DeepStruct: Pretraining of Language Models for Structure Prediction

ACL 2022

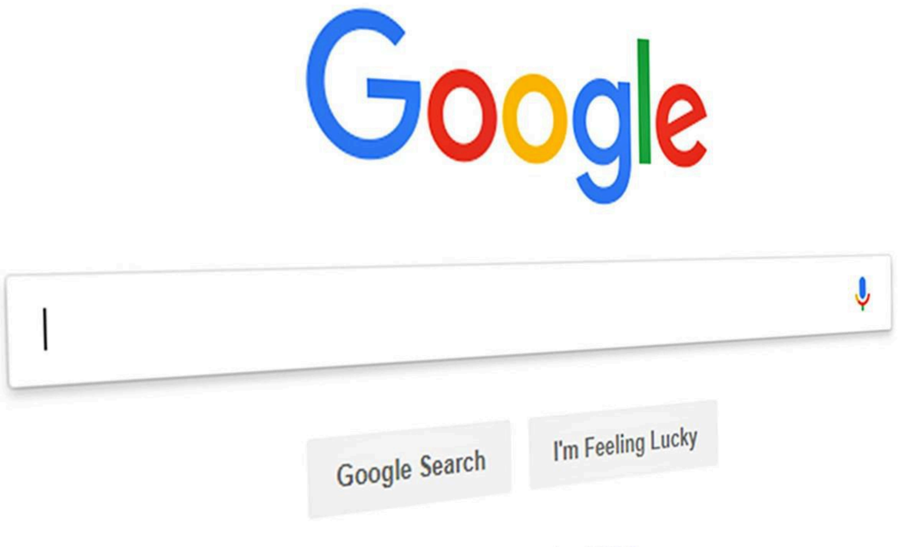
Chenguang Wang*, Xiao Liu*, Zui Chen*, Haoyun Hong, Jie Tang, Dawn Song

Structure prediction is important

Structure prediction has a wide range of applications in NLP area

Structure prediction is important

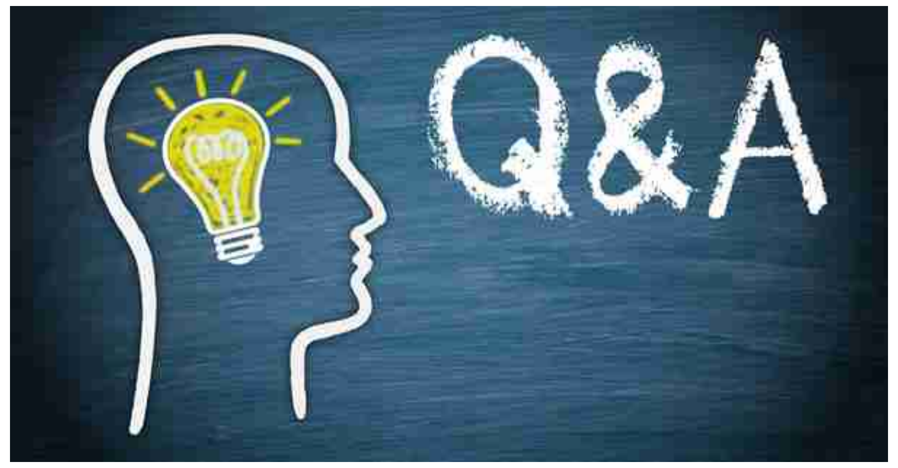
NLP Applications



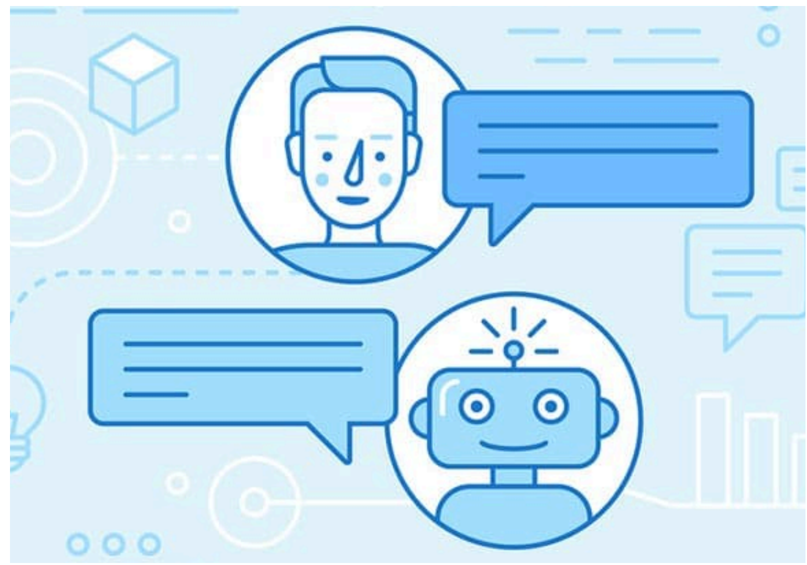
Search Engine

WIKIPEDIA
The Free Encyclopedia

Knowledge Base



QA System

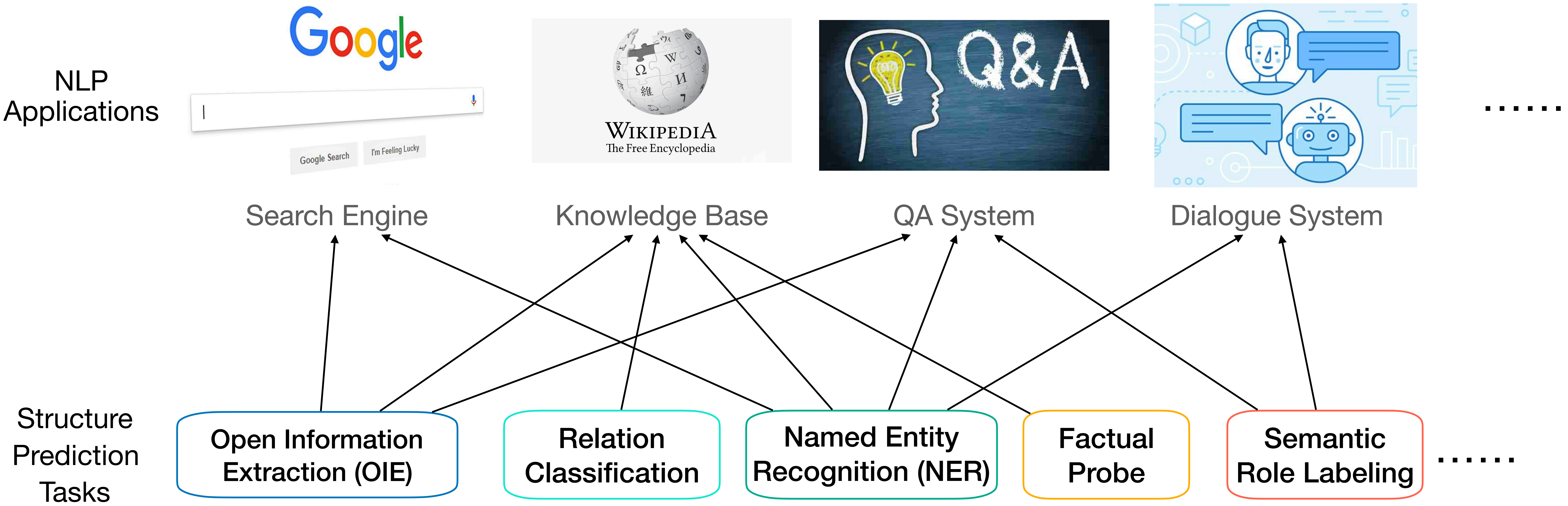


Dialogue System

……

Structure prediction has a wide range of applications in NLP area

Structure prediction is important



Structure prediction has a wide range of applications in NLP area

Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Structure prediction tasks could have flexible output formats

Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition (NER)

Born in 1951 in Tbilisi, Iago is a Georgian artist. ← predict entities

city person

Structure prediction tasks could have flexible output formats

Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition (NER)

Born in 1951 in Tbilisi, Iago is a Georgian artist. ← predict entities

city person

Joint-entity relation extraction (JER)

Born in 1951 in Tbilisi, Iago is a Georgian artist. ← predict entities and relations

city_of_birth city person

Structure prediction tasks could have flexible output formats

Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition (NER)

Born in 1951 in Tbilisi, Iago is a Georgian artist. ← predict entities

city person

Joint-entity relation extraction (JER)

Born in 1951 in Tbilisi, Iago is a Georgian artist. ← predict entities and relations

city_of_birth city person

Open information extraction (OIE)

Born in 1951 in Tbilisi, Iago is a Georgian artist. ← predict triples

Structure prediction tasks could have flexible output formats

Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Structural Understanding

Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Structural Understanding

Next Word Prediction

Born in 1951 in Tbilisi,
Iago is a Georgian artist.

Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Structural Understanding

Next Word Prediction

Born in 1951 in Tbilisi,
Iago is a Georgian artist.

Predict single words

Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Next Word Prediction

Born in 1951 in Tbilisi,
Iago is a Georgian artist.

Predict single words

Structural Understanding

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth city person

Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Next Word Prediction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Predict single words

Structural Understanding

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth

city

person

Predict structures

Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Structural Understanding

Next Word Prediction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Predict single words

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth

city

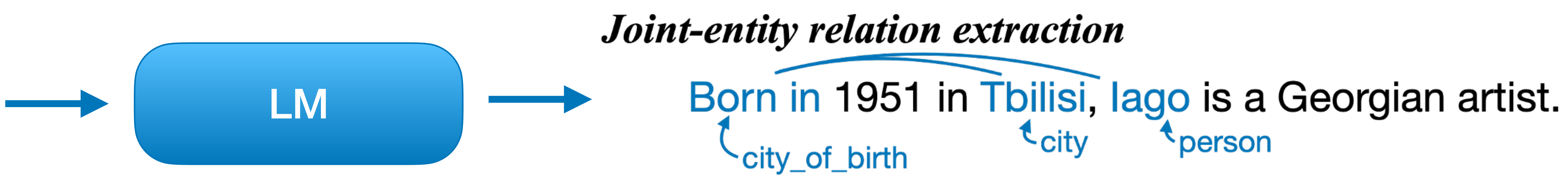
person

Predict structures

Structural understanding can be more difficult than traditional understanding

Why is structural understanding challenging for LMs?

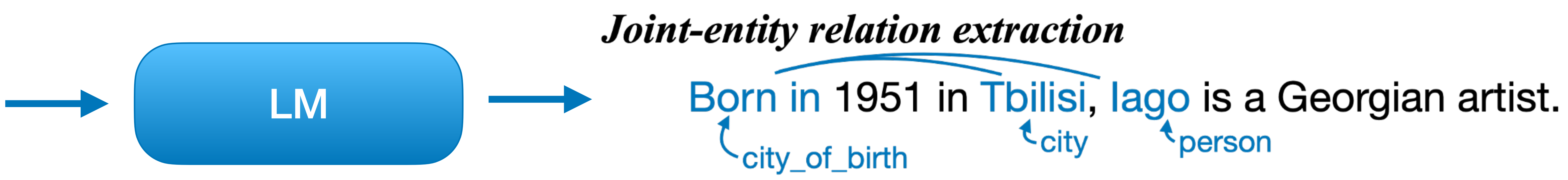
Born in 1951 in Tbilisi,
Iago is a Georgian artist.



Why is structural understanding challenging for LMs?

Challenge 1: Representation for structure

Born in 1951 in Tbilisi,
Iago is a Georgian artist.



Why is structural understanding challenging for LMs?

Challenge 1: Representation for structure

Born in 1951 in Tbilisi, Iago is a Georgian artist.

LM

Named entity recognition

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city person

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth city person

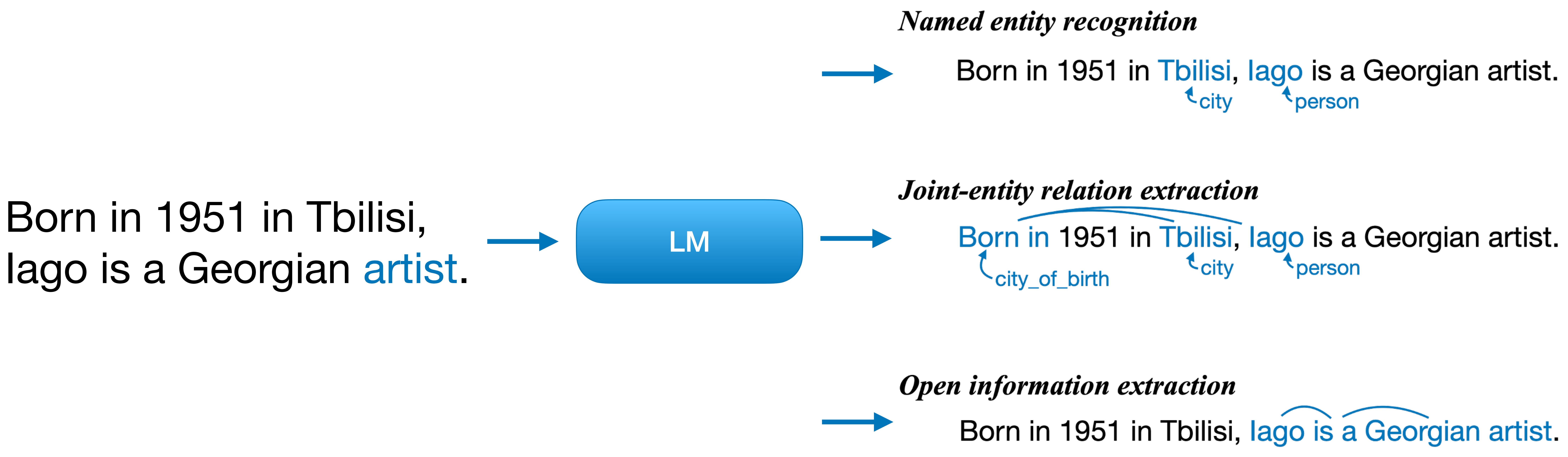
Open information extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Why is structural understanding challenging for LMs?

