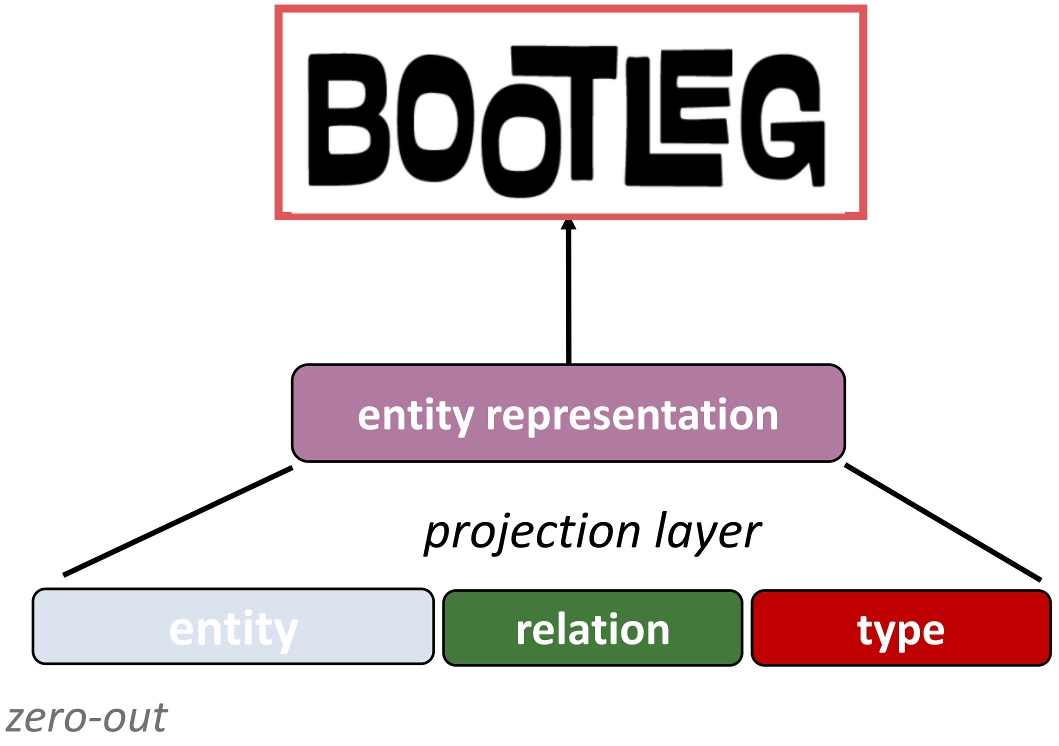


Inverse Hierarchical Regularization

Want an inductive bias to push toward the tail and more general signals.

But we do not want to lose performance over popular entities.

Regularize the ***entire*** entity embedding proportional to inverse popularity



Regularization scheme gains 13.6 F1 points on unseen entities

"Hands-Free" Training

Need self-supervised training data that does not have hand tuned features

WIKIDATA

band type

country of origin

Lincoln (Band)

Q6550456

Add structured data from Wikidata

WIKIPEDIA
The Free Encyclopedia

Although they toured together, Lincoln did not take its name from They Might Be Giants' album *Lincoln*.

Train by "reading" from Wikipedia

Using Wikipedia and Wikidata is easy to maintain with limited engineer effort and is vastly easier to extend to new languages.

Use Weak Supervision to Refine Labels

Wikipedia is known to be sparsely labeled and missing mentions. However, Wikipedia has internal structure: most sentences on an entity's Wikipedia page are referring to that entity.

Lincoln (Band)

Q6550456

WIKIPEDIA
The Free Encyclopedia

Although they toured together, Lincoln did not take its name from They Might Be Giants' album *Lincoln*.

Use Weak Supervision to Refine Labels

Wikipedia is known to be sparsely labeled and missing mentions. However, Wikipedia has internal structure: most sentences on an entity's Wikipedia page are referring to that entity.

WIKIPEDIA
The Free Encyclopedia

Lincoln (Band)
Q6550456

Although they toured together, Lincoln did not take its name from They Might Be Giants' album *Lincoln*.

Lincoln (Album)
Q1421879

Add missing labels by heuristic label functions

Simple heuristics add ~1.7x training data and give 2 F1 point improvement over unseen entities

Experiments

Bootleg outperforms existing SotA systems on NED benchmarks.

Benchmark	SotA system	SotA F1	Bootleg
RSS500	Phan et al., 2019	82.3	**82.5**
KORE50	Hu et al., 2019	79.9	**85.7**
AIDA CoNLL YAGO	Fevry et al., 2020	**96.7***	**96.7***

*SotA model reports test accuracy (v. F1), so for comparison we evaluate Bootleg on test accuracy for AIDA CoNLL YAGO.

But: Tail Performance is the game!

Wikipedia Dataset.

On the head, BERT-based baseline performs ~ 5 F1 points of Bootleg.

On the tail, Bootleg outperforms baseline by > 40 F1 points!

Evaluation Set	BERT NED Baseline	Bootleg	# Examples
All	85.9	**91.3**	4,066K
Torso Entities	79.3	**87.3**	1,912K
Tail Entities	27.8	**69.0**	163K
Unseen Entities	18.5	**68.5**	10K

Overall F1 can be really misleading!

Understanding the Performance Lift

Mined for subpopulations of examples where a reasoning pattern is present for the correct entities.

On tail, Bootleg demonstrates ability to capture the patterns.

Subpopulation where Signal is Present	BERT NED Baseline (Head/Tail)	Bootleg (Head/Tail)	Coverage (Head/Tail)
Entity	59 / 29	**66 / 47**	0.7% / 3.3%
Type	87 / 28	**93 / 73**	84% / 76%
KG	91 / 30	**98 / 92**	27% / 23%

The reasoning patterns are how to improve tail performance

Production Task

Included Bootleg embeddings into an Overton production task answering millions of users' factoid queries. We report relative lift.

Evaluation Set	English	Spanish	French	German
All Entities	1.08	1.03	1.02	1.00
Tail Entities	1.08	1.17	1.05	1.03

See 8-17% lift in production task, even over multiple languages.

Using Bootleg Downstream: SoTA on the TACRED Benchmark

Goal: extract the relationship between a subject and object pair.

Gold relation: per:employee_of

(subject) (object)

Mays worked with several other companies aside from **Media Enterprises** in his career.

Micro-Avg. F1 on TACRED Revised test dataset:

Model	Test F1 Score
SpanBERT	78.0
KnowBERT	79.3
Bootleg+SpanBERT	80.2 (SoTA)

Zhang et al., 2017 and Hennig et al., 2020.

Bootleg resolves errors made in by the prior SoTA model.

Gold relation: per:cause_of_death

(subj) (obj)

Vincent Astor , like **Marshall** , died unexpectedly of a **heart attack** in 1959 …

Q299001 Q12152

Wikidata relation: ['cause of death']

SpanBERT no_relation

BOOTLEG per:cause_of_death

Leveraging type and relation information downstream

Gold relation: org:alternate_names

(obj) (subj)

The **International Water Management Institute** or **IWMI** study said both …

Q868028 Q868028

Wikidata same entity

SpanBERT no_relation

BOOTLEG org:alternate_names

Understand that sub-strings relate to the same entity

Bootleg:

- Clean slate, simple, **open source** NED

Key ideas:

- Leverages a simple hierarchy to learn patterns that generalize to the tail
- Patch errors in the self-supervised model

https://hazyresearch.stanford.edu/bootleg/

lorr1@cs.stanford.edu

BOOTLEG