Challenge 1: Representation for structure

Challenge 2: Unifying different structure prediction tasks



DeepStruct: Produce triples from text

Born in 1951 in Tbilisi, Iago is a Georgian artist. → LM →

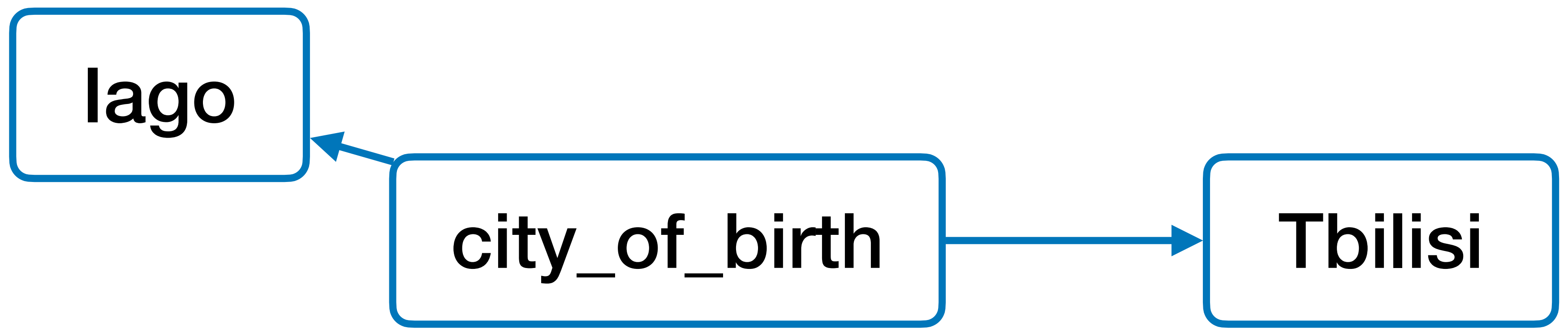
DeepStruct: Produce triples from text

Joint-entity relation extraction

Born in 1951 in Tbilisi,
Iago is a Georgian artist.

DeepStruct: Produce triples from text

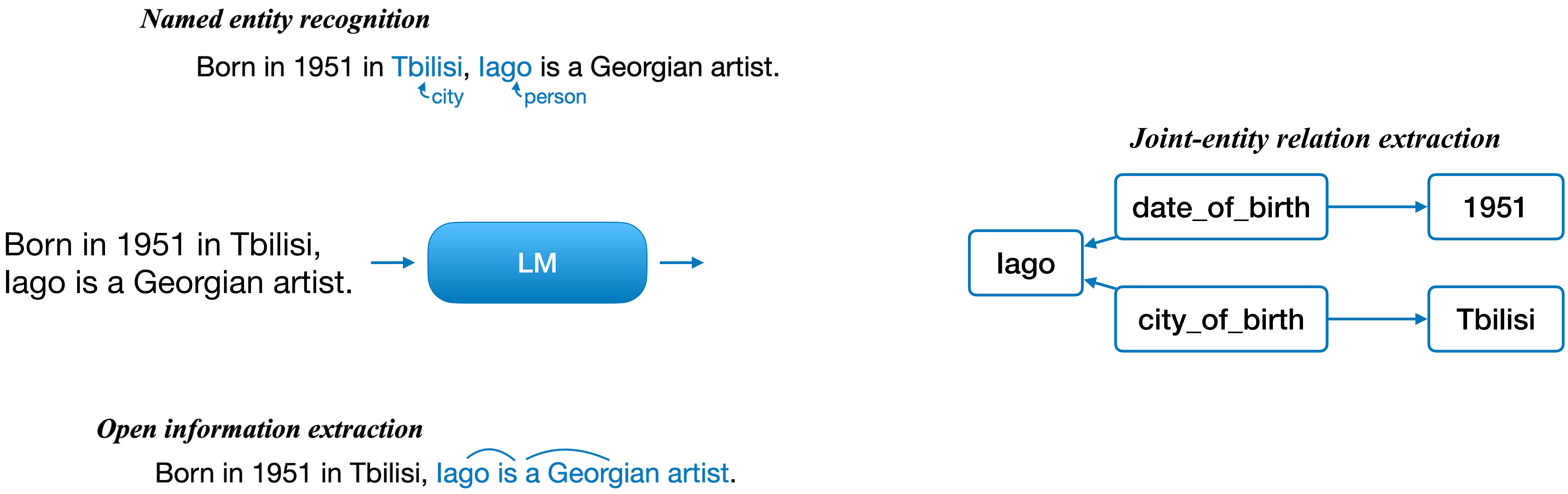
Born in 1951 in Tbilisi,
Iago is a Georgian artist.



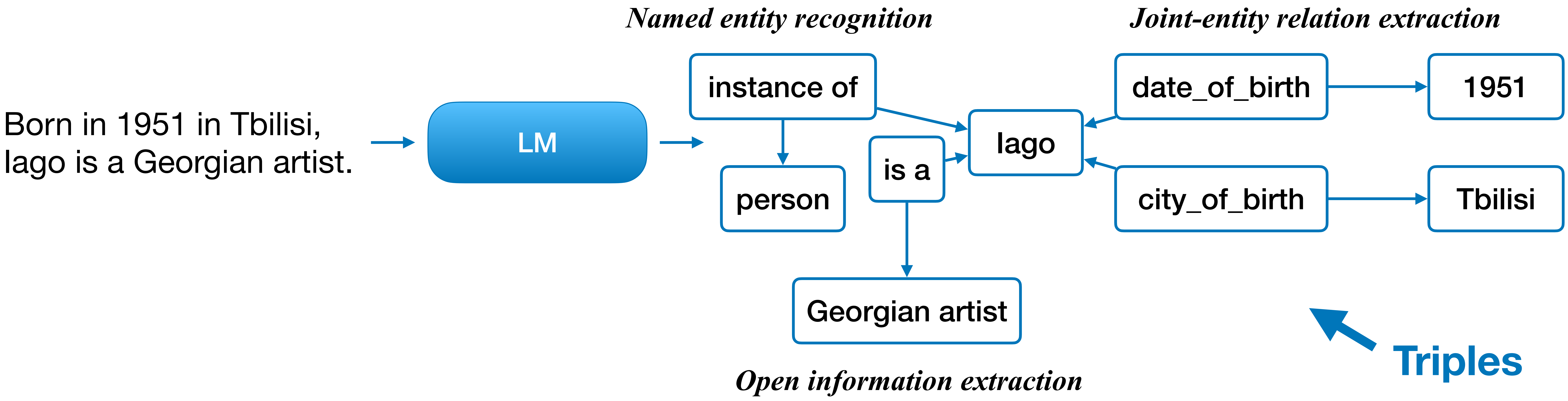
DeepStruct: Produce triples from text

Born in 1951 in Tbilisi,
Iago is a Georgian artist.

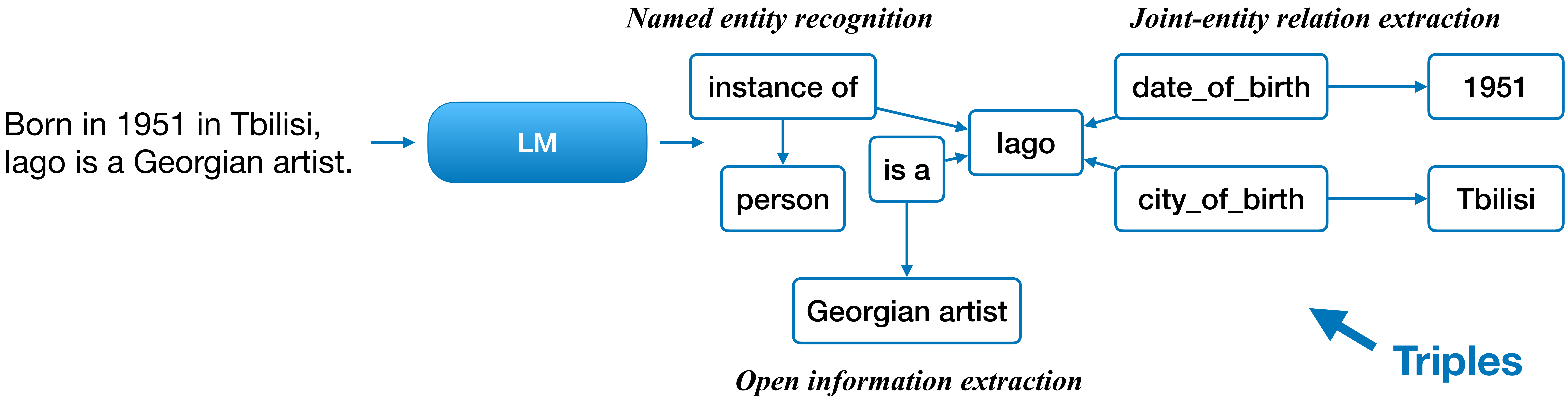
DeepStruct: Produce triples from text



DeepStruct: Produce triples from text

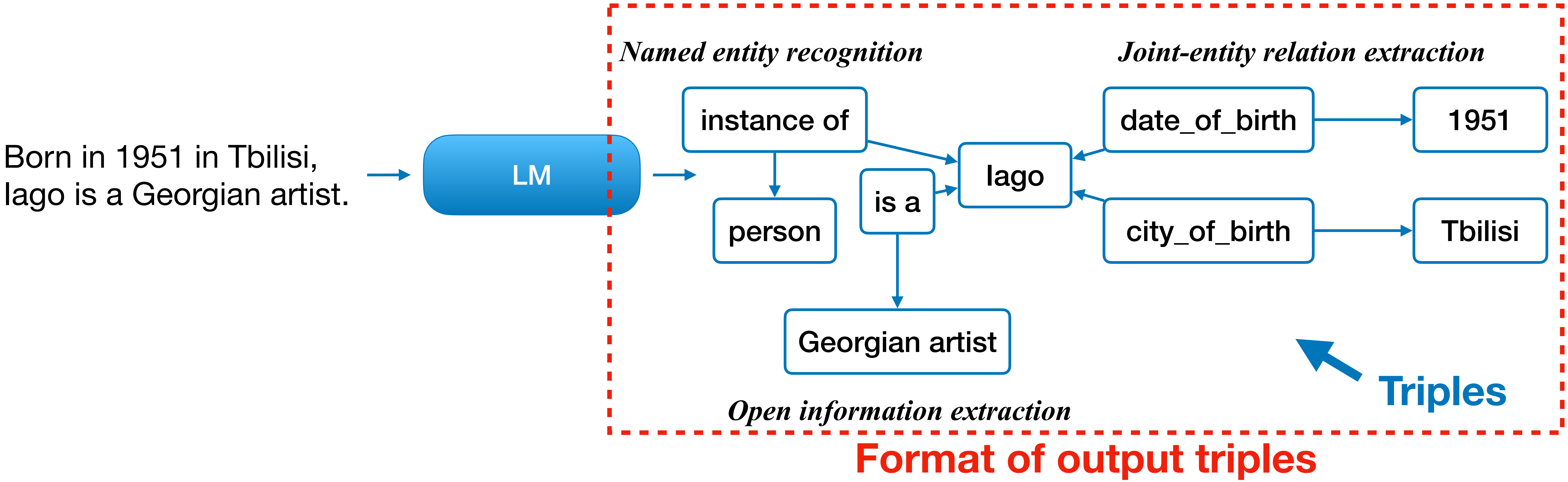


DeepStruct: Produce triples from text



Structure representation formulated as text-to-triple generation problem for LM

DeepStruct: Produce triples from text



Structure representation formulated as text-to-triple generation problem for LM

DeepStruct: Format of output triples

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth city person

(Iago; instance of; person)
(Tbilisi; instance of; city)
(Iago; city_of_birth; Tbilisi)

Named entity recognition

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city person

(Iago; instance of; person)
(Tbilisi; instance of; city)

Structure representation formulated as text-to-triple generation problem for LM

DeepStruct: Format of output triples

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth city person

(Iago; instance of; person)
(Tbilisi; instance of; city)
(Iago; city_of_birth; Tbilisi)

Named entity recognition

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city person

(Iago; instance of; person)
(Tbilisi; instance of; city)

Open information extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

(Iago; is a; Georgian artist)

Structure representation formulated as text-to-triple generation problem for LM

DeepStruct: Format of output triples

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth city person

(Iago; instance of; person)
(Tbilisi; instance of; city)
(Iago; city_of_birth; Tbilisi)

Named entity recognition

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city person

(Iago; instance of; person)
(Tbilisi; instance of; city)

Open information extraction

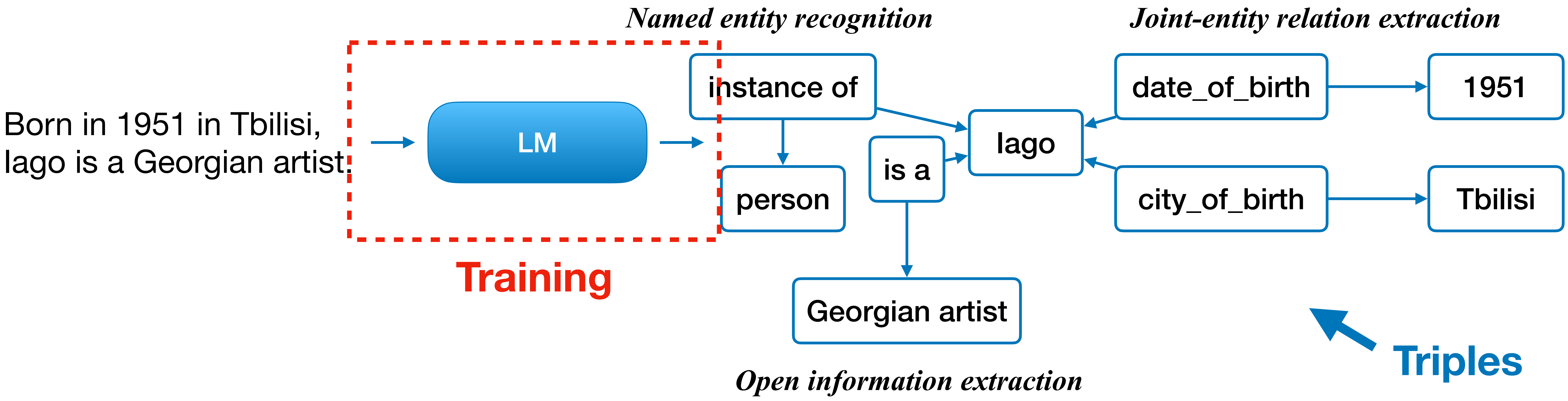
Born in 1951 in Tbilisi, Iago is a Georgian artist.

(Iago; is a; Georgian artist)

...

Structure representation formulated as text-to-triple generation problem for LM

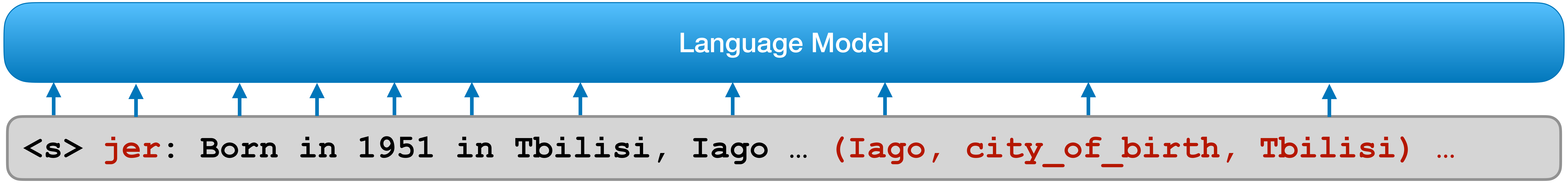
DeepStruct: Produce triples from text



Structure representation formulated as text-to-triple generation problem for LM

DeepStruct: Training

Task: ***Joint-entity relation extraction***
Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.
Desired Output: (Iago, city_of_birth, Tbilisi), …

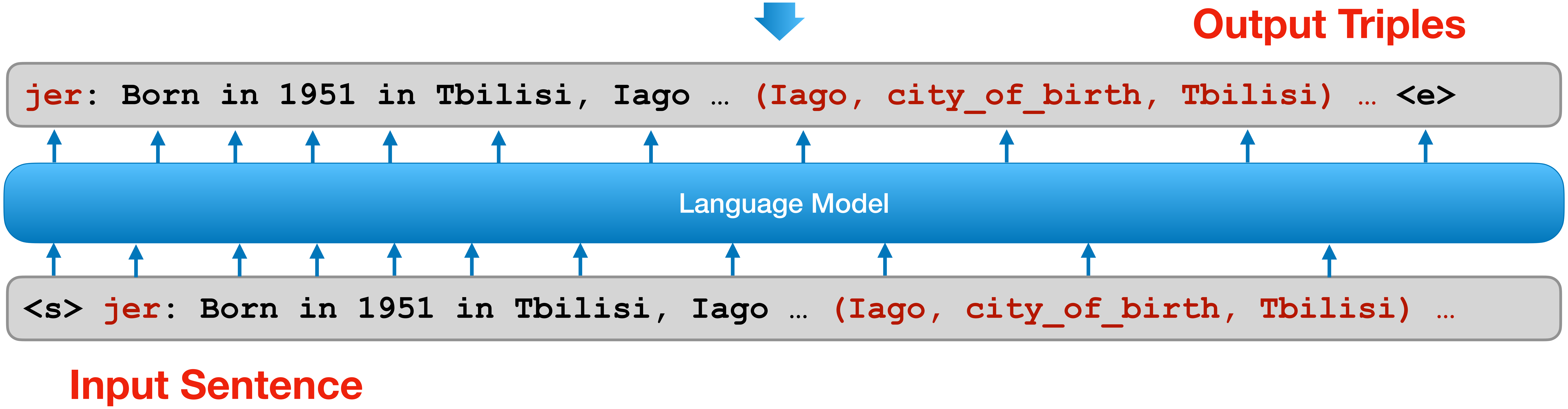


Input Sentence

DeepStruct concats input text and structure triple for autoregressive training

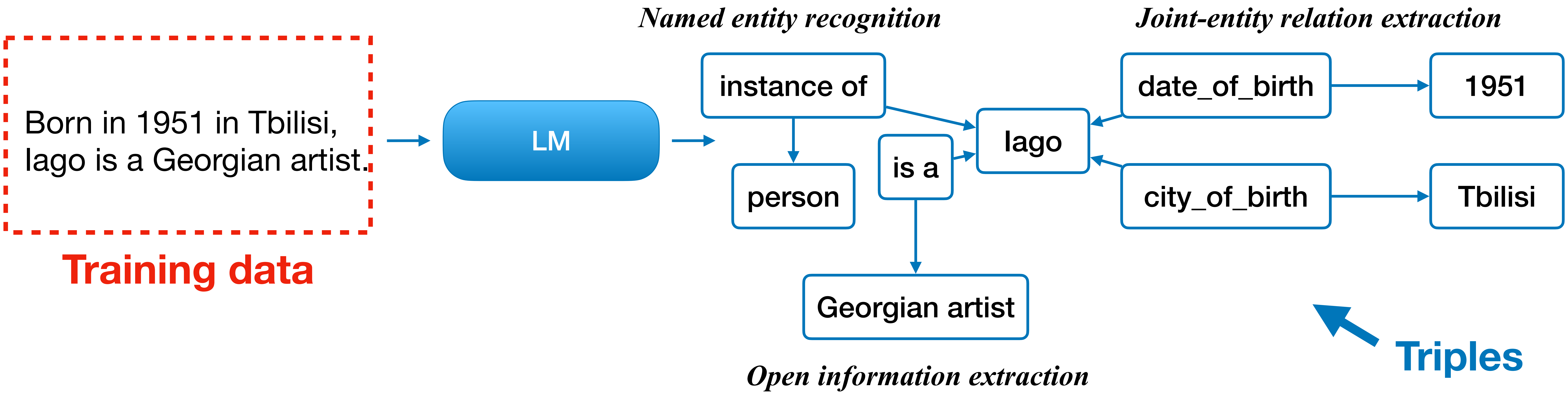
DeepStruct: Training

Task: ***Joint-entity relation extraction***
Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.
Desired Output: (Iago, city_of_birth, Tbilisi), …



DeepStruct concats input text and structure triple for autoregressive training

DeepStruct: Produce triples from text

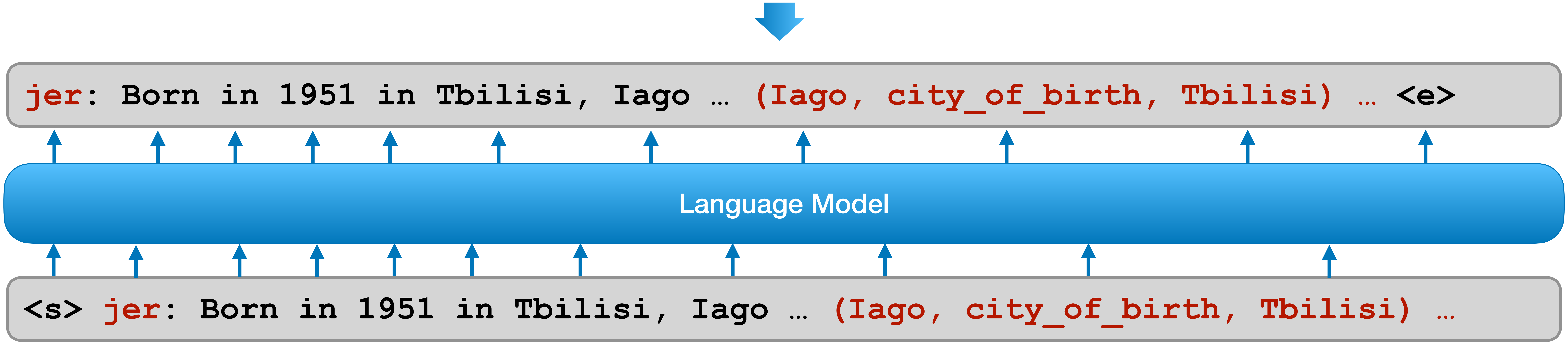


Structure representation formulated as text-to-triple generation problem for LM

DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets

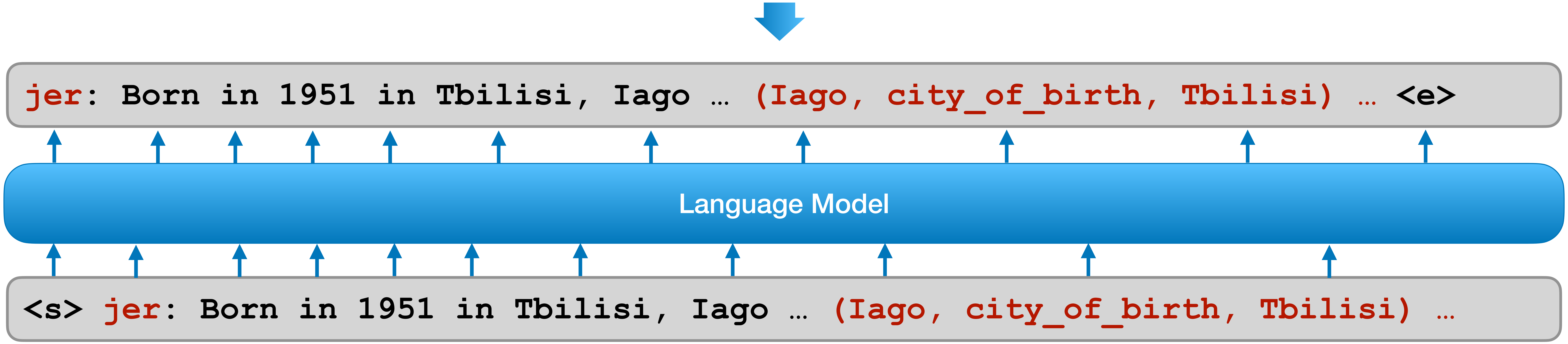


DeepStruct could incorporate both task-agnostic and multi-task data

DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets



DeepStruct could incorporate both task-agnostic and multi-task data

DeepStruct: Task-agnostic datasets

Dataset Source

6 publicly available datasets:

- T-REx
- TEKGEN
- KELM
- WebNLG
- ConceptNet
- OPIEC

Dataset Statistics

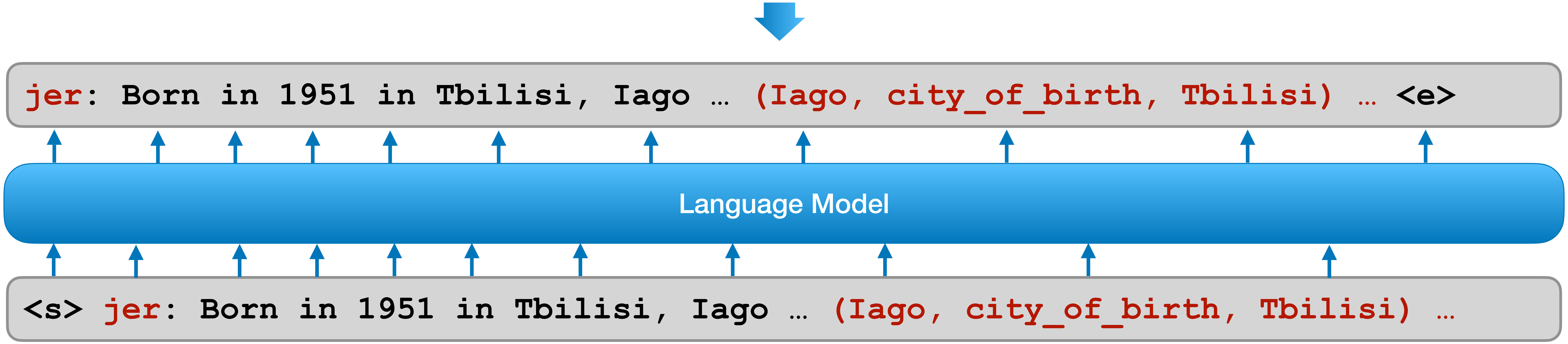
~ 51M sentences
~ 134M entities
~ 114M relations (triples)

DeepStruct is trained on a large task-agnostic corpus

DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets

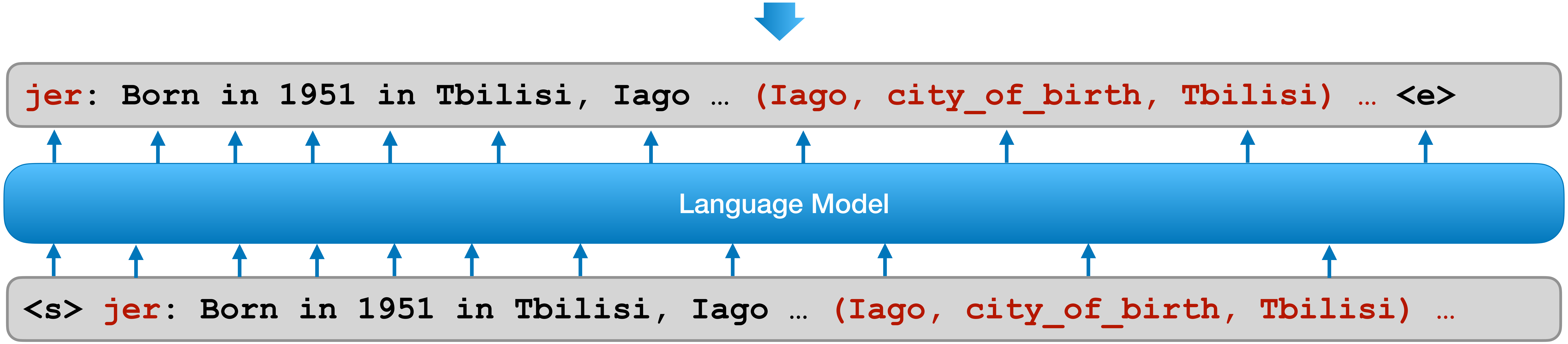


DeepStruct could incorporate both task-agnostic and multi-task data

DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets



DeepStruct could incorporate both task-agnostic and multi-task data

DeepStruct: Multi-task datasets

28 Datasets **10 Tasks** **~ 700K sentences**

Event Extraction
- ACE2005 Trigger Id
- ACE2005 Argument Id
- ACE2005 Trigger Cl
- ACE2005 Argument Cl

Semantic Role Labeling
- CoNLL05 WSJ
- CoNLL05 Brown
- CoNLL12

Coreference Resolution
- CoNLL12

Relation Classification
- TACRED
- FewRel 1.0

Open Information Extraction
- OIE2016
- WEB
- NYT
- PENN

Joint Entity and Relation Extraction
- CoNLL04
- ADE
- NYT
- ACE2005

Dialogue State Tracking
- MultiWOZ 2.1

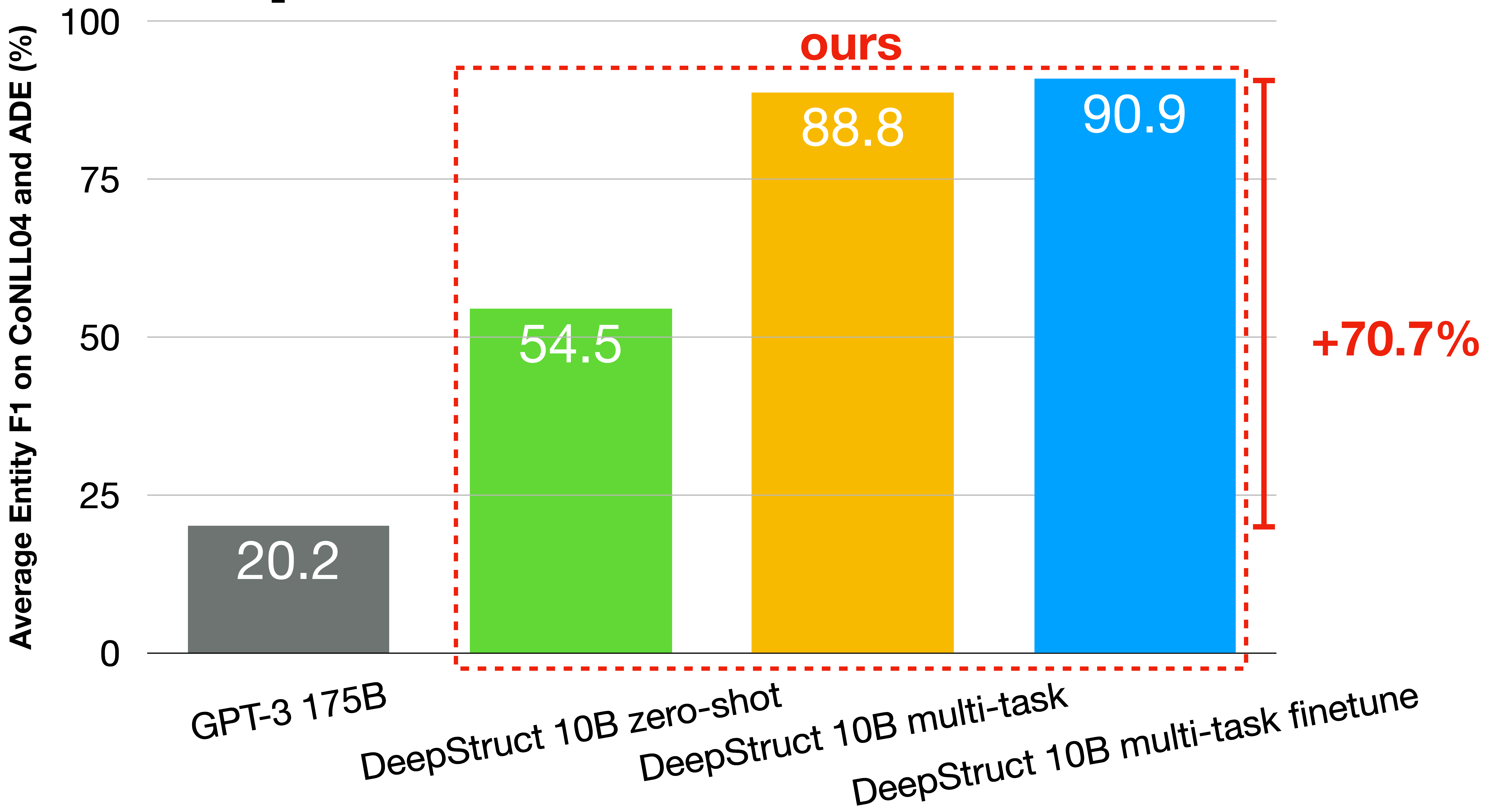
Intent Detection
- ATIS
- SNIPS

Named Entity Recognition
- OntoNotes
- CoNLL03
- GENIA
- ACE2005

Factual Probe
- Google-RE
- T-REx

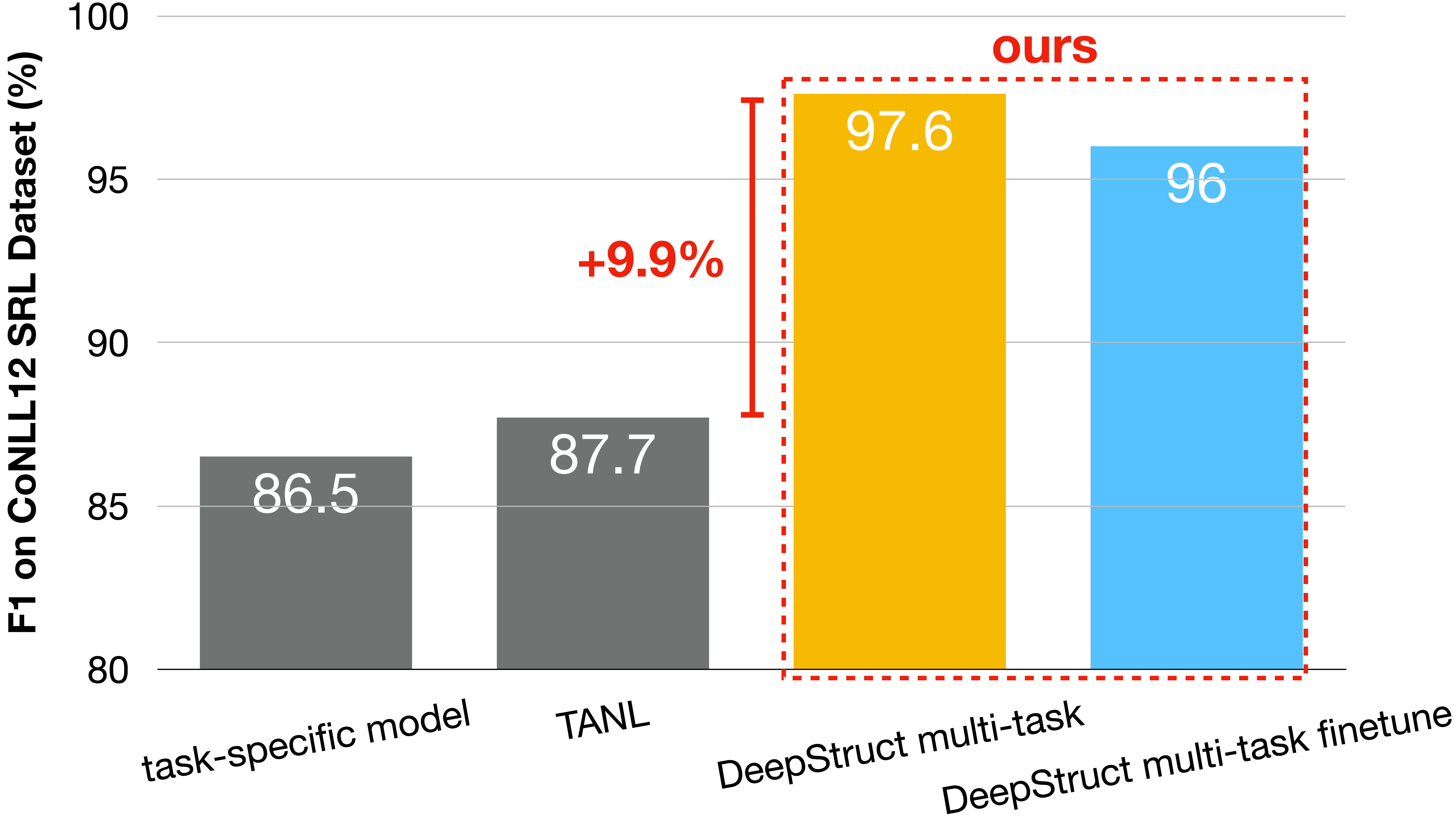
DeepStruct supports a wide range of downstream applications

Result: DeepStruct 10B vs GPT-3 175B



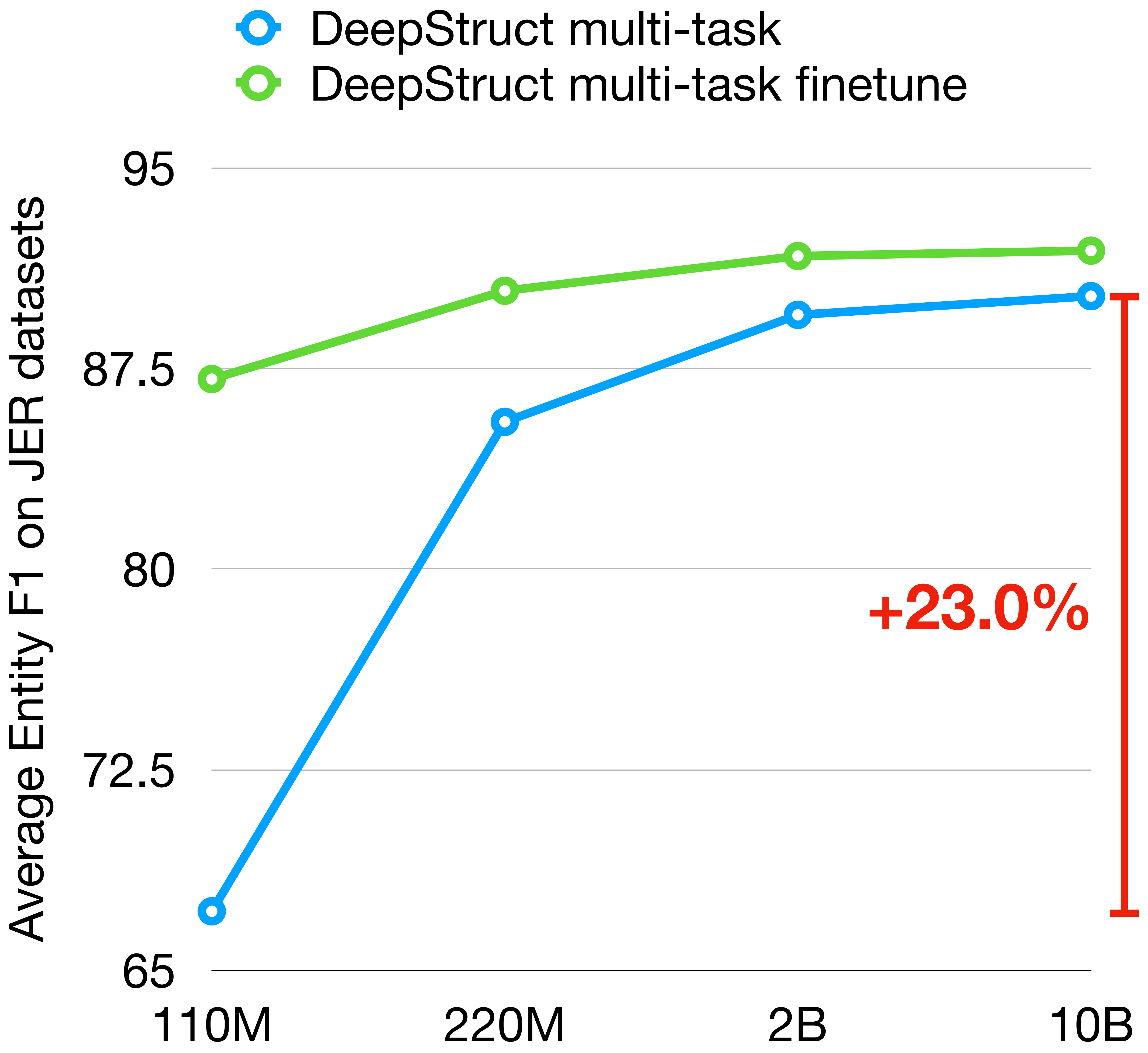
DeepStruct 10B model remarkably outperforms GPT-3 175B model

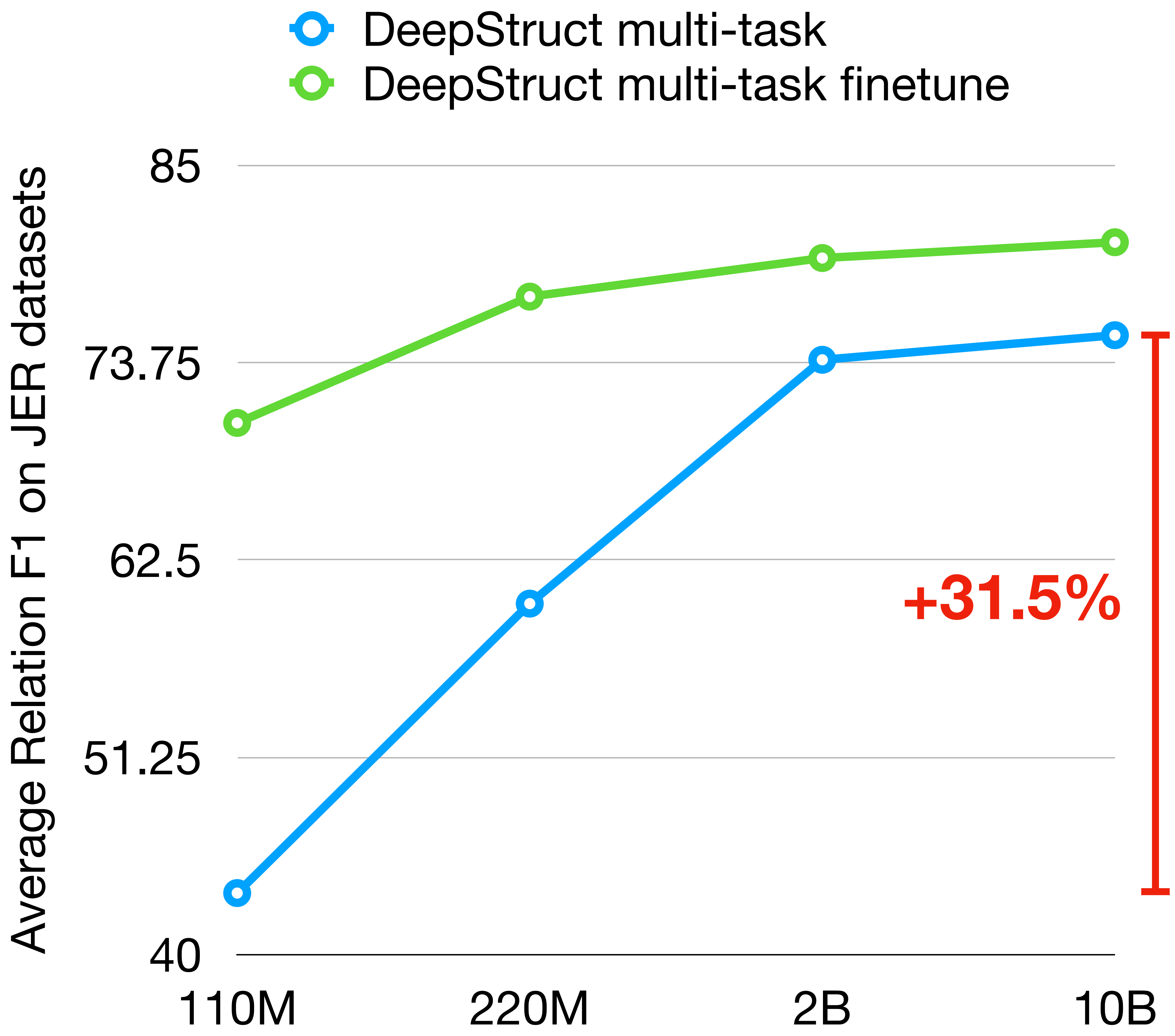
Result: 10 tasks and 28 datasets



DeepStruct achieved state-of-the-art result on 21 of 28 datasets over 10 tasks

Scaling Effect





Larger model further improves DeepStruct performance

Conclusion

DeepStruct: train LM to produce triples from text

DeepStruct 10B zero-shot model largely outperforms GPT-3 175B

State-of-the-art on 21 of 28 datasets over 10 tasks

Code: https://github.com/cgraywang/deepstruct

Thank you for your time and interest!